

Review

# Biomarker-Guided Non-Adaptive Trial Designs in Phase II and Phase III: A Methodological Review

Miranta Antoniou <sup>1,2,\*</sup>, Ruwanthi Kolamunnage-Dona <sup>1,2</sup> and Andrea L. Jorgensen <sup>1,2</sup>

<sup>1</sup> MRC North West Hub for Trials Methodology Research, Liverpool L69 3GL, UK;

Ruwanthi.Kolamunnage-Dona@liverpool.ac.uk (R.K.-D.); A.L.Jorgensen@liverpool.ac.uk (A.L.J.)

<sup>2</sup> Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, Liverpool L69 3GL, UK

\* Correspondence: Miranta.Antoniou@liverpool.ac.uk; Tel.: +44-(0)151-794-9752

Academic Editor: Stephen B. Liggett

Received: 8 October 2016; Accepted: 11 January 2017; Published: 25 January 2017

**Abstract:** Biomarker-guided treatment is a rapidly developing area of medicine, where treatment choice is personalised according to one or more of an individual's biomarker measurements. A number of biomarker-guided trial designs have been proposed in the past decade, including both adaptive and non-adaptive trial designs which test the effectiveness of a biomarker-guided approach to treatment with the aim of improving patient health. A better understanding of them is needed as challenges occur both in terms of trial design and analysis. We have undertaken a comprehensive literature review based on an in-depth search strategy with a view to providing the research community with clarity in definition, methodology and terminology of the various biomarker-guided trial designs (both adaptive and non-adaptive designs) from a total of 211 included papers. In the present paper, we focus on non-adaptive biomarker-guided trial designs for which we have identified five distinct main types mentioned in 100 papers. We have graphically displayed each non-adaptive trial design and provided an in-depth overview of their key characteristics. Substantial variability has been observed in terms of how trial designs are described and particularly in the terminology used by different authors. Our comprehensive review provides guidance for those designing biomarker-guided trials.

**Keywords:** biomarker-guided trial design; clinical research design; phase II; phase III; personalized medicine; predictive biomarker; prognostic biomarker; non-adaptive trial designs; clinical trials methodology; sample size

## 1. Introduction

The rapidly developing field of 'personalized medicine' [1], also known as 'individualized medicine', 'stratified medicine', or 'precision medicine' is allowing scientists to treat patients by providing them with a specific regimen according to their individual demographic, genomic or biological characteristics. The latter two aforementioned characteristics are collectively known as biomarkers [2]. The terms 'personalized medicine' and 'individualized medicine' often create confusion in literature, as in reality, the objective of this approach is to identify demographic- or biomarker-defined subgroups. Thus, as it still remains a population and not an individualized approach, the terms 'stratified' or 'precision' medicine are often considered to be more accurate. The National Institutes of Health Biomarkers Definitions Working Group [3] defined a biomarker to be "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [1,4–7]. Biomarkers related to clinical outcome which are measured before treatment commences can be classified as either prognostic or predictive biomarkers. Prognostic biomarkers provide information

regarding the likely progression of a disease without taking into account any specific treatment, whilst predictive biomarkers provide information about the patient's outcome given a certain treatment, i.e., their likely response to the treatment [4,7–34]. Prior to utilizing a patient's biomarker information in clinical practice, it is necessary that they have been robustly tested in terms of analytical validity (the results of testing a specific biomarker or biomarkers can be trusted), clinical validity (the results obtained from the test correlates with important clinical information) and clinical utility (the test will be useful in ameliorating patients' health) [9,13,19,25].

