Trial Designs for Personalizing Cancer Care: A Systematic Review and Classification

Parvin Tajik1,2, Aleiko H. Zwinderman1, Ben W. Mol2, and Patrick M. Bossuyt1

Abstract

There is an increasing interest in the evaluation of prognostic and predictive biomarkers for personalizing cancer care. The literature on the trial designs for evaluation of these markers is diverse and there is no consensus in the classification or nomenclature. We set this study to review the literature systematically, to identify the proposed trial designs, and to develop a classification scheme. We searched MEDLINE, EMBASE, Cochrane Methodology Register, and MathSciNet up to January 2013 for articles describing these trial designs. In each eligible article, we identified the trial designs presented and extracted the term used for labeling the design, components of patient flow (marker status of eligible participants, intervention, and comparator), study questions, and analysis plan. Our search strategy resulted in 88 eligible articles, wherein 315 labels had been used by authors in presenting trial designs; 134 of these were unique. By analyzing patient flow components, we could classify the 134 unique design labels into four basic patient flow categories, which we labeled with the most frequently used term: single-arm, enrichment, randomize-all, and biomarker-strategy designs. A fifth category consists of combinations of the other four patient flow categories. Our review showed that a considerable number of labels has been proposed for trial designs evaluating prognostic and predictive biomarkers which, based on patient flow elements, can be classified into five basic categories. The classification system proposed here could help clinicians and researchers in designing and interpreting trials evaluating predictive biomarkers, and could reduce confusion in labeling and reporting. Clin Cancer Res; 19(17); 4578–88. ©2013 AACR.

Introduction

Recent progress in understanding the molecular basis of cancer has redefined the landscape for achieving stratified and personalized medicine for patients with cancer. Ongoing efforts concentrate on the translation of these molecular insights into biomarkers that can reliably guide application of existing and new cancer treatments. Biomarkers that are informative for selecting treatment can be broadly classified as prognostic or predictive biomarkers. Prognostic biomarkers classify patients treated with standard therapies—including no treatment, if that is standard practice—into subgroups with distinct expected clinical outcomes. Predictive biomarkers identify patients whose tumors are likely to be sensitive or resistant to a specific agent (1). Valid and definitive evidence showing the clinical use of these biomarkers can be obtained by conducting well-designed clinical trials with adequate sample size (1–3). These clinical trials are planned to evaluate biomarker-based treatment strategies. Their design often differs from the more conventional comparative clinical trials of interventions. Following the recent dramatic developments in the technologies for identifying potential novel biomarkers and the huge increase in the number of claimed biomarkers, there has also been a great expansion in the number of trial designs proposed for validation of biomarkers for treatment selection purposes.

A rapid review of the biomarker trial designs in the oncology literature suggests substantial variability in the designs, as well as in the terms proposed by authors for labeling them. This makes retrieval, interpretation, comparison, and critical appraisal of these types of studies complicated for consumers of trial results, who can be practicing physicians, researchers, or policymakers. The variability in labeling is a phenomenon typical for many rapidly developing areas of science. However, now that the field is maturing, and experts in other fields of medicine, such as cardiovascular diseases and infectious diseases, have started to apply the methodologic achievements of the oncology biomarker trials, it is time for harmonizing the terminologies in use and framing a classification of the designs. A harmonized terminology for describing biomarker clinical trials and a simple classification scheme may help in speeding up the translation process of the
biomarker findings and could assist in paving the way for making personalized medicine a reality.

Here, we report on a systematic review of literature on trial designs for evaluating biomarkers for treatment selection. The review is based on a comprehensive and systematic search in multiple databases. We propose a classification scheme based on the flow of patients in these trials. The classification system is presented using recent biomarker clinical trials in oncology as examples, along with the main questions each category of designs can answer.

Materials and Methods

Literature search
We searched MEDLINE, EMBASE, Cochrane Methodology Register, and MathSciNet up to 15 January 2013, and hand searched references and cited articles of all included studies using the Web of Science database. The search filters that we developed and used in collaboration with a clinical librarian are presented in the Supplementary Materials and Methods. Eligible for inclusion were methodologic articles that described one or more trial designs for identification and/or validation of prognostic or predictive biomarkers for treatment selection. The search was not limited to oncology. No language restrictions were applied.

