

UCSF

UC San Francisco Previously Published Works

Title

Charting a Roadmap for Heart Failure Biomarker Studies

Permalink

<https://escholarship.org/uc/item/4cv442x9>

Journal

JACC Heart Failure, 2(5)

ISSN

2213-1779

Authors

Ahmad, Tariq
Fiuzat, Mona
Pencina, Michael J
et al.

Publication Date

2014-10-01

DOI

10.1016/j.jchf.2014.02.005

Peer reviewed



Published in final edited form as:

JACC Heart Fail. 2014 October ; 2(5): 477–488. doi:10.1016/j.jchf.2014.02.005.

Charting a Roadmap for Heart Failure Biomarker Studies

Tariq Ahmad, MD, MPH^{*,†}, Mona Fiazat, PharmD^{*,†}, Michael J. Pencina, PhD^{*}, Nancy L. Geller, PhD[‡], Faiez Zannad, MD, PhD[§], John G. F. Cleland, MD^{||}, James V. Snider, PhD[¶], Stephan Blankenberg, MD[#], Kirkwood F. Adams, MD^{**}, Rita F. Redberg, MD, MPH^{††}, Jae B. Kim, MD^{‡‡}, Alice Mascette, MD^{§§}, Robert J. Mentz, MD^{*}, Christopher M. O'Connor, MD^{*,†}, G. Michael Felker, MD, MHS^{*,†}, and James L. Januzzi, MD^{*,|||}

^{*}Division of Cardiology, Duke University Medical Center, Durham, North Carolina [†]Duke Clinical Research Institute, Durham, North Carolina [‡]Office of Biostatistics Research, National Heart, Lung, and Blood Institute, zBethesda, Maryland [§]Nancy University, Nancy, France ^{||}University of Hull, Yorkshire, United Kingdom [¶]Critical Diagnostics, San Diego, California [#]Johannes Gutenberg University, Mainz, Germany ^{**}Division of Cardiology, University of North Carolina, Chapel Hill, North Carolina ^{††}Division of Cardiology, UCSF, San Francisco, California ^{‡‡}Amgen, Thousand Oaks, California ^{§§}Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, xxBethesda, Maryland ^{|||}Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts.

Abstract

Heart failure is a syndrome with a pathophysiological basis that can be traced to dysfunction in several interconnected molecular pathways. Identification of biomarkers of heart failure that allow measurement of the disease on a molecular level has resulted in enthusiasm for their use in prognostication and selection of appropriate therapies. However, despite considerable amounts of information available on numerous biomarkers, inconsistent research methodologies and lack of clinical correlations have made bench-to-bedside translations rare and left the literature with countless publications of varied quality. There is a need for a systematic and collaborative approach aimed at definitively studying the clinical benefits of novel biomarkers. In this review, on the basis of input from academia, industry, and governmental agencies, we propose a systematized approach based on adherence to specific quality measures for studies looking to augment current prediction model or use biomarkers to tailor therapeutics. We suggest that study quality, rather than results, should determine publication and propose a system for grading biomarker studies. We outline the need for collaboration between clinical investigators and statisticians to introduce more advanced statistical methodologies into the field of biomarkers that would allow for data from a large number of variables to be distilled into clinically actionable information. Lastly, we propose the creation of a heart failure biomarker consortium that would allow for a comprehensive list of biomarkers to be concomitantly analyzed in a pooled sample of randomized clinical trials and hypotheses to be generated for testing in biomarker-guided trials.

Such a consortium could collaborate in sharing samples to identify biomarkers, undertake meta-analyses on completed trials, and spearhead clinical trials to test the clinical utility of new biomarkers.

Keywords

heart failure; biomarkers; studies

“Out of clutter, find simplicity. From discord, find harmony.”

Albert Einstein (1)

Heart failure is among the leading causes of death and disability worldwide (2). Among the challenges in treating this patient population are inadequacies in prediction of disease severity, with a resultant mismatch between risk stratification and intensity of therapy (3). Identification of biomarkers that allow measurement of the disease on a molecular level has resulted in considerable enthusiasm for their use in prognostication and selection of appropriate therapies (4–6). Illustrating this point, a PubMed search of the phrase “biomarkers in heart failure” resulted in close to 6,500 publications over the last decade (Fig. 1) (5). Reasons for this include high-throughput molecular biology techniques that allow increased availability of point-of-care, rapid-turnaround biomarker testing, and reductions in costs of analysis (7).

Nonetheless, the current state of biomarker research in heart failure is one of exponential gains in information that have far exceeded the ability to contextualize findings: more “interesting data” than “clinically actionable information”. As a result, despite considerable amounts of information available on numerous biomarkers, inconsistent research methodologies, insufficient study size, and lack of clinical correlations have made bench-to-bedside translations rare, leaving the literature with numerous publications of varied quality. This has led to slow adoption of established biomarkers, debates about the utility of biomarkers in standard clinical care, and delays in approval by regulatory agencies for clinical use (8,9). For example, the clinical efficacy of natriuretic peptide–guided therapy in heart failure remains unclear, despite 12 studies over the last decade that included >2,500 patients; fortunately, a multicenter trial is currently underway (GUIDE-IT [Guiding Evidence Based Therapy Using Biomarker Intensified Treatment]; NCT01685840) to provide clearer recommendations (10). Furthermore, considering the large number of biomarkers discussed in the literature, it is worth noting that beyond the natriuretic peptides, only galectin-3 and serum ST2 (sST2) are cleared by the Food and Drug Administration (FDA) for use as aids in assessing prognosis in heart failure. Even so, the appropriate clinical use of these 2 novel biomarkers is unclear due to a shortage of well-designed studies informing their proper clinical use, accompanied by a large number of studies repeatedly depicting their prognostic value.

This paper describes unanswered questions in the field of heart failure biomarkers and recommends a roadmap for further studies that will provide more definitive answers about the clinical role of biomarkers in diagnosis, prognosis, and treatment of heart failure. It is based on discussions among cardiologists, epidemiologists, clinical trialists, statisticians,

journal editors, and regulatory agency representatives at the ninth annual Global Cardio Vascular Clinical Trialists Forum (CVCT) in Paris, France.

What Qualifies as a Useful Heart Failure Biomarker?

In theory, a biomarker can be any measurement made on a biological system. In practice, however, biomarkers in heart failure typically refer to substances measured in the blood other than commonly used laboratory tests and imaging studies (6).

To understand the added clinical utility of biomarkers in heart failure, criteria have been previously recommended by van Kimmenade and Januzzi (Table 1) (5,11). When biomarkers are considered for clinical use (together with other clinical parameters and cardiac tests), currently only the natriuretic peptides meet the proposed standards. The majority of the remaining emerging biomarkers remain entangled in debates as to whether they provide any incremental value over established clinical measurements.