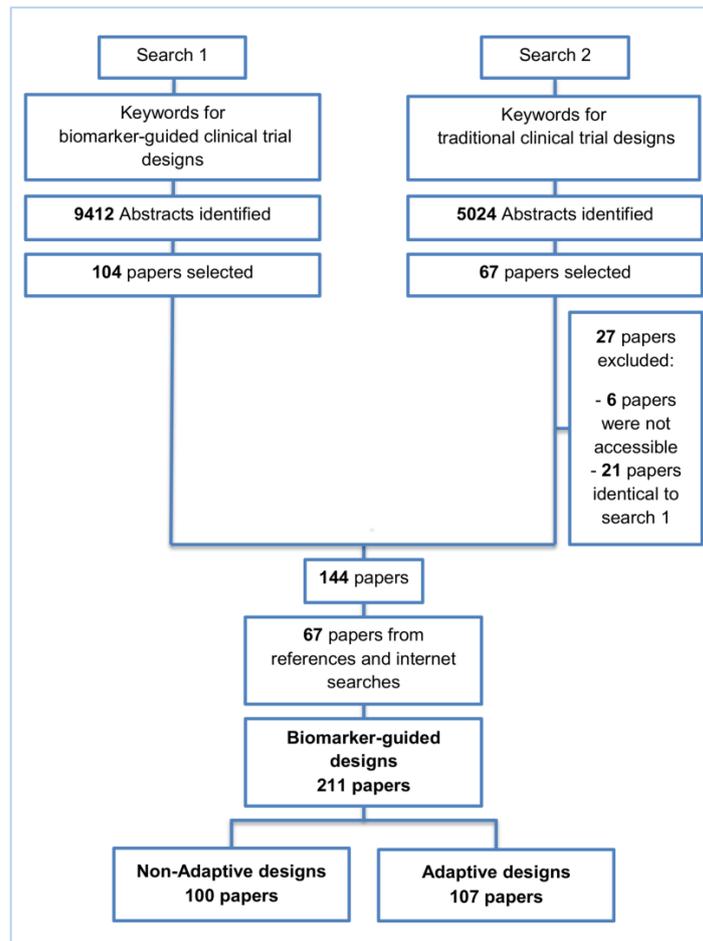
A number of phase II and phase III trial designs have been proposed for testing the clinical utility of prognostic biomarkers. Due to the large amount of literature in this field, we have split our review into two broad categories, i.e., the biomarker-guided non-adaptive trial designs which are presented in the current study and the biomarker-guided adaptive trial designs. The latter are extensively discussed in our published paper "Biomarker-Guided Adaptive Trial Designs in Phase II and Phase III: a Methodological Review", Antoniou et al., 2016 [35].

In this review we aim to communicate the different non-adaptive biomarker-guided trial designs, which can be either randomized or non-randomized designs (e.g., single-arm designs), proposed in the literature so far and to report on the potential advantages and weaknesses of each. Although not included in the paper by Antoniou et al., 2016 [35] which describes and discusses adaptive designs, some designs discussed in the current paper, although not adaptive in the traditional sense, they own an adaptive element.

## 2. Methods and Findings

We undertook a search of the MEDLINE (Ovid) database, restricted to published papers in the English language within the previous ten years aiming to identify articles which describe and discuss biomarker-guided trial designs. Traditional trial designs, i.e., designs which do not incorporate biomarkers aiming to aid in making treatment decisions (we will refer to as 'traditional' trial designs) are part of our literature review search strategy in order to help us identify and distinguish any potential reference to biomarker-guided designs, as the finding of the appropriate keywords in Medline database for biomarker-guided designs was challenging. Furthermore, the restriction of published papers within the past decade was made not only because of the large amount of literature in this field, but also for the identification of the most recent trial designs. Two separate strategies as illustrated in Figure 1 were used to identify relevant articles, and the keywords utilized in the search are presented in S1 Keywords. Our initial search resulted in 9412 and 5024 relevant titles for biomarker-guided clinical trial designs and traditional trial designs, respectively. From the 9412 papers, 104 articles were included based on their title and abstract. From the 5024 papers, 40 articles were included based on their title and abstract and after removing inaccessible articles or those already identified in the search for biomarker-guided trial designs. An additional 67 eligible papers were identified from searching both the reference list of included articles and the internet (the internet searches were performed using the same keywords as those for the Ovid strategy), making a total of 211 included papers. Of these 211 included papers, biomarker-guided non-adaptive trial designs were referred to in 100 papers; 107 papers for biomarker-guided adaptive trial designs were reviewed in our published paper Antoniou et al., 2016 [35]. In the total number of 211 papers, some papers are referred to both adaptive and non-adaptive designs. Articles from references and internet searches which did not provide further information on each broad category of biomarker-guided designs were not included. Cited books, web pages for actual trials and papers published before 2005 are also not included in these numbers. For each included paper, the following details were extracted: definition of the trial design(s) referred to in the paper, how patients were screened and/or randomized based on their biomarker status, treatment groups randomized to, as well as other key information relating to the trial design and methodology, including advantages and limitations. Where reference was made in the included papers to an actual trial which had adopted a particular biomarker-guided non-adaptive trial design, the clinical field with which the trial was associated was also recorded. However, a review of

