Perspective

For reprint orders, please contact: reprints@future-science.com

Biomarkers in early-phase trials: fundamental issues

Laura M Yee*,¹, Tracy G Lively¹ & Lisa M McShane¹

¹Division of Cancer Treatment & Diagnosis, National Cancer Institute, National Institutes of Health, Rockville, MD, 20850, USA *Author for correspondence: laura.yee@nih.gov

Biomarkers are frequently being included in early-phase clinical trials. This article is meant to introduce clinical investigators to the fundamentals of choosing a biomarker test for use in an early-phase trial. Steps to consider are briefly outlined including defining the role of the biomarker in the early-phase trial; selecting a fit-for-purpose biomarker test and laboratory; describing the test procedures; carrying out analytical validation testing appropriate for the research objectives and the risk involved in the trial; implementing the test in the trial; and planning for the future. Examples illustrate analytical validation approaches in the context of typical biomarker roles. The importance of collaboration between clinical investigators and laboratory researchers is emphasized.

First draft submitted: 5 January 2018; Accepted for publication: 19 February 2018; Published online: 20 June 2018

Keywords: analytical validation • biomarkers • clinical trials • drug development • early-phase trials

Biomarkers are playing increasingly critical roles in the development of new drugs and are therefore being incorporated earlier in the drug development pipeline – such as in early-phase trials. The FDA-NIH Biomarker Working Group describes a biomarker to be "a characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention" [1]. Biomarkers that are used in late phase trials are typically better characterized in analytical validation studies to aid in more rapid deployment into clinical settings, such as via an US FDA approved or cleared test. Biomarker tests differ from biomarkers in that a biomarker test is the entire set of materials, assay and method for measuring a biomarker [1]. Although it is acknowledged that it may not be feasible to have a fully validated biomarker test in an early phase setting – for example, Phase 0, I and early Phase II trials – there are still several factors to consider when deciding whether a biomarker test is ready for use within a trial. Many aspects of early-phase trials are designed on the basis of limited preliminary information obtained from preclinical testing. Planning the use of biomarker tests in these trials calls for the same thoughtful, step-wise approach as used to determine when a new therapeutic is ready for clinical evaluation.

Although clinical researchers design and lead academic clinical trials, laboratory researchers are the experts in biomarker test development and implementation. Sometimes there is a lack of communication between study clinicians and laboratory researchers leading to biomarker tests that may be poorly suited for use in the early-phase trial in which they are being proposed. There have been many other efforts to describe and improve the process of getting biomarkers ready for use in a clinical setting [1–7]. This document is not meant to provide an in-depth survey of such efforts. Instead, the goal of this document is to introduce academic clinicians to key considerations for incorporating biomarkers into early phase clinical trials, including discussion of biomarker test analytical validation in the early phase setting. Preparing a biomarker test for routine clinical use or seeking a broad qualification would require a more comprehensive and rigorous process than what is described here for use of a biomarker in very early stages of drug development. The considerations described here are motivated by experiences with use of biomarkers in early phase clinical trials supported by the US National Cancer Institute (NCI).

Overview of incorporating a biomarker into an early-phase trial

There are several important steps to prepare for use of a biomarker in an early phase clinical trial (Figure 1). These steps are described in further detail in the following sections.

newlands press





Step 1

The first step when incorporating a biomarker into a clinical trial is determining what role it will play (Figure 1). A variety of uses for biomarkers in clinical trials have been proposed [1,5,8–13]. The NCI makes a further broad distinction pertaining to whether a biomarker is integral or integrated. Integral biomarkers are those that are essential to conducting the study and must be performed in real time. Integrated biomarkers are those that are associated with a high priority scientific question and for which there is a clear statistically testable hypothesis, there are prespecified protocols for specimen collection, biomarker testing and statistical analysis, but the trial could proceed without the biomarker. Exploratory biomarkers are all others – neither integral nor integrated [14–16]. In general, integral biomarkers will require the most supporting data, while exploratory may require the least. The following sections will focus primarily on integral biomarkers in early-phase trials, as they require the most scrutiny. Three types of general uses of early phase biomarkers that will be the focus of discussion in the next section are biomarkers that are used for study eligibility, as primary end points and as stratification factors.

Further distinction of the biomarker can help describe its intended use. A diagnostic biomarker is used to determine the presence of a condition, disease or disease subtype [1]. A prognostic biomarker is used to indicate a subset of patients for whom a clinical event is more (or less) likely [1]. In addition to diagnostic and prognostic biomarkers, early-phase trials also frequently employ biomarkers as outcomes in target validation, early compound screening, pharmacodynamic studies and as surrogate end points [8].

Carefully defining a biomarker's use in the trial is sometimes overlooked or poorly communicated to laboratory and statistical collaborators, which has implications later in trial development. Therefore, this is a juncture at which the clinician, the laboratory researcher and the statistician must work together.