We argue that instead of the current piecemeal approach in which biomarkers are evaluated with a variety of statistical approaches with or without comparisons with other markers, there is a need for standardized methodologies for clinical assessment of biomarkers in heart failure. Presently, the vast majority of biomarker publications in heart failure are related to prognosis, and considerable opportunities exist for harmonization of research methods for these studies. Although demonstration of prognostic value is of importance, modification of therapeutics based on biomarker values in a time-sensitive and cost-effective manner that improves patient outcomes is the *sine qua non* of a useful biomarker. Unfortunately, studies attempting to address these questions are scarce.

Biomarkers for Diagnosis

To date, heart failure biomarkers have had their greatest impact in the realm of disease diagnosis. Prior to the widespread use of natriuretic peptides for this purpose, clinicians relied on data from subjective variables such as clinical symptoms and physical examination findings to diagnose patients with heart failure (7). In 2002, publication of the BNP (Breathing Not Properly) study triggered a paradigm shift toward a biomarker-based method of evaluating dyspnea, and widespread use of both B-type natriuretic peptide (BNP) and N-terminal proBNP (NT-proBNP) has since then fundamentally impacted the standard of care in heart failure (7,12). Confirmed by studies such as PRIDE (ProBNP Investigation of Dyspnea in the Emergency Department), natriuretic peptide assessment for diagnosis of heart failure in the setting of clinical uncertainty now has the highest level of recommendation by all major professional cardiology societies (13,14).

Despite the widespread use of BNP assays for diagnosis of heart failure, there remains a lack of well-defined and accepted diagnostic cutoffs. Additionally, elevations in natriuretic peptide levels can occur as a result of several cardiac and noncardiac disease states, making the negative predictive value of the test most clinically helpful (13). As a result of these limitations, there is a need for better diagnostic algorithms, potentially through addition of novel biomarker information, to objective clinical and natriuretic peptide data (15). Ideal diagnostic biomarkers would feature rapid sustained elevation, high tissue specificity

(myocardial origin), release proportional to disease extent, and assay features conducive to high quality point-of-care testing (16). Additionally, as discussed below, clinical diagnostic algorithms involving biomarkers should aim for high levels of sensitivity, specificity, predictive values, as well as low overall cost (16).

Biomarkers for Risk Stratification

Risk assessment is important in the care of patients with heart failure because many key therapeutic decisions depend on these evaluations. Effective therapies in heart failure such as implanted devices and cardiac transplantation are complex, costly, and rely on accurate risk stratification (17). Therefore, an ideal prognostic biomarker in heart failure should allow for early identification of individuals at risk for adverse clinical outcomes and should be relatively easy to measure, with acceptable costs (5,15). The biomarker measurement should display accuracy (the test is measuring what it is supposed to measure and generalizability (capacity to provide accurate predictions in population samples different than that in which the biomarker was originally validated). Lastly, the biomarker should not provide information that is already obvious at the bedside or through measurement of clinically available biomarkers.

Prognostic studies of novel biomarkers should have well-defined outcome measures, and when sufficiently powered, they should report 3 measures of prognostic accuracy: discrimination, calibration, and reclassification. These measures should be standardized as much as possible to provide consistency between studies (18). Although these approaches are becoming more routinely performed in modern heart failure studies, they are by no means considered mandatory—improvement in this regard is needed.

Other common shortcomings in biomarker studies include the use of inadequately modeled survival analyses, use of low-quality comparator assays, and lack of validation. We suggest that prognostic models examining a novel biomarker must include currently used clinical measures as covariates, such as New York Heart Association functional class, left ventricular ejection fraction, and natriuretic peptides using high-quality assays. Also, whenever possible, findings from novel biomarker studies should be verified in an independent sample of patients to demonstrate generalizability.

Statistical Considerations

Despite significant overlaps, statistical evaluation of novel biomarkers depends on intended clinical use. Detailed standards have been proposed for designing and reporting the results of studies evaluating the performance of bio-markers for diagnosis and for prognosis (11,16,19). Here, we provide an outline of these guidelines (Table 2, Fig. 2).

Diagnostic biomarkers.

For a diagnostic biomarker, the first step is evaluating test accuracy in terms of its sensitivity (detection of disease when disease is truly present) and specificity (recognition of absence of disease when disease is truly absent) at clinically relevant cut points. This information can be summarized using receiver-operating characteristic (ROC) curves to illustrate the trade-

off between sensitivity and specificity. Likelihood ratios are calculated with these data and address the likelihood of obtaining a positive test result in someone with disease compared with someone without disease, as well as the converse (16). Clinical use would depend on the balance between consequences of missing disease versus overdiagnosis.

Prognostic biomarkers.

For prognostic biomarkers used to track disease progression, the key factors to consider are discrimination, calibration, and reclassification. Table 2 summarizes the pros and cons of various statistical methods used to address these factors in biomarker studies.

Discrimination reflects the ability of a prognostic model to identify a clinical status (event vs. nonevent) and is of particular importance to clinicians whose decision making may hinge on the ability to predict an outcome. *Calibration* measures how much the model estimate of a specific outcome matches the “real” probability of the outcome (20,21). The key difference between discrimination and calibration is that the former reflects the ability of a given prognostic biomarker to distinguish between an event versus nonevent, whereas calibration measures how frequently the estimation by the model matches the real outcome. Whereas statistical methods for gauging calibration exist, we prefer to rely on visual displays: instead of categories, one might look at the plot of predicted versus smoothed observed risk (11). When logistic regression is used, it is preferable to look at the slope and intercept of a model that uses predicted risks as independent variables and true outcome status as the dependent variable (22). Below we summarize statistical tests of discrimination.

C-statistic

The most commonly used measure to evaluate the discriminatory ability of a new biomarker in heart failure is the area under the ROC curve (AUC), often referred to as the C-statistic (20,21,23,24). The ROC curve plots 1 the specificity (the false-positive rate) on the x-axis versus the sensitivity (the true positive rate) on the y-axis for different biomarker concentrations for a fixed follow-up time. The AUC reports the probability of correctly ranking cases and noncases appropriately: If the AUC has a value of 0.50, it is as if scores were randomly assigned to individuals and the marker has no better prognostic value than flipping a coin. As the AUC approaches 1.0, it demonstrates increasing discrimination or separation of diseased and nondiseased. An advantage of this approach is the ability to compare changes in AUC that result from the addition of a new biomarker to the predictive model that already contains clinical variables and other biomarkers: a significant increase in AUC might indicate significant “value” beyond the variables in the model. However, small increases in the AUC may be clinically irrelevant, whereas a lack of change in the AUC may result from an overly optimistic model containing variables not typically available to the clinician. For models that are already very good, the change in AUC can be an insensitive measure of model improvement; markers with large associations, and a large accompanying relative risk, may have little effect on the ROC curve. Furthermore, AUC tends to be heavily influenced by population admixture and the distribution of major risk factors (e.g., using the same model, if the sample of interest has a wide distribution of age, the AUC will be much larger than a case in which the distribution is narrow). Finally, AUC cannot adequately capture the clinical usefulness of new markers in situations in which treatment decisions are based on established categories. Thus, although ROC analyses are considered standard for

the general comparison of results from biomarkers, they are not conclusive and differences in AUC alone should not be considered sufficient to argue clinical utility.