all implementations of the different trial designs in practice is beyond the scope of this paper; however, and is an area for potential future work. Therefore, it is important to highlight that even where no evidence of the implementation of a particular design was found in the papers included in our review, the design may well be currently in use in ongoing trials.



**Figure 1.** Flow diagram of the review process. From our search strategy a total number of 211 papers have been identified giving information regarding not only the biomarker-guided designs but also general information about personalized medicine and biomarkers. Before arriving at 211 papers, books, web pages for actual trials and papers published before 2005 were excluded. The 211 papers are split into two overlapping sets of 100 and 107 papers. The total of 207 is less than 211 due to overlap of papers, and also due to the fact that some articles referring to general information about personalized medicine and biomarkers and articles which do not provide further information on each broad of biomarker-guided designs were excluded. The 107 papers for biomarker-guided adaptive trial designs were reviewed in our published paper Antoniou et al. (2016) [35].

In our review, we identified five main biomarker-guided non-adaptive trial designs namely: (i) single-arm designs; (ii) enrichment designs; (iii) randomize-all designs; (iv) biomarker-strategy designs and (v) other designs. Within each main design several subtypes and extensions were also identified. Graphical representations of the main designs and subtypes are given in Figures 2–16. Graphical representations of the extensions are given in Figures S1–S4 included in File S1–S4. The characteristics and methodology of the main design types and subtypes are discussed below and are summarized in Table 1, whilst information on the extensions are discussed in File S1–S4. Furthermore, sample size formulae for each biomarker-guided design are provided in Table 2.

**Table 1.** Types of Biomarker guided non-adaptive designs proposed within the last ten years.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<p><b>Single arm designs</b> (7 papers) [30,36–41] (see Figure 2)</p>	<p>Useful for initial identification and/or validation of a biomarker.</p>	<p>(A1) Considered as a simple statistical design as there is no need for randomization of patients.</p>	<p>(L1) There is no distinction between prognostic and predictive biomarker as patients are not randomized to experimental and control treatment arms.</p>
<p><b>Also called:</b> Nonrandomized clinical trial design, Uncontrolled Cohort Pharmacogenetic Study design</p>		<p>(A2) Simple logistics.</p>	
<p><b>Examples of actual trials:</b> None identified <sup>a</sup></p>		<p>(A3) Not complex statistical design</p>	
		<p>(A4) In some cases, these designs may be viewed as ethical as all patients are given the opportunity to experience the experimental treatment. However, they may be viewed as unethical if the novel treatment does not benefit a subgroup of patients or causes adverse events.</p>	
<p><b>Enrichment designs</b> (71 papers) [1,4,7–9,11,13,15,16,18,19,21,23,25–33,36,42–86] (see Figure 3)</p>	<p>Useful when we aim to test the treatment effect only in biomarker-positive subset for which there is prior evidence that the novel treatment is beneficial, but the candidate biomarker requires prospective validation.</p>	<p>(A5) Evaluates the effect of the experimental treatment in the biomarker-positive subgroup in a simple and efficient way.</p>	<p>(L2) Do not assess whether the experimental treatment benefits the biomarker-negative patients, thus we cannot obtain information about this subgroup. Also unable to demonstrate whether the targeted treatment is beneficial in the entire study population.</p>
<p><b>Also called:</b> Targeted design, Selection design, Efficient Targeted design, Biomarker-Enrichment design, Marker-enrichment design, Gene enrichment design, Enriched design, Clinically enriched Phase III study design, Clinically Enriched Trial design, Biomarker-Enriched design, Biomarker Enriched design, Biomarker Selected trial design, Screening enrichment design, Randomized Controlled Trial (RCT) of test positive design, Population enrichment design</p>	<p>Useful when it is not ethical to assign biomarker-negative patients to the novel treatment for which there is prior evidence that it will not be beneficial for this subpopulation, or that it will harm them.</p>	<p>(A6) Provides clear information about whether the novel treatment is effective for the biomarker-positive subgroup, thus these designs can identify the best treatment for these patients and confirm the usefulness of the biomarker.</p>	<p>(L3) Do not inform us directly about whether the biomarker is itself predictive because the relative treatment efficacy may be the same in the unevaluated biomarker-negative patients. Since these designs only enrol a subgroup of patients, they do not allow for full validation of the marker’s predictive ability. For full validation, a trial would need to randomize all patients in order to test for a treatment–biomarker interaction.</p>