Step 2

Once the biomarker's role in the trial is identified, it is critical to select a biomarker test that is fit for this role and a laboratory that is equipped to perform it [3]. A test that is fit for one purpose may not be adequate for another. For example, an immunohistochemistry test that can provide a binary distinction between biomarker positive and negative cases to determine eligibility for a trial may not have the precision or sensitivity required to measure small quantitative differences in biomarker expression. A biomarker test that has been developed on one type of sample may need significant adaptation to perform successfully on another.

This step requires clear and extensive communication with the laboratory so that expectations are understood and requirements can be met. Will a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory be required, so that results can be returned to patients [17]? Alternatively, does the trial require reliable quantitative assessment of a pharmacodynamic end point by means of a test that has already undergone extensive validation in a suitably equipped research laboratory? Investigators without laboratory experience sometimes underestimate the length of time and the resources required to bring a biomarker test to readiness.

Step 3

The next step is to carefully define the biomarker test and the laboratory operating procedures for performing it (Figure 1). The clinical investigator's main role in this step may be to choose a laboratory with previous experience running the test and to include test specifics in the study protocol. For pharmacokinetic and pharmacodynamic assays, it is advisable to seek out laboratories that are already equipped and experienced with the desired type of assay. Other uses of biomarkers such as for patient eligibility determination or stratification might make use of less mature assays because less is known about the relevance of the biomarker for the investigational drug. While this step primarily involves input from laboratory collaborators, it is important that the primary clinical investigator understand basic considerations of describing sample collection and test procedures, as they may have significant impact on performance of the test [18].

There are many important factors that may affect how the test performs. Therefore, it is best to fully describe the test, including aspects such as the target analyte, specimen type, pre-analytic processing, platform, test components, positive and negative controls, where the test will be performed, and the scoring procedure/cut-points of the test [18–20]. The scoring procedure may be quantitative (e.g., continuous biomarker value proportional to a reference standard), semiquantitative (e.g., an underlying continuous measurement that is not proportional to a reference standard to which a cut-off is applied to report results in two or more categories such as low/intermediate/high values), or qualitative (e.g., binary – positive/negative) [3,4]. For tests such as omics tests that involve mathematical models or algorithms for prediction or classification, additional aspects must be described to fully define the test, including methods of data normalization, bioinformatic processing and the model or algorithm required to combine various omics features to obtain the test result [21,22].

At a minimum, the clinical protocol must provide instructions for the trial sites for the collection, handling and processing of the specimens for the assay, so that the pre-analytic requirements are met by the time the materials arrive at the laboratory [23]. The protocol and its informed consent must also describe the purpose of the assay in the trial and its limitations, so that Institutional Review Boards and participants can understand any risks posed by use of the test [23,24]. Risks may be interpreted broadly to include risk inherent in collection of the specimen needed to run the test as well as risk if the test result is incorrect. It is also advisable to include an appendix or separate laboratory manual that describes the assay in more detail. An integral biomarker assay is just as important a component of the trial as the investigational treatment [25]. As a guiding principle, investigators should ask: "If the current trial is successful, could a second qualified laboratory perform the same test and find the same group of patients or provide results comparable to those obtained in this trial?"

Step 4

Analytical validation refers to the process of establishing the performance of the biomarker test [1]. For early phase clinical trials, biomarker test analytical validation should be performed to justify that the test can obtain a result with sufficient accuracy and reliability to reasonably mitigate risks involved in use of the test for the intended purpose that was defined in Step 1 (Figure 1).

Table 1 defines several important analytical validation performance characteristics, along with the information that should be provided from validation studies to support each one [26]. It is important to emphasize that analytic validation studies should be performed on test samples that resemble as closely as possible the specimen types that will be obtained from the trial and should cover the entire range of values to be expected in the early-phase trial. Although there is no substitute for performing test runs on samples as similar as possible to those that will be obtained on the trial, in the setting of a rare cancer, it may be difficult to obtain such samples. In this case, it may be appropriate to consider what other approaches might be used to show that the test is adequate for use in the trial,