One author (P. Tajik) independently identified potentially eligible articles by reading titles and abstracts, whereas a second author (P.M. Bossuyt) independently screened a random sample of 400 abstracts to ensure that no abstracts were missed. There was a 99% agreement on the selection of abstracts and disagreements were resolved in consensus discussions. Thereafter, full text copies of all potentially eligible articles were obtained and read in full. Articles were included if they satisfied the inclusion criteria.

Data extraction
From all included articles and for each proposed trial design, we extracted detailed data on the proposed design label, trial objectives, patient flow elements, and the analysis plan. Our definition of patient flow is composed of the biomarker status of participants deemed eligible for the study, the intervention participants are assigned to (whether or not biomarker status is used for assigning study participants to the experimental treatment) and the comparator (standard treatment or both standard and experimental treatments).

Analysis
We started our analysis by developing the list of study labels from all included studies. For each label in the list, we explored the participant–intervention–comparator components of patient flow. We then clustered all study labels with identical components into disjoint categories. The most commonly used label for describing designs of each category was selected for labeling the corresponding category.

Results

Search results
Our initial search yielded 2,506 abstracts, of which 136 were deemed eligible based on title and abstract. After assessing the full texts of these 136 articles, 71 were included. Seventeen other articles were added by hand searching references and citing articles, resulting in a total of 88 articles in the final analysis (Supplementary Table S1). A summary of the search process is outlined in Fig. 1.

Trial design labels
From the included articles, we could extract 315 design labels. The identified labels and the definitions as presented in the studies are listed in the Supplementary Table S2. By analyzing the extracted labels along with their patient flow components, we found 134 unique combinations of label and patient flow elements. There were many trial designs with the very same patient flow elements, which had received different labels. There were also a few labels used for describing designs with completely different patient flows. The 134 unique label and patient flow combinations are presented in Table 1. By comparing the patterns of all the included designs, we could distinguish four basic and distinct patient flow categories, as well as a fifth category, consisting of combinations of two or more of the four basic patient flow elements (Table 1). In a later section, we discuss these categories in more detail.

Figure 1. Flow chart of search results.
**Trial design categories**

To ease the presentation, we present the flow elements for two treatment options, labeled as experimental (Exp) and control (Ctrl), in the presence of a single, binary biomarker (Fig. 2). However, the flow diagrams are generalizable to conditions with more than two treatment options and in which the biomarker has more than two levels, or is numeric. For sake of simplicity and consistency, we define biomarker positivity as the biomarker status that is associated with a better outcome on the experimental treatment. Therefore, in cases where overexpression of the biomarker is associated with a worse response to treatment, we consider the normal expression as biomarker-positive and overexpression as biomarker-negative.

**Single-arm.** In single-arm trials, all patients, irrespective of their biomarker status, are included in the trial and all undergo the experimental treatment (Fig. 2A; refs. 4–8). This trial has no control group, and no random assignment.

**Enrichment.** With an enrichment design, all potentially eligible patients are first tested for the biomarker and only biomarker-positives are randomly assigned to the experimental or control treatment. Other patients are in principle excluded from further investigation in the study (Fig. 2B).

We found 12 labels for describing this design, which were all interchangeable and referred to the main feature of the design: biomarker status performs as a key trial eligibility criterion (Table 1).

The pivotal trial for trastuzumab is a well-known example of an enrichment design (9). Patients with HER2-positive breast adenocarcinoma were enrolled and randomly allocated to chemotherapy with or without trastuzumab. This study provided strong evidence that trastuzumab combined with chemotherapy improves outcomes among women with HER2-positive breast cancer.

**Randomize-all.** In designs in the randomize-all category, all patients meeting the trial eligibility criteria, irrespective of their biomarker status, are randomly allocated to either experimental or control treatment. Thereafter, associations between biomarker status and treatment response are evaluated (Fig. 2C).

Because its eligibility criteria is not restricted by biomarker status, it has commonly been labeled as "randomize-all" (2, 10–13), "all-comers" (14–18), or "unstratified" (19–22). The design is also called "traditional" (23–25) or "conventional" (1, 26) because it has the same patient flow elements as routine randomized controlled trials (RCT) for evaluating treatment options. It is possible that researchers evaluate a biomarker retrospectively, using data and stored biospecimens collected in previously completed RCTs. In such scenarios, trials are commonly labeled as "biomarker analysis within an existing RCT" (18) or "prospective/retrospective" (27). All these label variations mainly refer to the timing of introducing the biomarker question to the trial.