Reclassification metrics

The integrated discrimination improvement (IDI) offers an alternative way to quantify the incremental value of new biomarkers (25). It has a simple and intuitive interpretation as an increase in the distance between the mean risks for events versus nonevents. However, judging the magnitude of the IDI may be context specific because the values depend on the incidence rate of the outcome of interest. To account for this, Pencina et al. (26) proposed heuristic benchmarks based on the relative IDI that put the increment offered by the new biomarker in the context of the average value contributed by each of the predictors already incorporated into the model.

The IDI shares with the AUC the issue of unequal clinical implications of various risk classifications. As a response to this, risk reclassification measures have been introduced and are being increasingly used for evaluation of novel biomarkers. Of these, the net reclassification improvement (NRI) is most commonly used (20,21,25,27). The NRI summarizes net changes of allocation in clinically meaningful risk categories for events and nonevents when a novel predictor (biomarker) is added to an existing model. It is computed by summing up the proportions of correctly upward classified events (correctly qualifying for treatment) and downward classified nonevents (correctly qualifying for no treatment) subtracted by the proportions of incorrectly downward classified events (incorrectly qualifying for no treatment) and upward classified nonevents (unnecessarily qualifying for treatment). This approach is readily applied in situations in which categories of risk are well established.

Limitations of reclassification analyses include the fact that a lack of “gold standard” categories of risk may result in the use of arbitrary risk strata that may bias the result either in favor of a novel biomarker or against it. However, even if the risk models are valid and the risk categories chosen are clinically relevant, current methods of reporting NRI may not meaningfully summarize improvements in risk reclassification in all instances because changes in risk categories are counted equally and such changes may not reflect clinically equivalent events. Additional work is needed to develop a consensus for quantifying improvements in risk prediction performance and linking these to clinically important events.

Competing risks

Although the statistical methodology described is not specific to it the issue of competing risks is of particular importance in heart failure, in which patients can suffer from severe comorbid conditions and different treatments exist for the risk of sudden death versus pump failure. The simplest approach to competing risks is to censor follow-up at the time of their occurrence. However, this approach implicitly assumes that those censored could still develop the primary event of interest, which is not true in the example of death due to other causes. We recommend the competing risk model proposed by Fine and Gray (28). The use of competing risk models also raises the question of appropriate performance metrics. This

choice might depend on the role of the competing event in the analysis. If not considered clinically important, the method outlined might still be applicable. On the other hand, if we have multiple and equally important outcomes that we try to model, performance methods specific to multinomial outcomes might be used (29).

In summary, when reporting results of biomarker studies reporting on prognostication, we propose the following statistical steps be followed. First, the study design and levels of standard risk factors as well as the joint prognostic strength should be reported. Against this backdrop, the effect of the new marker on relative risk should be presented, with P values and CIs. Then, impact of model discrimination should be quantified using AUCs and the IDI. If meaningfully risk categories exist, the NRI can be presented to quantify the impact on risk reclassification. We prefer that the components of the NRI are presented separately for events and nonevents. Furthermore, following the newest research developments in this area, we recommend that no statistical testing be conducted and no P values reported for measures of incremental value in model performance (30). Instead, once the new biomarker has been shown to be significantly associated with the outcome in a multivariable model, CIs should be given to quantify the precision of the estimate of incremental value.

If the biomarker appears promising according to these statistics, one should consider more formal assessment of clinical utility. Simple and appealing choices of metrics in this regard include the increment in net benefit and weighted NRI (31,32). Both measures quantify the net gain offered by the new biomarker: the net benefit does it in terms of gains in true positives and weighted NRI in terms of monetary cost or quality-adjusted life-years.

How Can We Use Biomarkers in Heart Failure?

The most relevant question in the field of heart failure biomarkers is whether the addition of a biomarker to clinical work flow improves clinical decision making to the degree that the added cost and complexity are justified. Such an approach has been successfully used in the treatment of patients suspected to have acute coronary syndromes: cardiac troponin measurements routinely influence aggressiveness of care. In preventive cardiology, cholesterol levels and high-sensitivity C-reactive peptide (CRP) levels influence the use of medical therapies (33). In the field of heart failure, however, beyond the use of natriuretic peptides for diagnosis, the concept that a biomarker can influence therapeutic decisions remains in its infancy. The reasons for this range from complexity of the disease state to a lack of definitive evidence from studies. Because establishing the therapeutic utility of prognostic biomarkers is more challenging and in need of study than evaluating their use as a diagnostic test (which can be done via a cross-sectional design), we will focus our discussion on the former topic.

Prognostic Biomarkers for Treatment Guidance

Despite the considerable number of heart failure biomarker studies in the literature, very few address the question of how prognostic biomarkers might be used for therapeutic guidance. We suggest that this should be a pivotal consideration for future studies of novel heart failure biomarkers—a prognostic biomarker may be of no clinical consequence if a meaningful response cannot be triggered from its measurement. We can use lessons learned

from the evaluation of natriuretic peptides as well as other biomarkers in cardiology to recommend guidelines for biomarker-targeted studies that help demonstrate whether a prognostic biomarker can also classify patients into distinct subgroups that respond differently to therapy. To address this question, we describe the 3 most commonly used methods: biomarker-stratified, enrichment, and biomarker strategy designs. Elements of each approach may be combined (Fig. 3) (34).

Biomarker-stratified designs can be used if there is no definitive evidence to suggest that treatment efficacy may depend on biomarker levels. All patients are randomly assigned to treatments, but the results are analyzed according to biomarker status. This design maximizes the advantage of randomization by providing unbiased estimates of benefits across biomarker categories as well as for the entire population. This design is the most commonly reported biomarker analysis in heart failure, and most contemporary clinical trials of testing therapeutics with evaluable plasma have been subject to post hoc biomarker analysis.

Example 1: In the CORONA (Controlled Rosuvastatin Multinational Trial in Heart Failure) study, patients with chronic heart failure were randomly assigned to rosuvastatin or placebo. It was observed that those with galectin-3 levels less than the median (19.0 ng/ml) benefited more from statin therapy (35,36). These results raised the possibility that galectin-3, a biomarker of myocardial fibrosis, might be used to define heart failure subtypes that respond differently to rosuvastatin (37).

Example 2: In the Val-HeFT (Valsartan Heart Failure Trial) study, valsartan caused a significant reduction in heart failure hospitalizations in patients with galectin-3 levels below the median level of 16.2 ng/ml but not in patients with levels above the median (38). These results suggested that galectin-3 might be used to predict benefit from valsartan therapy in patients with heart failure.