Table 1. Early phase biomarker test analytical validation definitions and recommendations.		
Validation factor	Definition	Supporting information
Accuracy	"The closeness of agreement between the test results obtained using the new biomarker test and results obtained using a reference standard method widely accepted as producing 'truth' for the analyte The observed level of agreement will depend on both the bias and precision of the new test" [26]	 Fully describe the reference standard. When no reference standard exists, describe the nonreference standard employed Describe the type and source of the samples (e.g., patient samples vs contrived, number of samples that are positive or negative for the clinical condition of interest) Describe summary measures for the study (scatter plots, average bias, mean squared deviation for continuous data; 2 × 2 tables, overall percent agreement, sensitivity and specificity for binary data) and maintain raw data Measurements of precision are discussed below [26]
Bias	"The amount by which an average of many repeated measurements made using the new test systematically over- or underestimates the reference standard method result" [26]	
Repeatability	Precision of the biomarker test under essentially unchanged conditions ('within-series precision' or 'within-run precision') [26]	 Describe the study design (number of replicates, number of samples, number of factors considered and number of that factor included, range of assay values studied) Describe the type and source of the samples (e.g., patient samples vs contrived, number of samples that are positive or negative for the clinical condition of interest) Describe summary measures for each sample (mean, SD and CV% for continuous data; overall, positive and negative percent agreement for binary data) and maintain raw data. Test results should be summarized according to the output that will be used in the trial. Binary tests with underlying quantitative or semiquantitative data (e.g., imaging tests) should also provide quantitative or semiquantitative analyses when possible [26]
Intermediate precision	Precision of the biomarker test when there is "variation in one or more factors, such as time, calibration, operator and equipment – usually within a laboratory" [26]	
Reproducibility	Precision of the biomarker test between laboratories that also "relates to changes in conditions such as different operators and measuring systems (including different calibrations and reagent batches)" [26]	
LoD	"The smallest amount of analyte that an analytical method can detect with a specified probability" [26]	- Further discussion of LoD can be found elsewhere [27,28]
LoQ	"The smallest amount of an analyte in a sample that can be quantitatively determined with acceptable precision, and trueness as measured by bias" [26]	- Further discussion of LoQ can be found elsewhere [27,28]
Linearity	"The ability to provide measured quantity values that are directly proportional to the value of the measurand in the experimental unit" [28]	- Further discussion of linearity can be found elsewhere [28,29]
LoD: Limit of detection; LoQ: Limit of quantitation.		

such as testing in contrived samples followed by a feasibility biomarker study in a subset of the samples accrued early in the trial.

To give better intuition as to what analytical performance characteristics are most critical for a given biomarker purpose, additional discussion and examples are provided in Section 3.

Step 5

Once the test has been deemed adequate for the purpose within the early-phase trial, it can be implemented in the trial according to the test procedures defined in Steps 2 and 3 (Figure 1). During the trial, it may be advantageous to collect and retain specimens for future retrospective analyses. For example, these specimens may later be used retrospectively to assess hypotheses with new biomarker tests or to bridge an improved version of the biomarker test to the previous one [30,31].

Step 6

Next, it is advantageous to consider how the biomarker may be used in the future to be prepared should the biomarker test advance to use in a new trial or clinical setting (Figure 1). After trial completion, the study team should convene to assess whether the trial confirms hypotheses related to the biomarker, and whether a future trial incorporating the biomarker may be warranted. In early-phase trials, it is useful to be clear as to whether a biomarker is being used primarily to answer questions about the drug (Is the target engaged? Are the pharmacodynamic effects as expected?) versus clinical questions about the patient (Is this person's type of tumor more likely to respond?). These two kinds of questions call for different laboratory approaches and have very different implications for the future of the biomarker test. In early-phase trials, it is often possible to obtain specimens under specialized conditions meeting the stringent requirements of pharmacodynamic tests for quantitative changes in labile analytes. Tests planned for use in large, later phase trials or for development as FDA-approved *in vitro* diagnostic devices must perform adequately on materials collected during routine clinical care and often with the assay performed in multiple laboratories. Although investigators often look to the future in terms of drug/therapy development,

frequently insufficient attention is devoted to biomarkers, leading to inadequately developed biomarker tests in late phase trials or delays in drug development while waiting for analytical validation of an integral biomarker.

Therefore, if the same biomarker may be used in a later phase trial or in a more important role (e.g., moving the biomarker from integrated to integral), it may be of interest to continue to develop the biomarker test. For a later phase trial, the test may need to undergo more extensive analytical validation to ensure that it will perform well in a multicenter setting or to meet regulatory requirements (e.g., FDA and CLIA). Any changes to the test to improve its analytical performance might require that new analytical validation studies be performed. Similarly, in a later phase trial where the biomarker test might be returned to the physician and/or patient, the test will have to be run in a CLIA-certified laboratory. However, because each laboratory will have its own test specifics and operating procedures, all analytical validation might need to be redone.

Examples: deciding whether biomarker analytical results are acceptable for the proposed trial

The scope of analytical validation required will depend on the purpose that the biomarker serves in the trial. The following examples illustrate matching the scope of analytical validation studies to the biomarker's role in the early-phase trial (Step 4) and how to avoid choosing an inappropriate biomarker test. A common thread through all the examples is the need to acquire preliminary data on the performance of the test as it will be used in the trial to verify that it will be fit-for-purpose.

Study eligibility

A biomarker test that is used to determine study eligibility delivers a binary result, although it can be based on underlying semiquantitative or quantitative measurement with a cut-point to determine study eligibility. Studies of accuracy and precision are particularly of interest when considering whether the biomarker test's analytical performance is sufficiently good to mitigate the risk in determination of patient eligibility for the trial.