The type of randomization is another source of variability in labeling of randomize-all trials. In cases where a simple 1:1 randomization procedure is applied to all patients, trials are labeled as "simple randomization" (28). However, in cases where the biomarker under evaluation is binary or categorical with few categories, randomization can be done separately for each biomarker category through stratified randomization. Labels such as "biomarker-stratified" (29, 30), "stratified randomized" (31), "non-targeted RCT (stratified by marker)" (32), "stratified analysis" (25), "stratification" (33, 34), and "separate randomization" (35) all refer to this type of randomization. A special case of stratified randomization is when randomization is conducted by means of a Bayesian response-adaptive procedure, rather than a standard equal randomization procedure. The "Bayesian adaptive randomization design" by Zhou and colleagues is an example (36). It has also been called "outcome-based adaptive randomization" (5, 37).

The next source of variability in labeling the trials in the randomize-all category is the trial's statistical analysis plan. In many cases, authors have coined a design label for referring to a randomize-all design when analyzed on the basis of their novel analysis plan. Examples of such labels are "biomarker analysis" (38), "sequential testing" (39), "prospective subset" (2), "adaptive threshold" (40), "adaptive biomarker" (33), "adaptive signature" (41), "cross-validated adaptive signature" (42), and "generalized adaptive signature" (33). All these designs have a randomize-all patient flow structure, but they differ in their analysis plan.

The Sequential Tarceva in unresectable non–small cell lung cancer (NSCLC) trial (SATURN; ref. 43) has been labeled as "prospective-subset design" (2). In the SATURN trial, all eligible patients were randomly allocated to erlotinib or placebo plus standard supportive care. The trial had two primary objectives: evaluating the effectiveness of erlotinib separately in all patients and in patients with EGFR receptor (EGFR) immunohistochemistry-positive tumors. To address the multiple comparisons issue, authors used an $\alpha$ splitting technique; the $\alpha$ level of 5% was split between the two primary objectives: 3% for all patients and 2% for patients with EGFR immunohistochemistry-positive tumors.

There are five "adaptive" designs in this category, which are labeled after adaptive elements in their analysis plan. Each of these adaptive plans evaluates the efficacy of experimental treatment in all patients and, if not significant, tries to define a biomarker-defined subset that is responsive to the experimental treatment. In settings where a single continuous candidate biomarker is available but its positivity threshold is not predefined, adaptive threshold plans try to find such a threshold (40). Adaptive biomarker designs have been proposed to evaluate multiple binary biomarkers defined in advance (33). Adaptive signature (41) and cross-validated adaptive signature (42) designs develop a predictive combination of biomarkers in a training set of the trial and consequently evaluate it in a test set. The proposed "generalized adaptive signature design" uses a training set of the trial to select among several candidate biomarkers and optimize cutoff points, and thereafter evaluates the selected biomarkers in a test set (33).

**Biomarker-strategy.** The distinguishing feature of designs in the biomarker-strategy category is the inclusion of a new
Table 1. Patient flow elements of the extracted design labels for trials evaluating prognostic and predictive biomarkers. Labels clustered in the same bullet describe identical trial designs and are interchangeable.