Enrichment designs should be performed to test the hypothesis that a biomarker-defined subgroup of patients will benefit from a treatment. This should only be done when there is evidence to suggest that treatment benefit is limited to this subgroup; therefore, treatment is randomized only among a subset of patients. In some cases, natriuretic peptide levels have been used in the inclusion criteria for heart failure trials (e.g., ASTRONAUT [Aliskiren Trial on Acute Heart Failure Outcomes]); in these cases, the study design is a variation along the enrichment design theme as the biomarker is being used as a diagnostic test to confirm presence of heart failure rather than to define a subgroup that might benefit from therapy.

Example 3: The GRAVITAS (Gauging Responsiveness With a VerifyNow Assay-Impact on Thrombosis and Safety) trial evaluated the effects of high-dose compared with standard-dose clopidogrel in patients with high on-treatment platelet reactivity after percutaneous coronary intervention. Patients with stable angina/ischemia or non-ST-segment elevation acute coronary syndrome with high residual platelet reactivity on clopidogrel therapy (P2Y₁₂ reaction units [PRU] ≥ 230) post-intervention were randomized to standard-of-care clopidogrel therapy (75 mg) or high-dose clopidogrel therapy (150 mg) (39). This study

design allowed the investigators to gauge the efficacy of an experimental therapy (high-dose clopidogrel) only in the subset of patients in whom there was clinical equipoise: high PRU on therapy.

Example 4: The JUPITER (Justification for the Use of Statins in Primary Prevention Intervention Trial Evaluating Rosuvastatin) study randomized patients with low-density lipoprotein (LDL) cholesterol levels of less than 130 mg/dl and high-sensitivity CRP (hsCRP) levels of 2.0 mg/l or higher to rosuvastatin or placebo. Prior to the study, it was unknown if patients with LDL cholesterol levels of less than 130 mg/dl would benefit from statin therapy. This trial showed that, among these patients, hsCRP levels might identify those who would derive the clinical benefit from statin therapy (33).

A biomarker strategy design may be used in the presence of reliable data that support varied levels of therapeutic effectiveness according to biomarker concentrations. Patients are randomly assigned to a treatment arm that uses biomarker values to determine therapy. For example, the use of natriuretic peptides to guide the management of chronic heart failure treatments has been recently explored (7,40). Based on favorable findings from a meta-analysis of previous studies and a recently published single-center study, the strategy of including natriuretic peptide reduction as part of the standard approach to heart failure care may soon become standard of care (40,41).

Example 5: The PROTECT (Pro-BNP Outpatient Tailored Chronic Heart Failure Therapy) study followed a biomarker strategy methodology by randomizing patients with systolic heart failure to standard guideline-based care or biomarker-guided care, with the goal to reduce NT-proBNP concentrations below 1,000 pg/ml (40). The study found that patients randomized to the biomarker-guided arm had fewer adverse cardiovascular events, better quality of life and improved echocardiographic parameters at study conclusion (42,43).

Example 6: The STOP-HF (St. Vincent's Screening to Prevent Heart Failure) study randomized patients with risk factors for heart failure to a BNP-guided intervention arm versus usual care. The study found that BNP-based screening and collaborative care reduced rates of adverse events (44).

Example 7: The GUIDE-IT trial (Guiding Evidence Based Therapy Using Biomarker Intensified Treatment; NCT01685840), a prospective, randomized trial of 1,100 patients randomized to usual care versus biomarker-guided therapy is underway to provide definitive information about the efficacy of this approach.

The approaches discussed here may be modified and combined according to the clinical question, *a priori* knowledge about biomarkers being evaluated, and consideration of the pros and cons of each approach (Fig. 3). With consideration of the limitations of these study designs, heart failure clinical studies should employ and modify the designs as needed; they may be imperfect but are a prerequisite to move forward with clinical implementation of new biomarkers.

The Path Beyond Natriuretic Peptides in Heart Failure

A plethora of candidate biomarkers now exist that encapsulate varying aspects of heart failure pathology, with theoretical roles in therapeutic tailoring, but lack data from testing within clinical trials. Instead of a fragmentary approach toward answering these questions, there is a need for a collaborative effort that involves pooling together data from large randomized, controlled heart failure trials with blood samples available for prospective analyses. To achieve this goal, researchers, clinicians, government, and regulatory agencies, as well as the in vitro diagnostic industry will need to align their goals.

In this regard, we recommend the creation of a heart failure biomarker consortium. We envision this being a public-private biomedical research partnership with broad participation from a variety of stakeholders, including government, industry, academia, and potentially patient advocacy groups. An initial goal of this consortium would be hypothesis-driven treatment interaction testing of “first-tier” biomarkers (i.e., those with the most convincing data in support of potential value). Concomitant testing of bio-markers, something that is rarely done—yet so crucially needed—would allow for an appraisal of which ones are best suited as surrogates for specific pathophysiological pathways (e.g., cystatin C vs. neutrophil gelatinase-associated lipocalin for acute kidney injury). This approach would also allow the study of whether efficacy of therapies in heart failure varies according to biomarker levels. The consortium could establish consensus criteria for meaningful differences in prognosis and treatment response; based on these results, future studies could be performed in patients with heart failure with therapeutic choices based on biomarker testing. Involvement of investigators and cohorts from diverse geographic backgrounds would be important because the clinical implications of biomarkers might vary according to patient characteristics. The long-term goal of this consortium would be to identify novel biomarkers that can be transitioned to the bedside as well as spearhead the creation of biobanks within ongoing and future clinical trials. This effort would also facilitate the development of biomarkers using newer technology, help decide which biomarkers should be used for specific decisions in heart failure, provide advice to regulatory bodies, and disseminate information.

Examples of consortia that might serve as models are: 1) HOMAGE (Heart Omics in Ageing), a project aiming to identify and validate specific biomarkers of heart failure; 2) EDRN (Early Detection Research Network), an initiative of the National Cancer Institute that brings together dozens of institutions to help accelerate the translation of biomarker information into clinical applications and to evaluate new ways of testing cancer in its earliest stages; and 3) the Bio-markers Consortium, a public-private biomedical research partnership managed by the National Institutes of Health that aims to discover, develop, and qualify biomarkers to support new drug development, preventive medicine, and medical diagnostics.

Recommendations on Grading Biomarker Studies

Although the use of biomarkers in heart failure has the potential to improve care of patients, large variations in study quality have overwhelmed the literature with contradictory findings, causing confusion, rather than increasing clarity about the value of novel

biomarkers, and added resistance to clinical uptake of established biomarkers. Therefore, we believe that the field would benefit from more consistent methodologies and greater collaboration.

Although general criteria for evaluation of biomarkers have been proposed, they are vague and thus lack necessary impact. In an attempt to provide specific guidelines for biomarker studies in heart failure, we propose a systematized approach based on adherence to specific quality measures: The Paris Criteria (Table 3). We have focused our recommendations on studies designed to augment current prediction models or tailor therapeutics.