Evaluation of the accuracy of the biomarker test is important both for patient safety and for demonstration of efficacy of the agent under study. In terms of patient safety, a false positive test result will expose patients to unnecessary and ineffective treatment, while a false negative result will withhold potentially effective treatment. Depending on the clinical risks of the study agent, false positive and false negative tests might not have equal importance. For example, a false positive test result might be concerning when the treatment has high toxicity and there is evidence that the drug will not work in the biomarker negative population. In some early-phase trials, a false negative test result may pose less risk when the only consequences are that it unnecessarily screens the patient away from the investigational treatment.

An inaccurate biomarker test used for eligibility will also impede the ability of a trial to show efficacy of the drug, compared with a highly accurate test. Consider an example of a binary test where the presence of a specific mutation is required for eligibility in a Phase II-targeted therapy cancer trial because there is evidence that the targeted therapy only works for patients with that mutation. If the test yields 25% false positive results, then 25% of patients eligible for the trial are undergoing potentially toxic treatment that will not work for them, which is a patient safety concern. In addition, suppose this agent yields a response rate of 35% in patients who truly have the mutation of interest, but no patients whose tumors do not carry the mutation respond. With this inaccurate test, the response rate will be diluted by 25% due to the presence of patients with false positive mutations – yielding an observed response rate of only 26% (=100% × 0.35 × [1–0.25]), and therefore reducing the chance that the trial will be positive.

Analytical validation studies should be carefully designed to consider a case mix that mimics the range of biomarker values that is expected in the clinical trial. For a binary biomarker test, agreement studies should report results separately for true positive and true negative samples. To illustrate these points, consider a binary test for study eligibility that is based on an underlying continuous measurement. This test can report within the range of 0-10, but patients with the disease in question are entered into the study only if they have a biomarker test result >5. An intermediate precision study was completed on ten contrived (nonpatient) samples. Each sample was split into two aliquots with the two aliquots run on different days by different operators. Investigators concluded that the test was adequately precise based on a reported 90% concordance (same side of cut-point 5) between measurements obtained on different days. Results of the experiment are presented in Figure 2. For each sample, the average biomarker test score (averaged between the two operators) is depicted by a closed circle if the two operators agreed and by an 'X' if the two operators disagreed, where agreement is based on both operators observing results on the same side of the cut-point 5. When defining sample positivity/negativity based on this average value, we see



Figure 2. Results from an intermediate precision study involving two operators evaluating replicate samples from each of ten contrived (nonpatient) specimens. For each specimen, the average biomarker test score (averaged between the two operators) obtained on the replicate samples is depicted by a closed circle if the two operators agreed and by an 'X' if the two operators disagreed, where agreement is based on both operators observing results on the same side of the cut-point 5. A separate study looked at the range of the biomarker in 11 patients with the disease of interest, which is depicted by the histogram in gray.

that 100% (8/8) of negative samples were concordant, while only 50% (1/2) positive samples were concordant. The 50% concordance in positive samples was masked when reporting the overall concordance (90%). Because negative and positive test errors may have different risks depending on the use of the biomarker in the trial, it is important to report concordance in negative and positive samples separately [32]. In this example, it also may be informative to complete another study with more positive samples and more operators to better understand the pattern of discordance in those samples. It would be important to determine whether the 50% concordance in positive samples was an aberration due to the small number of positive samples in the study, or whether the positive concordance truly was poor.

Because this study was completed in contrived samples, a separate study looked at the range of the biomarker in 11 patients with the disease of interest, which is depicted by the histogram in gray (Figure 2). This study found that patients with the disease of interest had biomarker values between 3 and 9. The precision study used samples that were mainly outside this expected range in patient samples, which is problematic because it did not allow examination of test variability at the ranges expected in the early-phase trial. In addition, because most of the samples that were used in the first study were further from the cut-point, they would be easier to call, which potentially led to overly optimistic estimates of day-to-day (operator-to-operator) concordance.

Primary end point

Biomarker tests that are used as primary end points – for example, as surrogate end points, monitoring biomarkers or pharmacodynamic biomarkers [1] – are often continuous markers, and are frequently described by their absolute value or in measurements relative to a baseline value. Depending on the proposed use of the biomarker test in the early-phase trial, it may be important to consider test precision, bias, linearity and limit of detection/quantitation, as described below. Just as the statistical plan for a protocol will include a calculation of the trial's statistical power to detect an effect on the prespecified clinical end point, responsible planning for the use of a biomarker test to evaluate a primary protocol objective requires demonstration that the assay is capable of reliably detecting the biomarker-based outcome of interest. For example, if the outcome of interest is a change in the level of an intracellular protein following treatment, does the biomarker assay have the necessary analytic sensitivity, precision and dynamic range to detect that change if it occurs? It will not be possible to demonstrate this without some preliminary data on the magnitude of the expected change, the baseline distribution of expression levels before treatment and the variability inherent in the biomarker measurement process.