<table>
<thead>
<tr>
<th>Design labels</th>
<th>Patients</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Patient flow category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrichment (1, 3, 11, 15, 16, 18, 29, 33, 37, 39, 44, 48, 61–75), targeted (2, 10–13, 15, 18, 20, 48, 64, 67, 69, 76, 77), enriched (23, 76), marker enrichment (78), biomarker-enrichment (56, 68), biomarker enriched (14, 23, 79, 80), biomarker selected (79), efficient targeted (28), clinically enriched (79), selection (2, 10, 81), RCT of test positives (48), screening enrichment (82)</td>
<td>M⁺</td>
<td>Exp</td>
<td>Std</td>
<td>2. Enrichment</td>
</tr>
<tr>
<td>Randomize-all (2, 10–13), traditional (23–25), standard (33), conventional (1, 26), simple randomization (28), untargeted (19, 19–22), unselected (18, 64), all-comers (14–18), all-comers (stratified by marker status; ref. 67), including test negatives and positives (73), test as baseline in RCT (48), prospective/retrospective (27), biomarker analysis (30, 38), biomarker analysis within an existing RCT (18), controlled cohort pharmacogenetic study (50), retrospective biomarker (79), indirect predictive biomarker-based (48), hybrid (77)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
<tr>
<td>Sequential testing strategy (16, 18)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
<tr>
<td>Prospective subset (2)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
<tr>
<td>Adaptive threshold (29, 64, 76, 78), biomarker adaptive threshold (74)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
<tr>
<td>Adaptive biomarker (33)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
<tr>
<td>Marker stratified (28), marker-stratified (16), biomarker-stratified (29, 30), biomarker-stratified with fall back analysis (69), stratified randomized (31), separate randomization (35), stratification (33, 34), stratified (69), stratified analysis (28), non-targeted RCT (stratified by marker; ref. 32), marker by treatment interaction (30, 53), marker-by-treatment interaction (8), treatment by marker interaction (3), treatment-by-marker interaction (64), marker x treatment interaction (85), interaction (13), treatment-marker interaction (11)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
<tr>
<td>Outcome-based adaptive randomization (5, 36), adaptive randomization (15, 76, 86), Bayesian adaptive randomization (2, 28, 62), combined dynamic multarm (80)</td>
<td>M⁺, M⁻</td>
<td>Exp</td>
<td>Std</td>
<td>3. Randomize-all</td>
</tr>
</tbody>
</table>

(Continued on the following page)
### Table 1. Patient flow elements of the extracted design labels for trials evaluating prognostic and predictive biomarkers. Labels clustered in the same bullet describe identical trial designs and are interchangeable (Cont’d)

<table>
<thead>
<tr>
<th>Design labels</th>
<th>Patients</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Patient flow category</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Biomarker-strategy, with biomarker assessment in the control arm (1)</td>
<td>M⁺, M⁻</td>
<td>M-based strategy</td>
<td>Std</td>
<td>a. With biomarker assessment in the control arm</td>
</tr>
<tr>
<td>- Biomarker-strategy without biomarker assessment in the control arm (1)</td>
<td>Unknown</td>
<td>M-based strategy</td>
<td>Std</td>
<td>b. Without biomarker assessment in the control arm</td>
</tr>
<tr>
<td>- Modified marker-based strategy (31, 33, 53, 68), with randomized control (2)</td>
<td>M⁺, M⁻</td>
<td>M-based strategy</td>
<td>Exp, Std</td>
<td>c. With treatment randomization in the control arm</td>
</tr>
<tr>
<td>- Combination of biomarker trial designs (1)</td>
<td></td>
<td></td>
<td></td>
<td>5. Combination of patient flow categories</td>
</tr>
<tr>
<td>- Intermediate risk randomized (2), two-way stratified (49, 89), modified marker strategy (33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Discordant risk randomization (2), discrepant case randomized (31), discordant test results RCT (49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Two-stage sample-enrichment (54), two-stage enrichment (16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Tandem two-step phase II predictor biomarker evaluation (7), tandem two step (2, 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Adaptive patient enrichment (59), adaptive accrual (18, 37, 78), adaptive accrual based on interim analysis (6-6), adaptive modifying types of patients accrued (33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Adaptive parallel Simon two-stage (88), biomarker-adaptive parallel two-stage (7), adaptive parallel (2), two-stage Bayesian (63)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Bayesian covariate adjusted response-adaptive (BCARA) randomization (91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Adaptive (16, 23, 62, 67, 69), adaptive analysis (39), adaptive biomarker driven (14), multiarm multistage (67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Management strategy. This strategy is not the standard or experimental treatment, but a prespecified maker-based treatment strategy. For example, biomarker-positive patients would receive experimental therapy, whereas all biomarker-negative patients get standard of care. Eligible patients are randomized to this biomarker-based treatment strategy or to control treatment. In our review, we could identify three subtypes of this category.

**Biomarker-strategy, with biomarker assessment in the control arm.** Biomarker status is assessed in all patients enrolled in the trial, who are then randomly allocated to either the biomarker-strategy arm or to standard treatment (ref. 1; Fig. 2D). Some other labels for this design type were biomarker-strategy (15, 38, 44), "marker-based strategy I" (45), "customized strategy" (12), "direct predictive biomarker-based" (46), and "biomarker-guided" (47). There were also other labels in use, such as "random disclosure" (48), "classifier randomization" (49), or "parallel controlled pharmacogenetic study" (50).