Our classification system rates studies based on tiers of depth and quality. Of course, the ability of studies to achieve the highest tier may be limited in analyses of novel bio-markers or new indications. In cases such as these, novelty of the approach should be valued, but the authors should be requested to provide concrete future steps. To avoid false-positive leads, negative findings should be evaluated in the same manner as positive findings and receive equal weight when considered for publication.

For diagnostic and prognostic biomarkers, third-tier studies should be expected to adhere to minimum statistical requirements (Fig. 2). They should all follow the guidelines laid out by the STARD (Standards for Reporting of Diagnostic Accuracy) statement (45,46). The objective of the STARD initiative is to improve the quality of reporting of studies involving diagnostic testing across the spectrum of medical research, allowing readers to detect the potential for bias in the study and to judge the generalizability and applicability of the results. Furthermore, the study should include clinically impactful previously established variables into prediction models (including results from high-quality natriuretic peptide assays), and testing should be performed in a representative cohort. The addition of validation analyses, preferably in an external dataset, would elevate the analysis to the second tier. Highest-tiered studies would include multiple novel biomarkers in unbiased comparisons, rather than the current piecemeal approach of 1 to 2 novel assays. Additional aspects of value would include interaction testing of subgroups to identify specific areas of value for novel biomarker applications, as well as carefully designed analyses to detect influence of heart failure therapeutics on biomarker-derived risk assessment.

Once a biomarker is being considered for clinical use, lowest-tiered studies should show differential effects of therapy according to treatment in a well-defined randomized, controlled trial of patients with heart failure. Results that are validated in another cohort merit a middle-tier rating. Both tiers of studies may be retrospective. However, we suggest that the highest tier of quality be reserved for prospectively designed randomized, controlled trials of therapeutic assignments based on biomarker concentrations.

Simultaneous Use of Multiple Biomarkers

Testing multiple biomarkers when each has not been appropriately validated adds additional complexity to inferring their usefulness in a clinical setting. On the one hand, a multimarker approach may allow us to integrate various aspects of the disease process such as renal disease, inflammation, and myocardial fibrosis. Such an approach has the potential to improve prognostication and selection or titration of therapeutics. However, there is the risk

that increasing the number of variables to describe heart failure may greatly increase complexity without generating actionable knowledge that improves clinical outcomes. To strike a balance between inclusion of information and minimizing complexity, studies must be carefully designed and use of more advanced statistical approaches might be required.

Thus far, despite the ability to simultaneously ascertain dozens of biomarkers in patients with heart failure, the majority of studies have evaluated only pairs or triplets in a piecemeal approach. Two recent studies represent important initial steps toward using biomarkers reflecting different pathophysiological pathways for clinical decision making. The first, performed in patients with heart failure, combined a group of 7 biomarkers that represented different disease pathways into a multimarker score and found that addition of the multimarker score to the Seattle Heart Failure Model led to a significantly improved AUC and appropriate reclassification of patients (47). The second study, performed in patients without known heart failure, showed that a multimarker score consisting of BNP, hsCRP, growth differentiation factor-15, high-sensitivity troponin I, and sST2 provided incremental information for predicting onset of heart failure (48). In these studies, each biomarker was subjected to thorough evaluation via C-statistic analysis, Cox proportional hazards modeling, and NRI and IDI assessment to verify their additive nature prior to the assembly of the multimarker score. These exhaustive methods are an example of the necessary approaches to ascertain the appropriate role of biomarkers beyond the natriuretic peptides.

Although these are important “proof-of-concept” studies, further investigations are required to understand the best combinations of biomarkers for use in clinical practice and the potential therapeutic implications of this approach. Also, because unbiased biomarker screens continue to reveal an exponentially increasing number of candidates, there is a need for more advanced statistical methodologies, such as the use of artificial neural networks and support vector machines to analyze intricate changes in multiple biomarkers at the same time. An ideal approach would allow for algorithms to recognize clinically relevant patterns in data without prior knowledge of relationships (49,50). However, these data-driven techniques tend to underperform in validation settings; therefore, it is essential that results are appropriately validated.

Conclusions

Over the last decade, significant advances in understanding the pathophysiology of heart failure have provided us with numerous biomarkers that allow for a more complete understanding of the disease process. Whether integration of this information into clinical decision-making algorithms may improve care of patients with heart failure rather than simply increase complexity is unknown. To answer these questions, there is a need for a systematic and collaborative approach aimed at definitively studying the clinical benefits of novel biomarkers.

Despite numerous potential candidates, few novel bio-markers of heart failure have successfully made the transition from bench to routine clinical practice; to accelerate this process will require assiduously designed studies with a clear focus on quality. We propose a systematized approach based on adherence to specific quality measures, both for studies

looking to augment current diagnostic and prediction models and for studies using biomarkers to tailor therapeutics. We suggest that study quality, rather than results, should determine publication because this would minimize false publication bias (51). There is an unmet need for clinical investigators and statisticians to collaborate, with the goal of introducing more advanced statistical methodologies into the field of biomarkers that would allow for data from a large number of variables to be distilled into clinically actionable information. Lastly, we propose the creation of a heart failure biomarker consortium that would allow for a comprehensive list of bio-markers to be concomitantly analyzed in a pooled sample of randomized clinical trials and hypotheses to be generated for testing in biomarker-guided trials. Such a consortium could collaborate in sharing samples to identify biomarkers, undertaking meta-analyses on completed trials, and undertaking new clinical trials to test the clinical utility of new biomarkers.

Acknowledgments

Drs. Fiuzat and O'Connor have received research funding from BG Medicine, Critical Diagnostics, and Roche Diagnostics. Dr. Pencina served as a data safety and monitoring board member for DC Devices and the Cardiovascular Clinical Science Foundation; and served as a consultant for BioVentrix. Dr. Zannad has received research funding from BG Medicine and Roche Diagnostics; and served as a consultant for BG Medicine. Dr. Cleland has received research funding from Roche Diagnostics. Dr. Snider is an employee of Critical Diagnostics. Dr. Kim is an employee of Amgen. Dr. Mentz has received honoraria from Bristol-Myers Squibb and Novartis. Dr. Felker has received research funding from BG Medicine, Critical Diagnostics, and Roche Diagnostics; and served as a consultant for BG Medicine and Roche Diagnostics. Dr. Januzzi has served as an advisor or consultant for Critical Diagnostics; and received research funding from Critical Diagnostics, Roche Diagnostics, BG Medicine, Thermo Fisher, Singulex, and Siemens AG. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose. John R. Teerlink, MD, served as Guest Editor for this article.