Precision study design and performance goals should be chosen relative to their purpose in the early-phase trial. For example, suppose an early phase study's primary end point is to detect a 20-unit increase in biomarker value between baseline and post-treatment measurements within a given patient. Because the test will be interpreted by different observers, investigators complete an intermediate precision study with three separate observers running the test in its entirety in samples from ten different patients. Each patient specimen was split into three samples for independent analysis by the three observers. Figure 3A & B illustrates hypothetical results of the precision study, where each set of three vertically stacked, horizontally aligned points represents the readings from the three



Figure 3. Interpretation of results from observer-to-observer intermediate precision studies (three observers) depends on the purpose of the biomarker in the trial. Each of ten patient specimens was split into three samples for independent analysis by the three observers. Each set of three vertically stacked, horizontally aligned points represents the readings from the three different observers (y-axis) for a given patient (x-axis represents reference standard value for each patient). Brackets indicate a 20-unit decrease to 20-unit increase in biomarker value. (A) (Left): the biomarker must be adequately precise to call a 20-unit increase in biomarker value; however, it is not possible to distinguish between interobserver variability and a true 20-unit increase. (B) (Right): the biomarker must be adequately precise in biomarker value. Results of the precision study indicate that this biomarker can generally distinguish between interobserver variability and a 20-unit increase.

different observers (y-axis) for a given patient (x-axis represents reference standard value for each patient). Brackets mark a 20-unit decrease to 20-unit increase at each reference standard value. In Figure 3A, interobserver variability is high. In fact, it was common for observers to call biomarker values beyond a 20-unit difference from the reference standard value, indicating it is not possible to distinguish between interobserver variability and a true 20-unit increase with this test. In Figure 3B, interobserver variability is much lower, indicating that it is possible to distinguish a 20-unit increase in biomarker test value. In this study, it is also important to note that the primary end point (20-unit increase in biomarker test value) was chosen based on previous preclinical data. This example illustrates the necessity of understanding the analytical performance of the biomarker assay and how the required performance depends on the role the biomarker will play in the trial.

In addition, analytical precision study reports should carefully describe the study design and levels of replication, and the results must be analyzed consistent with the study design. Levels of replication refer to the points from which the biological sample preparation or technical precision testing occurs. There may be several different levels of replication represented within a single precision study; these generally represent different steps in the sample preparation and assay procedure. For example, Geschwind describes general steps that occur in cDNA microarray experiments, which include, among other factors, steps for sample preparation, RNA extraction, fluorescent labeling, hybridization to the array and image scanning [33]. An intermediate precision study in which only one extraction is performed to create a single RNA batch from which aliquots are assayed on five different days will produce a smaller estimate of variability than a study in which independent RNA extractions are performed on each day. Any time more than one level of replication is present in a precision study, the statistical analysis of the data generated must properly account for the hierarchy of replication levels, for example by use of variance component models that can partition total variability according to contributions of variability from the various levels of replication [34,35]. When reporting precision estimates from such studies, it is also important to clearly describe the sources of variation represented.

The linear range of the assay refers to the portion of the assay reportable range for which linearity has been experimentally demonstrated. Ideally, linearity is established by demonstrating proportionality to values obtained



Figure 4. A linearity study shows observed biomarker test values (y-axis) against the known amount of analyte as calculated by the dilution scheme (x-axis). Squares designate values observed in the linearity study, with a best-fit line (solid line). Based on the observed data, investigators concluded the test had acceptable linearity. A dashed line connects points (denoted by an 'X') that would have been observed if investigators had continued the dilutions below 2. Because biomarker values in the target clinical population often fall in the lower end of the range (below 2), the test would not be adequate for interpreting biomarker change values in the early phase clinical trial.

using a reference standard method for measuring the true amount of analyte in a sample (Table 1). In the absence of an accepted reference standard assay method, it might be possible only to demonstrate proportionality maintained in a dilution series of samples. Linearity is important to consider when proportional changes in the biomarker test are of interest. For example, Figure 4 illustrates a hypothetical small linearity study. Investigators made five dilutions of a sample with a known high amount of biomarker analyte, and reported the observed biomarker test values (y-axis) against the known amount of analyte as calculated by the dilution scheme (x-axis) as represented by the solid line and squares. Because these points formed a linear relationship for the points considered (x-values between 2 and 5), investigators determined the test had acceptable linearity. However, the biomarker test values often fall between 0 and 2 in clinical samples from the population of interest. Dashed lines connect points that would have been observed if investigators had continued the dilutions below 2. As can be seen, a 1-point increase in test score between 2 and 3 or 4 and 5 can be interpreted the same with respect to the reference standard, while a 1-point increase between 1 and 2 cannot be interpreted similarly. In fact, the test may not be able to report any values as low as 1. Because the test is not linear in this important range, the test would not be adequate for this early phase study. Although not discussed in detail here, it is also sometimes important to characterize the limit of detection when it is important to interpret values at the low end of the test [27,28].

Stratification factor

Some early phase studies enroll patients regardless of biomarker value but plan primary analyses to determine the effect of the investigational drug separately within strata defined by values of the biomarker. Assay analytical performance considerations overlap with those for biomarkers used for eligibility determination (because there is a cut-point) and for primary end points (because often there is an underlying continuous biomarker value). As discussed for the other uses of biomarkers, things may go awry if the biomarker has poor analytical performance. Estimates of treatment effects of the investigational drug for different biomarker-defined subgroups may be seriously distorted if the biomarker is poorly measured.