**Biomarker-strategy, without biomarker measurement in the control arm.** In settings where it is not feasible or ethical to evaluate the biomarker in all patients, biomarker status is only acquired in patients allocated to the biomarker-strategy arm (Fig. 2E). This design is also labeled as "biomarker-strategy with standard control" (2), "direct predictive biomarker-based" (46), "RCT of testing" (48), "test-treatment" (51), or "parallel controlled pharmacogenetic diagnostic study" (50).

**Biomarker-strategy, with treatment randomization in the control arm.** Sargent and Allegra (52) have proposed a modification of the biomarker-strategy design, wherein a second randomization between experimental versus control therapy replaces the control arm (Fig. 2F). They called it "modified marker-based strategy" design (53). Some other authors referred to the design as "marker-based strategy design II" (45), "augmented strategy" (13), or simply "marker-strategy" (28).

**Combination of patient flows.** A final, fifth category consists of trial designs in which two or more of the aforementioned basic patient flow structures are combined to form a new design. Combination of designs is usually required when the trial aims at evaluating multiple hypotheses, multiple biomarkers, multiple treatments, or when the trial has several stages. Freidlin and colleagues have similarly proposed a category called "combination of biomarker trial designs" in a comparable review of study designs (1).

The simplest combination is when in enrichment trials biomarker-negative patients are not excluded from the study but assigned to control treatment(s) for which the outcomes are assessed. Here, an enrichment flow is combined in parallel with a single-arm trial of standard therapy in biomarker-negative patients. Most authors have referred to this flow as "hybrid" design. An example of a hybrid design is the Trial Assigning Individualized Options for Treatment (TAILORx). The study was designed to evaluate Oncotype Dx, a 21-gene recurrence score, in tamoxifen-treated patients with breast cancer. In this trial, patients are divided into three subgroups of low, intermediate, and high risk based on their Oncotype Dx recurrence score. Low-risk patients receive hormonal therapy, high-risk patients receive both hormonal and chemotherapy, whereas patients at intermediate risk are randomized to hormonal therapy or chemotherapy plus hormonal therapy. This trial is a parallel combination of enrichment trial in intermediate-risk group and two single-arm trials in the low- and high-risk groups. Other labels used for this design are "intermediate risk randomized" (2) or "two-way stratified" (49) design.

The Microarray In Node-negative Disease may Avoid Chemotherapy (MINDACT) trial also has an enrichment element in its patient flow. Here, patients with discordant risk estimations by tumor’s clinicopathologic features and a
70-gene signature (MammaPrint) are eligible for randomization. Patients with concordant risk estimations receive control treatment. Consenting discordant patients are randomized to treatment determined on the basis of the clinocopathologic risk category versus MammaPrint risk category. This way, the MINDACT is a trial combining three flow types: enrichment, single-arm, and biomarker-strategy. This combination has been described in the literature as “discrepant case randomization” (31), “discordant risk randomization” (2), and “discordant test results RCT” (48). Simon refers to this trial as a modified marker-based strategy (33), as only patients for whom the treatment assignment is influenced by biomarker results are randomized.

Staged trial designs are also, in essence, a combination of basic patient flows. For instance, the proposed “two-stage sample-enrichment” design by Liu and colleagues (54) starts with accruing only biomarker-positive patients during the initial stage of the trial. At the end of the first stage, an interim analysis is conducted comparing the outcome of the experimental versus control treatment in biomarker-positives. If the results are not promising for the new treatment, accrual stops and no treatment benefit is claimed. Otherwise, accrual continues with recruiting unselected population. This design is a combination of an enrichment and a traditional flow, conditional on the result of the interim analysis.

Contrariwise, in “adaptive patient enrichment” design (55)—also labeled as “adaptive accrual” (18, 37)—the trial begins with a biomarker-stratified first stage in which it accrues both biomarker-positive and -negative patients. If the results of an interim analysis comparing the outcome of the experimental versus control treatment in biomarker-negatives are not promising, accrual to biomarker-negative subgroup is terminated and the second stage continues as an enrichment trial in biomarker-positive patients until the planned total sample size is reached.