Abbreviations and acronyms

AUC	area under the receiver-operating characteristic curve
BNP	B-type natriuretic peptide
hsCRP	high-sensitivity C-reactive protein
IDI	integrated discrimination improvement
NRI	net reclassification improvement
ROC	receiver-operating characteristic

REFERENCES

1. Calaprice, A. *The Ultimate Quotable Einstein*. Princeton University Press; Princeton, NJ: 2010.
2. Schocken DD, Benjamin EJ, Fonarow GC, et al. Prevention of heart failure: a scientific statement from the American Heart Association Councils on Epidemiology and Prevention, Clinical Cardiology, Cardiovascular Nursing, and High Blood Pressure Research; Quality of Care and Outcomes Research Interdisciplinary Working Group; and Functional Genomics and Translational Biology Interdisciplinary Working Group. *Circulation*. 2008; 117:2544–65. [PubMed: 18391114]
3. Lee DS, Tu JV, Juurlink DN, et al. Risk-treatment mismatch in the pharmacotherapy of heart failure. *JAMA*. 2005; 294:1240–7. [PubMed: 16160132]
4. Ahmad T, Fiuzat M, Felker GM, O'Connor C. Novel biomarkers in chronic heart failure. *Nat Rev Cardiol*. 2012; 9:347–59. [PubMed: 22450126]

5. van Kimmenade RR, Januzzi JL Jr. Emerging biomarkers in heart failure. *Clin Chem*. 2012; 58:127–38. [PubMed: 22086968]
6. Braunwald E. Biomarkers in heart failure. *N Engl J Med*. 2008; 358:2148–59. [PubMed: 18480207]
7. Maisel AS, Daniels LB. Breathing Not Properly 10 years later: what we have learned and what we still need to learn. *J Am Coll Cardiol*. 2012; 60:277–82. [PubMed: 22813603]
8. Januzzi JL, Troughton R. Are serial BNP measurements useful in heart failure management? Serial natriuretic peptide measurements are useful in heart failure management. *Circulation*. 2013; 127:500–7. [PubMed: 23357662]
9. Desai AS. Are serial BNP measurements useful in heart failure management? Serial natriuretic peptide measurements are not useful in heart failure management: the art of medicine remains long. *Circulation*. 2013; 127:509–16. [PubMed: 23357663]
10. Savarese G, Trimarco B, Dellegrataglie S, et al. Natriuretic peptide-guided therapy in chronic heart failure: a meta-analysis of 2,686 patients in 12 randomized trials. *PloS One*. 2013; 8:e58287. [PubMed: 23472172]
11. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009; 119:2408–16. [PubMed: 19364974]
12. Maisel AS. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. *N Engl J Med*. 2002; 347:161–7. [PubMed: 12124404]
13. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2013; 62:e147–239. [PubMed: 23747642]
14. Januzzi JL Jr, Camargo CA, Anwaruddin S, et al. The N-terminal Pro-BNP Investigation of Dyspnea in the Emergency Department (PRIDE) study. *Am J Cardiol*. 2005; 95:948–54. [PubMed: 15820160]
15. Allen LA. Use of multiple biomarkers in heart failure. *Curr Cardiol Rep*. 2010; 12:230–6. [PubMed: 20424966]
16. Vasan RS. Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation*. 2006; 113:2335–62. [PubMed: 16702488]
17. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006; 113:1424–33. [PubMed: 16534009]
18. Morrow DA, Cook NR. Determining decision limits for new bio-markers: clinical and statistical considerations. *Clin Chem*. 2011; 57:1–3. [PubMed: 21078839]
19. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards for Reporting of Diagnostic Accuracy*. *Clin Chem*. 2003; 49:1–6. [PubMed: 12507953]
20. Cook NR. Assessing the incremental role of novel and emerging risk factors. *Curr Cardiovasc Risk Rep*. 2010; 4:112–9. [PubMed: 20640227]
21. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med*. 2009; 150:795–802. [PubMed: 19487714]
22. Steyerberg EW, Pencina MJ. Reclassification calculations for persons with incomplete follow-up. *Ann Intern Med*. 2010; 152:195–6. [PubMed: 20124243]
23. Spevack DM, Chandra R. Changes in cardiac geometry and function after gastric bypass surgery. *J Am Coll Cardiol*. 2011; 58:1924. [PubMed: 22018304]
24. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004; 23:2109–23. [PubMed: 15211606]
25. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27:157–72. [PubMed: 17569110]
26. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol*. 2012; 176:473–81. [PubMed: 22875755]

27. Pencina MJ, D'Agostino RB, Vasan RS. Statistical methods for assessment of added usefulness of new biomarkers. *Clin Chem Lab Med.* 2010; 48:1703–11. [PubMed: 20716010]
28. Fine J, Gray R. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999; 94:496–509.
29. Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW. Extending the c-statistic to nominal polytomous outcomes: the polytomous discrimination index. *Stat Med.* 2012; 31:2610–26. [PubMed: 22733650]
30. Seymour CW, Cooke CR, Wang Z, et al. Improving risk classification of critical illness with biomarkers: a simulation study. *J Crit Care.* 2013; 28:541–8. [PubMed: 23566734]
31. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006; 26:565–74. [PubMed: 17099194]
32. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011; 30:11–21. [PubMed: 21204120]
33. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med.* 2008; 359:2195–207. [PubMed: 18997196]
34. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst.* 2010; 102:152–60. [PubMed: 20075367]
35. Kjekshus J, Apetrei E, Barrios V, et al. Rosuvastatin in older patients with systolic heart failure. *N Engl J Med.* 2007; 357:2248–61. [PubMed: 17984166]
36. Gullestad L, Ueland T, Kjekshus J, et al. Galectin-3 predicts response to statin therapy in the Controlled Rosuvastatin Multinational Trial in Heart Failure (CORONA). *Eur Heart J.* 2012; 33:2290–6. [PubMed: 22513778]
37. Wollert KC. Tailored therapy for heart failure: the role of biomarkers. *Eur Heart J.* 2012; 33:2246–8. [PubMed: 22611133]
38. Anand IS, Rector TS, Kuskowski M, Adourian A, Muntendam P, Cohn JN. Baseline and serial measurements of galectin-3 in patients with heart failure: relationship to prognosis and effect of treatment with valsartan in the Val-HeFT. *Eur J Heart Fail.* 2013; 15:511–8. [PubMed: 23291728]
39. Price MJ, Berger PB, Teirstein PS, et al. Standard- vs high-dose clopidogrel based on platelet function testing after percutaneous coronary intervention: the GRAVITAS randomized trial. *JAMA.* 2011; 305:1097–105. [PubMed: 21406646]
40. Januzzi JL Jr, Rehman SU, Mohammed AA, et al. Use of amino-terminal pro-B-type natriuretic peptide to guide outpatient therapy of patients with chronic left ventricular systolic dysfunction. *J Am Coll Cardiol.* 2011; 58:1881–9. [PubMed: 22018299]
41. Felker GM, Hasselblad V, Hernandez AF, O'Connor CM. Biomarker-guided therapy in chronic heart failure: a meta-analysis of randomized controlled trials. *Am Heart J.* 2009; 158:422–30. [PubMed: 19699866]
42. Weiner RB, Baggish AL, Chen-Tournoux A, et al. Improvement in structural and functional echocardiographic parameters during chronic heart failure therapy guided by natriuretic peptides: mechanistic insights from the ProBNP Outpatient Tailored Chronic Heart Failure (PROTECT) study. *Eur J Heart Fail.* 2013; 15:342–51. [PubMed: 23132825]
43. Bhardwaj A, Rehman SU, Mohammed AA, et al. Quality of life and chronic heart failure therapy guided by natriuretic peptides: results from the ProBNP Outpatient Tailored Chronic Heart Failure Therapy (PROTECT) study. *Am Heart J.* 2012; 164:793–799. e1. [PubMed: 23137512]
44. Ledwidge M, Gallagher J, Conlon C, et al. Natriuretic peptide-based screening and collaborative care for heart failure: the STOP-HF randomized trial. *JAMA.* 2013; 310:66–74. [PubMed: 23821090]
45. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med.* 2003; 138:40–4. [PubMed: 12513043]
46. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med.* 2003; 138:W1–12. [PubMed: 12513067]