For example, consider a study where a primary goal is to test the difference in response rates between patients with positive versus negative biomarker test values, as defined by a cut-point. Suppose that using the reference standard measurement (assumed 100% accurate), the response rate is 0.1 in the biomarker negative subgroup, and 0.4 in the biomarker positive subgroup, yielding a difference in response rates of 0.3 using the reference standard. The observed difference in response rates are diluted to 0.24, 0.18, 0.12 and 0.06 (= $0.3 \times [1 - False Positive Rate - False Negative Rate]$) when the percentage of false negative and false positive results are equal at 10, 20, 30 and 40%, respectively [7]. Therefore, a study that does not take level of biomarker assay inaccuracy into account may be underpowered for this primary goal due to a diminished apparent effect size, while a study that does take it into account will require significantly larger sample size if there is a large amount of inaccuracy in the chosen test. Although not discussed in detail here, observed effects will also be diluted when an inaccurate test is used in a randomized trial where the goal is to assess the biomarker-by-treatment interaction [36].

Cut-points are often used to determine stratification groups (e.g., positive or negative) when a biomarker test is quantitative or semiquantitative. Ideally, a cut-point for a study should be predefined based on preliminary data, when available. Use of different cut-points may alter conclusions about strength of association between the biomarker and treatment effect, as shown in Figure 5. Figure 5 demonstrates how moving a cut-point can produce



Figure 5. Moving a cut-point can produce different conclusions from observed data. The vertical dashed lines A and B indicate two possible choices of cut-point. Points falling to the right of a dashed line indicate a positive test result, while points falling to the left are negative. With cut-point A, the median, the response rate in the test positive group is 8/15 = 53% while the response rate in the test negative group is 2/15 = 13%, yielding a difference in response rate of 40%. Using cut-point B, the response rate in the test positive group is 7/9 = 78% while the response rate in the test negative group is 3/21 = 14%, yielding a larger difference in response rate of 64%.

different conclusions from observed data. The vertical lines A and B shown in Figure 5 indicate two different cut-point selections; points falling to the right of a dotted line indicate a positive test result, while points falling to the left are negative. With cut-point A, the median, the response rate in the test positive group is 8/15 = 53% while the response rate in the test negative group is 2/15 = 13%, yielding a difference in response rate of 40%. Using cut-point B, the response rate in the test positive group is 7/9 = 78% while the response rate in the test negative group is 3/21 = 14%, yielding a larger difference in response rate of 64%. Because different cut-points can yield vastly different results for any given dataset, it is problematic to choose a cut-point after assessing the observed data, as results of the same magnitude will not be reproducible in future studies [37]. Although a summary of various cut-point selection methods are beyond the scope of the present discussion, there are a number of methods to choose cut-points [38–40]. Furthermore, if preliminary data are not available to justify the use of a specific cut-point, the decision to stratify the analysis may be premature [26].

Molecular signatures that combine information from high-dimensional biomarker data (e.g., genomic, proteomic or metabolomic data) to stratify patients into groups are a type of biomarkers that are often used as stratification factors [21,22]. For integral aims, it is important to utilize a signature that can be applied to a single patient's sample, as methods that rely on processing multiple patient samples at once for normalization purposes cannot be used in a clinical setting as patients are accrued, and their samples are analyzed, one at a time.

Conclusion & future perspective

We presented a discussion of the steps to consider when preparing a biomarker for use in an early phase clinical trial of an investigational therapy. Emphasis was given to tailoring the biomarker test analytical validation requirements to the role that the biomarker will serve. In order to devise a biomarker test analytical validation plan, some minimal information about the biomarker and its association with outcome in the context of the new therapeutic is needed. We acknowledge that in some situations it may simply be premature to use a laboratory test as an integral biomarker in a clinical trial if the necessary preliminary data do not exist. The trial itself may offer the first opportunity to acquire the information that will permit development of a fit-for-purpose assay for a subsequent trial by using the test in an integrated or exploratory fashion. Nevertheless, it is important even in these more preliminary settings to ensure that the essential data are collected during the trial to support moving the biomarker forward if the investigational therapy looks promising.

Clinical validation of the biomarker – the process that establishes in clinical samples that the biomarker test identifies its concept of interest [1] – was not discussed at length, but is an important topic itself. For example, for biomarkers that are meant to identify subgroups of patients (or distinguish individuals with a disease from those

without), there are many considerations that can greatly affect study conclusions including obtaining an unbiased sample of the populations of interest and external validation or cross-validation of the subgroups [19].