**Effects assessed by each category of designs**

There are four types of effects we are commonly interested in when designing a biomarker trial: the treatment effect, the biomarker effect, the biomarker by treatment effect, and the strategy effect. These are presented with 8 study questions in Table 2. The treatment effect (experimental vs.

---

**Table 2.** List of effects that can be assessed and questions that can be answered by the trials of each design category

<table>
<thead>
<tr>
<th>Questions trial can answer</th>
<th>Single-arm</th>
<th>Enrichment</th>
<th>Randomize-all</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment effects</strong></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q1. How does the experimental treatment compare with the control treatment in biomarker-positives?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q2. How does the experimental treatment compare with the control treatment in biomarker-negatives?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q3. How does the experimental treatment compare with the control treatment in overall study population?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Biomarker effects</strong></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q4. Is the biomarker status associated with the outcome in the standard of care group? (Is the biomarker prognostic?)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q5. Is the biomarker status associated with the outcome in the experimental treatment group?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Biomarker by treatment effect</strong></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q6. Is the biomarker status associated with a benefit of experimental treatment? (Is the biomarker is predictive?)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Strategy effects</strong></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q7. How does the biomarker-based treatment strategy compare with the control treatment in the overall study population?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Q8. How does the biomarker-based treatment strategy compare with the experimental treatment in the overall study population?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
control) can be estimated separately for biomarker-positive patients (Q1), for biomarker-negative patients (Q2), and in the overall population (Q3). Single-arm trials do not answer any of these questions, as they lack a control arm to allow comparisons. Enrichment trials recruit biomarker-positive patients and allocate them to experimental or control treatment, thus letting us evaluate the effect of treatment, but only in biomarker-positive patients (Q1; ref. 1). Randomize-all trials recruit patients from the whole spectrum of the biomarker values and allocate them randomly to either of the two treatment options. Therefore, they allow estimation of the treatment effect in biomarker-positives, -negatives, and in the overall population (Q1–3).

To evaluate if a biomarker is prognostic (Q4), one needs to compare the outcome of biomarker-positive and -negative patients on control treatment, a comparison which is possible in randomize-all study designs but not in single-arm or enrichment trials (1, 28). The only effect one can estimate in the single-arm trials is the biomarker effect in the experimental arm, whether biomarker status is associated with the outcome of experimental treatment (Q5).

To assess the predictive capacity of a biomarker (Q6)—the biomarker by treatment effect—ones needs to have the outcome of biomarker-positive and -negative patients separately after experimental and control treatments. Randomize-all trials allow such an assessment by providing estimates for all four aforementioned outcomes (1).

In the biomarker-strategy trials, patients are randomized to treatment strategies, instead of treatments. This feature permits direct comparison of the outcomes of a biomarker-based strategy versus those in a strategy of control treatment for all (Q7). Addition of a randomization to the control arm of a biomarker-strategy trial allows direct comparison with a strategy of experimental treatment for all (Q8; ref. 56). With a randomize-all trial, one cannot estimate these effects using randomization properties. By combining marker-positivity rate, outcome of experimental treatment in biomarker-positives and outcome of control treatment in biomarker-negatives, an indirect estimate of strategy effects, can be obtained (57), though it might be biased considering confounding effects of other prognostic factors and it may miss the additional effects of testing, apart from treatment selection (58). In biomarker-strategy designs, there are biomarker-positive and -negative patients who are treated with the control treatment, so one can assess prognostic capacity of the biomarker in all three subtypes. However, the evaluation of biomarker-predictive capacity is only possible when treatment is randomized in the control arm allowing estimation of experimental treatment effect in biomarker-negative patients (1).

**Discussion**

This systematic review documented a substantial variability in the labeling of clinical trial designs for evaluating biomarkers for treatment selection in individual patients. We identified more than 300 labels, half of which were unique. In evaluating the heterogeneity in design labels, we used a classification scheme based on patient flow components of the corresponding trial designs. Under each of the four basic patient flow categories, several labels could be categorized, where some labels are completely interchangeable terms. Other designs, while having the same patient flow elements, carried specific objectives or had diverging analysis plans, which authors have labeled differently, to emphasize these distinctive aspects.

Using our patient flow scheme, one would be able to classify biomarker trial designs into a small set of basic categories. This could be useful for identifying similarities between novel designs and existing ones, or to evaluate proposed modifications of existing designs. It could also be helpful in reducing the confusion around the design of biomarker trials and help with standardizing the reporting of biomarker clinical trials. Because the classification is based on patient flow, it is directly connected to the possible comparisons that can be made in the trial and, consequently, the questions that could be potentially answered by the trial.