47. Ky B, French B, Levy WC, et al. Multiple biomarkers for risk prediction in chronic heart failure. *Circ Heart Fail.* 2012; 5:183–90. [PubMed: 22361079]
48. Wang TJ, Wollert KC, Larson MG, et al. Prognostic utility of novel biomarkers of cardiovascular stress: the Framingham Heart Study. *Circulation.* 2012; 126:1596–604. [PubMed: 22907935]
49. Berriz GF, King OD, Bryant B, Sander C, Roth FP. Characterizing gene sets with FuncAssociate. *Bioinformatics.* 2003; 19:2502–4. [PubMed: 14668247]
50. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011; 12:56–68. [PubMed: 21164525]
51. Demaria AN. Publication bias and journals as policemen. *J Am Coll Cardiol.* 2004; 44:1707–8. [PubMed: 15489107]

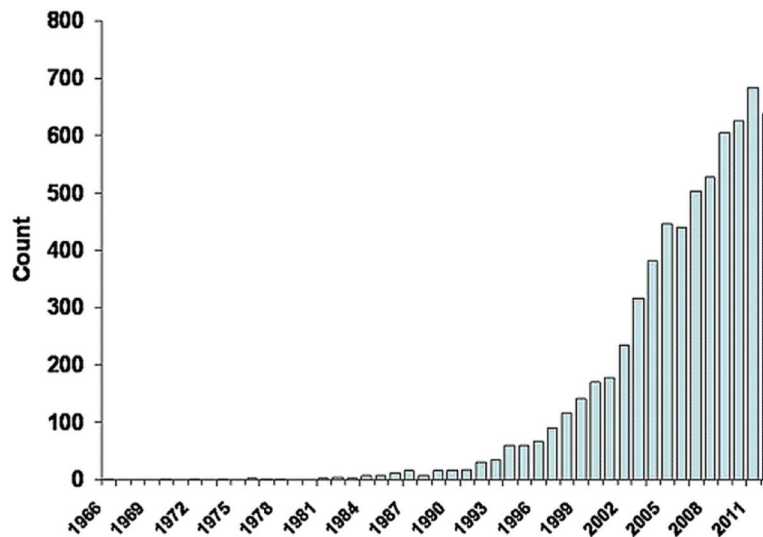


Figure 1. Publications Involving Heart Failure Biomarkers

The number of publications involving heart failure biomarkers has dramatically increased over the last decade. This has likely resulted from high-throughput molecular biology techniques that allow increased availability of rapid-turnaround biomarker testing and reductions in costs of analysis. The majority of these studies report associations between novel biomarkers and prognosis from heart failure.

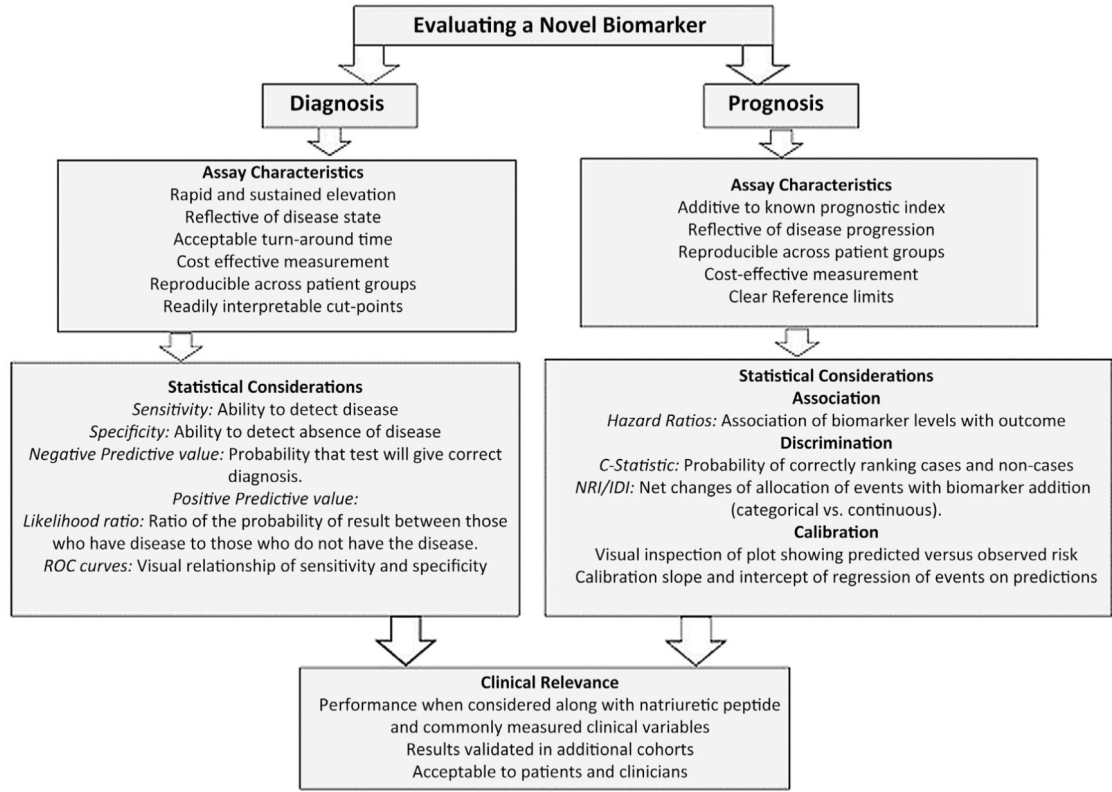


Figure 2. Approaches to Evaluation of Novel Biomarkers in Heart Failure

The methodology depends on proposed use of biomarkers, especially whether they are for diagnosis or prognostication. IDI = integrated discrimination improvement; NRI = net reclassification improvement; ROC = receiver-operating characteristic.