Investigators may find helpful the materials developed by the NCI that provide guidance about information required to support use of biomarker tests in early-phase trials, including a checklist for early phase biomarker tests [41], as well as separate checklists for tests of DNA-based *in situ* hybridization (FISH/CISH), immunohisto-chemistry and DNA-based mutation tests [20]. Additional NCI-provided references may also be valuable in this setting [14,15,26,42]. The discussion here was geared toward clinicians with limited laboratory experience aiming to incorporate biomarkers into early-phase trials with the goal to bridge the gap between clinical and laboratory investigators' understanding of biomarker development. Collaboration between principal clinical investigators and laboratory researchers is key to successful inclusion of biomarkers in early-phase trials. Working together, the research team will be more likely to choose a biomarker test that advances superior treatments and leads to improved patient outcomes.

Executive summary

• To introduce academic clinicians to key considerations for incorporating biomarkers into early phase clinical trials, including discussion of biomarker test analytical validation in the early phase setting.

Overview of incorporating a biomarker into an early-phase trial

- Define the role of the biomarker in the trial.
- Select a fit-for-purpose biomarker test and laboratory.
- Describe the test and operating procedures.
- Carry out and report analytical validation appropriate for the purpose and setting of the trial.
- Implement testing in the trial.
- Plan for the future of the biomarker test.
- Examples: deciding whether biomarker analytical results are acceptable for the proposed trial
- Examples demonstrate that the scope of analytical validation required will depend on the purpose that the biomarker serves in the trial.
- Conclusion/future perspective
- Collaboration between principal clinical investigators and laboratory researchers is key to successful inclusion of biomarkers in early-phase trials.

Disclosure

The views presented in this article are those of the authors and should not be viewed as official opinions or positions of the National Cancer Institute, National Institutes of Health, or U.S. Department of Health and Human Services.

Financial & competing interests disclosure

All authors of this manuscript are US federal government employees of the National Cancer Institute, NIH, so this article will be deposited in NIHMS system/PMC. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Guidance on author sequence

Author sequence is at the authors' discretion; however, Future Medicine journals suggest following the recommendations in GPP3 Appendix Table 2 (http://www.ismpp.org/gpp3), whereby authors are listed either in order of the level of their contribution, or alphabetically. The corresponding author should always be indicated.

References

Papers of special note have been highlighted as: •• of considerable interest

- 1. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. US FDA, MD, USA (2016). www.ncbi.nlm.nih.gov/books/NBK326791/
- 2. US FDA. Biomarker Qualification Program: Biomarker Guidances and Reference Materials www.fda.gov/Drugs/DevelopmentApprovalPro cess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/ucm536018.htm#pubs