By comparing the designs in Table 2, a biomarker-strategy design with treatment randomization in the control arm seems a very attractive option, allowing for direct estimation of all biomarker-related effects, yet this feature might come at the cost of a large sample size. Nevertheless, in situations where the biomarker-strategy is complex—having a large number of treatment options or biomarker categories—or when the trial is planned primarily for confirmatory assessment of a specific biomarker-based strategy, a biomarker-strategy trial can be the design of choice. A limitation of biomarker-strategy trials is that assessments are restricted to a prespecified biomarker-treatment combination strategy and they cannot be used for further identification or validation of other biomarkers.

Randomize-all trials also allow assessment of all biomarker-related effects, but provide indirect estimates of strategy effects. An attractive aspect of randomize-all trials is that they allow identification and evaluation of biomarkers that were not specified in the design phase of the trial. Single or multiple biomarkers can be studied and multimarker models can be developed and tested in trials of randomize-all category. If one collects and stores biologic specimens from participants of a randomize-all trial in biobanks, the trial data can be used later to identify or evaluate single or multiple biomarkers, possibly not even known at the time of trial design (59). However, as all analyses that emerge after designing the trial are considered post hoc and exploratory, cross-validation and/or independent validation approaches are required to establish the use of biomarkers.

Enrichment design can be selected when there is a strong prior biologic evidence that the experimental treatment has no effect in biomarker-negative patients (3, 20). Yet, a positive trial does not prove the use of the biomarker because there may exist a positive treatment effect in the unevaluated biomarker-negative patients (1, 3).

Biomarker trials have predominantly been proposed and discussed in the setting of phase III trials in oncology. All categories identified in this review apply when designing a phase III biomarker trial, though some designs have been
suggested primarily for a phase II setting. These include randomize-all designs with adaptive randomization, such as the outcome-based trials with Bayesian adaptive randomization (5, 36), as well as some of the designs in the combined category, such as tandem two-phase II predictor marker evaluation (60), which aim at finding a promising treatment/biomarker pair that can be moved forward to a phase III evaluation.

To our knowledge, this review is the first systematic review of trial designs for evaluation of prognostic and predictive biomarkers. Several other narrative reviews are available, most often written by experts (1, 2, 33, 35, 53), who commonly have discussed a selected series of biomarker trial designs and used their personally preferred design labels. Yet our review was also not without limitation, which was the shortcoming of our search strategy in the electronic databases. The retrieval of methodologic articles is challenging because no specific keywords are available to distinguish articles that have described or presented a method from those which have just applied that method. Most of the search terms we could use were nonspecific. Even the terms “prognostic” and “predictive” are commonly used by authors in other situations. To compensate for these challenges, we designed our search strategy to be broad and sensitive. We also looked for relevant articles by checking the references and citations of the selected retrieved articles.

Biomarkers are changing the way doctors handle many cancers, but there is still a long way to go before biomarkers reach their full potential in cancer management. With the advent of rapidly growing technologies for measuring new biomarkers, there is a parallel need for validating the clinical use of such biomarkers for selection treatment of individual patients. Well-designed and properly conducted trials may support the timely introduction of new biomarkers in clinical management, to the benefit of cancer patients.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: P. Tajik, P.M. Bossuyt
Development of methodology: P. Tajik, A.H. Zwinderman, B.W. Mol, P.M. Bossuyt
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): P. Tajik, B.W. Mol, P.M. Bossuyt
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis): P. Tajik, A.H. Zwinderman, B.W. Mol, P.M. Bossuyt
Writing, review, and/or revision of the manuscript: P. Tajik, A.H. Zwinderman, B.W. Mol, P.M. Bossuyt
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P. Tajik
Study supervision: P.M. Bossuyt

Acknowledgments
The authors thank Ms. Faridi van Eten-Jamaludin for her help with developing the search strategy and searching the databases.

Grant Support
Funding for this research was provided by The Netherlands Organization for Health Research and Development (ZonMW), the Hague, the Netherlands (grant number 152002026).

Received December 3, 2012; revised May 24, 2013; accepted June 9, 2013; published OnlineFirst June 20, 2013.

References