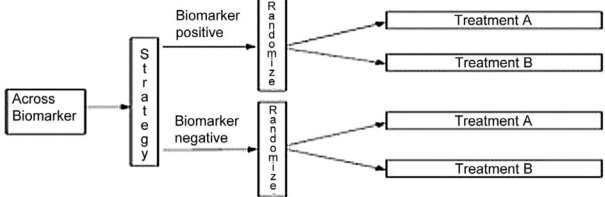
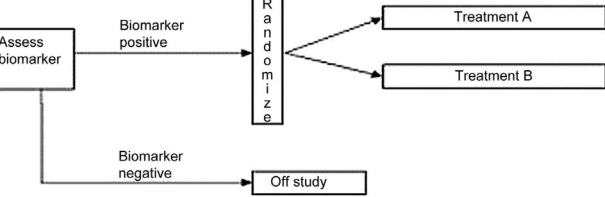
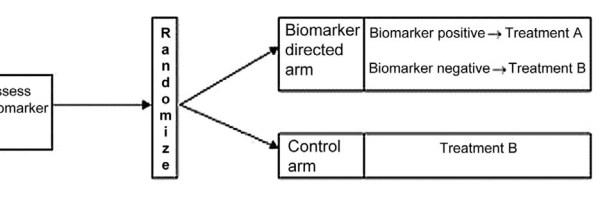
BIOMARKER STUDY DESIGNS	PROS/CONS	EXAMPLE STUDIES
<p style="text-align: center;">BIOMARKER STRATIFIED DESIGN</p> 	<p>PROS</p> <ul style="list-style-type: none"> • Modest <i>a priori</i> data • All patients randomized to same therapy • Results analyzed according to biomarker values <p>CONS</p> <ul style="list-style-type: none"> • Prospective: results not definitive • Difficult when multiple treatment options 	<ul style="list-style-type: none"> • CORONA • HF-ACTION • ValHEFT • WGHS • PHFS • CARE-HF • COACH • PRIDE
<p style="text-align: center;">ENRICHMENT DESIGN</p> 	<p>PROS</p> <ul style="list-style-type: none"> • Addresses question about best therapy for biomarker positive group <p>CONS</p> <ul style="list-style-type: none"> • Requires strong <i>a priori</i> data • Therapy efficacy in biomarker negative cohort remains unknown 	<ul style="list-style-type: none"> • GRAVITAS • JUPITER • ASTRONAUT • PARADIGM-HF
<p style="text-align: center;">BIOMARKER STRATEGY DESIGN</p> 	<p>PROS</p> <ul style="list-style-type: none"> • Most definitive study design • Can be used with multiple therapies <p>CONS</p> <ul style="list-style-type: none"> • Inefficient especially when considering a number of biomarkers • Patients and physicians are unblinded to biomarker levels 	<ul style="list-style-type: none"> • PROTECT • PONTIAC • STOP-HF • GUIDE-IT [on-going]

Figure 3. Strategies for Biomarker-Guided Trials in Heart Failure
Data from Freidlin et al. (34).

Table 1**Evaluation of Heart Failure Biomarkers****Method**

The method by which a novel biomarker is judged should be thorough: novel tests should be evaluated across a wide range of patients typical of the diagnosis for which the biomarker will be applied, and the statistical methods used to evaluate the biomarker should be contemporary, rigorous, standardized, and fair.

Measurement

Measurement of a novel heart failure biomarker should be easily achieved within a short period of time and provide acceptable accuracy. Assays for its measurement should have defined biological variation and low analytical imprecision.

Mechanism

The biomarker should primarily reflect important pathophysiological processes involved in heart failure presence and progression. Use of a biomarker that is reflective of heart disease but originates outside the myocardium is acceptable as long as such a biomarker provides independently useful information involved in the diagnosis, prognosis, progression, or therapy of heart failure syndromes.

Clinical Information

The biomarker must provide clinically useful information for caregivers and patients to facilitate more swift and reliable diagnosis accurate estimation of prognosis and to inform more successful therapeutic strategies. The information from such a biomarker should not recapitulate clinical information already available at the bedside and must be additional to information provided by established biomarkers.

Table 2

Statistical Approaches to Evaluation of Novel Biomarkers in Heart Failure

Property	Metric	Strengths	Weaknesses
Discrimination	<ul style="list-style-type: none"> • Area under the curve • Survival C-statistic • Discrimination slope 	<ul style="list-style-type: none"> • Familiar • Allows for comparison of different variables and additive value of variables beyond a base model through statistical change in discrimination • Based on absolute risks • Intuitive measure of separation 	<ul style="list-style-type: none"> • Ignores the potentially random nature of prognosis • May be influenced by case mix of the cohort studied and does not take distribution of risk into account • May underestimate the importance of variables when added to robust base model • Dependent on model calibration
Calibration	<ul style="list-style-type: none"> • Visual inspection of prediction and observation risk • Slope and intercept of model using predicted risk as independent variable and true outcome status as the dependent variable 	<ul style="list-style-type: none"> • Allows for assessment of areas of poor calibration • Measures amount of average bias of prediction and degree of overdispersion 	<ul style="list-style-type: none"> • Interpretation might be arbitrary • May be difficult to understand
Reclassification	<ul style="list-style-type: none"> • NRI • IDI • Increment in net benefit • Weighted NRI 	<ul style="list-style-type: none"> • Provides clinical context • NRI is particularly useful when predefined categories of risk exist • Weighted NRI and increment in net benefit provide measure of consequences 	<ul style="list-style-type: none"> • When used with categories, NRI result may be influenced by the arbitrary choice of risk thresholds • Choice of costs and weights might be controversial

IDI = integrated discrimination improvement; NRI = net reclassification improvement.

Table 3

Proposed Tier System for Quality of Biomarker Studies: The Paris Criteria

Improved Diagnosis and Prognostication	Tailoring Therapeutics
<p>Tier 1</p> <ul style="list-style-type: none"> • Satisfies criteria for tiers 2 and 3 • Interaction testing of subgroups • Influence of therapeutics on predictive abilities 	<p>Tier 1</p> <ul style="list-style-type: none"> • Satisfies criteria for tiers 2 and 3 • Randomized, controlled trial in which biomarker levels determine therapeutic choices
<p>Tier 2</p> <ul style="list-style-type: none"> • Satisfies criteria for tier 3 • Compares multiple biomarkers concomitantly • Validates results in a representative cohort 	<p>Tier 2</p> <ul style="list-style-type: none"> • Satisfies criteria for tier 3 • Validates results in a representative cohort
<p>Tier 3</p> <ul style="list-style-type: none"> • Follows STARD statement • Pathophysiological link to heart failure • Confirms association with well-defined outcomes in a representative cohort of patients with heart failure after controlling for clinical factors and natriuretic peptides • Reports AUC with and without addition of biomarker and shows that there is statistically significant improvement with the addition of the biomarker • Demonstrates that addition of the new biomarker improves the classification of those who had the biomarker event 	<p>Tier 3</p> <ul style="list-style-type: none"> • Shows differential effect of treatment based on biomarker levels in a retrospective analysis of randomized, controlled trials

AUC = area under the receiver-operating characteristic curve; STARD = Standards for Reporting of Diagnostic Accuracy.