- Chau CH, Rixe O, McLeod H, Figg WD. Validation of analytical methods for biomarkers employed in drug development. *Clin. Cancer Res.* 14(19), 5967–5976 (2008).
- Cummings J, Raynaud F, Jones L, Sugar R, Dive C. Fit-for-purpose biomarker method validation for application in clinical trials of anticancer drugs. Br. J. Cancer 103(9), 1313–1317 (2010).
- Dancey JE, Dobbin KK, Groshen S et al. Guidelines for the development and incorporation of biomarker studies in early clinical trials of novel agents. Clin. Cancer Res. 16(6), 1745–1755 (2010).
- Kessler LG, Barnhart HX, Buckler AJ et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat. Methods Med. Res. 24(1), 9–26 (2015).
- 7. Pennello GA. Analytical and clinical evaluation of biomarkers assays: when are biomarkers ready for prime time? *Clin. Trials* 10(5), 666–676 (2013).
- Institute of Medicine (US) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease. Evaluation of biomarkers and surrogate endpoints in chronic disease. National Academies Press (US), Washington, DC, USA. www.ncbi.nlm.nih.gov/books/NBK220297/
- 9. Salgado R, Moore H, Martens JWM *et al.* Societal challenges of precision medicine: bringing order to chaos. *Eur. J. Cancer* 84, 325–334 (2017).
- McShane LM, Hunsberger S. An overview of Phase II clinical trial designs with biomarkers. In: *Design and Analysis of Clinical Trials for Predictive Medicine*. Matsui, Buyse, Simon (Eds). Chapman and Hall/CRC, Boca Raton, FL, USA, 71–87 (2015).
- 11. McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into Phase II trials. *Clin. Cancer Res.* 15(6), 1898–1905 (2009).
- 12. Workman P. How much gets there and what does it do? The need for better pharmacokinetic and pharmacodynamic endpoints in contemporary drug discovery and development. *Curr. Pharm. Des.* 9(11), 891–902 (2003).
- Lee JW, Devanarayan V, Barrett YC et al. Fit-for-purpose method development and validation for successful biomarker measurement. Pharm. Res. 23(2), 312–328 (2006).
- 14. National Cancer Institute. Biomarker, Imaging and Quality of Life Studies Funding Program (BIQSFP). www.cancer.gov/about-nci/organization/ccct/funding/biqsfp
- 15. Zweibel JA, Lively TG. Guidelines for biomarker assays used in CTEP-sponsored, early phase clinical trials performed under CTEP IND (2013). https://ctep.cancer.gov/protocoldevelopment/docs/biomarker_review_guidelines.doc
- National Cancer Institute. Phase 1, 2, Or 1/2 Letter Of Intent Submission Form V8.2 (2017).https://ctep.cancer.gov/protocoldevelopment/docs/loi_form.docx
- 17. Centers for Medicare & Medicaid Services (CMS). Clinical Laboratory Improvement Amendments (CLIA) (2017). www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html?redirect=/CLIA/05_CLIA_Brochures.asp
- 18. Moore HM, Kelly AB, Jewell SD *et al.* Biospecimen reporting for improved study quality (BRISQ). *Cancer Cytopathol.* 119(2), 92–101 (2011).
- Bossuyt PM, Reitsma JB, Bruns DE et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Radiology (2015). pubs.rsna.org/doi/abs/10.1148/radiol.20151515156
- 20. Templates for Clinical Assay Development. Cancer Diagn. Program CDP (2015). https://cdp.cancer.gov/resources/templates.htm
- NCI-provided templates for clinical assay development for immunohistochemistry, DNA-based *in situ* hybridization and DNA-based mutation assays.
- 21. McShane LM, Cavenagh MM, Lively TG *et al.* Criteria for the use of omics-based predictors in clinical trials. *Nature* 502(7471), 317 (2013).
- 22. McShane LM, Cavenagh MM, Lively TG *et al.* Criteria for the use of omics-based predictors in clinical trials: explanation and elaboration. *BMC Med.* 11, 220 (2013).
- Hall JA, Salgado R, Lively T, Sweep F, Schuh A. A risk-management approach for effective integration of biomarkers in clinical trials: perspectives of an NCI, NCRI, and EORTC working group. *Lancet Oncol.* 15(4), e184–e193 (2014).
- Kimmelman J, Resnik DB, Peppercorn J, Ratain MJ. Burdensome research procedures in trials: why less is more. J. Natl Cancer Inst. 109(4), doi:10.1093/jnci/djw315 (2017) (Epub ahead of print).
- 25. Hayes DF, Allen J, Compton C et al. Breaking a vicious cycle. Sci. Transl. Med. 5(196), 196cm6 (2013).
- 26. Program for the Assessment of Clinical Cancer Tests (PACCT) Strategy Group Members. Performance standards reporting requirements for essential assays in clinical trials. https://cdp.cancer.gov/scientific_programs/pacct/assay_standards.htm
- •• National Cancer Institute summary of performance standards for assays that are essential in clinical trials.
- 27. Clinical and Laboratory Standards Institute (CLSI). EP17A2: Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures – Second Edition.clsi.org/standards/products/method-evaluation/documents/ep17/
- Kessler LG, Barnhart HX, Buckler AJ et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat. Methods Med. Res. 24(1), 9–26 (2015).

- Clinical and Laboratory Standards Institute (CLSI). EP06A: Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach.clsi.org/standards/products/method-evaluation/documents/ep06/
- Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J. Natl Cancer Inst. 101(21), 1446–1452 (2009).
- 31. Denne JS, Pennello G, Zhao L, Chang S-C, Althouse S. Identifying a subpopulation for a tailored therapy: bridging clinical efficacy from a laboratory-developed assay to a validated *in vitro* diagnostic test kit. *Stat. Biopharm. Res.* 6(1), 78–88 (2014).
- 32. Center for Devices and Radiological Health (CDRH). *Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests.* US FDA, MD, USA (2007). www.fda.gov/RegulatoryInformation/Guidances/ucm071148.htm
- 33. Geschwind DH. Sharing gene expression data: an array of options. Nat. Rev. Neurosci. 2(6), 435 (2001).
- 34. Searle SR, Casella G, McCulloch CE. Variance Components. John Wiley & Sons, NJ, USA (2006).
- 35. Clinical and Laboratory Standards Institute (CLSI). EP05A3: evaluating quantitative measurement precision (2014). clsi.org/standards/products/method-evaluation/documents/ep05/
- Liu C, Liu A, Hu J, Yuan V, Halabi S. Adjusting for misclassification in a stratified biomarker clinical trial. *Stat. Med.* 33(18), 3100–3113 (2014).
- 37. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies. J. Clin. Oncol. 23(36), 9067–9072 (2005).
- 38. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, NY, USA (2006).
- 39. Youden WJ. Index for rating diagnostic tests. Cancer 3(1), 32-35 (1950).
- 40. Zhou X-H, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine (2nd Edition).* John Wiley & Sons, NJ, USA (2011).
- Study Checklist for CTEP-Supported Early Phase Trials with Biomarker Assays.ctep.cancer.gov/protocoldevelopment/docs/Study_Checklist_Early_Phase_Trials_Biomarker_Assays.docx
- 42. National Cancer Institute Division of Cancer Treatment & Diagnosis. The Clinical Assay Development Program (CADP) (2017). cdp.cancer.gov/scientific_programs/pacct/cadp.htm
- •• A list of resources compiled by NCI's Clinical Assay Development Program.