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<p><b>Examples of actual trials:</b> CRYSTAL [49], BRIM 3 [49–51], EURTAC [49], CLEOPATRA [49], PROFILE 1007 [49,50], LUX-Lung [49], NSABP B-31 and NCCTG N9831 [4,15,16,18,19,28–31,36,44,46,52–60], CALGB-10603 [61], CATNON [62], CODEL [62], Evaluation of epidermal growth factor receptor variant III (EGFRvIII) peptide vaccination [62], N0923 [7,21], Flex study [64], TOGA trial [47], IPASS [33,43], N0147 [29], PetaCC-8 [29,47], C80405 [29], ECOG E5202 [29]</p>	<p>Recommended when both the cut-off point for determination of biomarker-status of patients and the analytical validity of a biomarker are well established.</p>	<p>(A7) Reduced sample size as the assessment of treatment effect is restricted only to biomarker-positive subgroup. Therefore, if the selected biomarker is “biologically correct” and reliably measured, the used enrichment strategy could result in a large saving of randomized patients.</p>	<p>(L4) Researchers should carefully decide whether or not to follow this strategy as it may be of limited value due to the exclusion of biomarker-negative patients. It may be that the entire population could benefit from the experimental treatment equally irrespective of biomarker status, in which case enrolling only the biomarker-positive patients will result in slow trial accrual, increase of expenses and unnecessary limitation of the size of the indicated patient population.</p>
		<p>(A8) Enables rapid accumulation of efficacy data.</p>	<p>(L5) Concern over an ethical problem as we cannot include individuals in a clinical trial if it is believed that the treatment is not effective for them, as raised by the US Food and Drug Administration (FDA) [50]. It was based on the facts that the experimental treatment can only be approved for a particular biomarker-defined subpopulation (i.e., biomarker-positive patients) if a companion diagnostic test is also approved, and how the test can be approved if the Phase III trial does not show that the novel treatment does not benefit the biomarker-negative patients.</p>
		<p>(A9) Allow us to avoid potential dilution of the results due to the absence of biomarker-negative patients. For example, if the design had included the biomarker-negative population and the biomarker positivity rate was low as compared to the biomarker negative rate, then the estimation of the overall treatment effectiveness could be diluted as it would be driven by the biomarker-negative subset.</p>	<p>(L6) The accuracy of diagnostic devices used to identify the biomarkers, e.g., biomarker assays, is not always correct [45]. This can result in incorrect selection of biomarker-positive patients and therefore these patients will erroneously be enrolled in a trial yielding biased treatment effect estimates. For example, even when the experimental treatment works well for a specific subgroup, if the biomarker assay is not able to identify this subgroup robustly then a promising treatment may be abandoned.</p>
		<p>(A10) Can be attractive in terms of speed and cost, meaning that patients are provided with tailored treatment sooner.</p>	

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<p><b>Marker Stratified designs</b> (45 papers) [4,10,12,13,15–19,21,25–27,30,31,33,44–46,49–51, 53,58,61,62,66,68,71–74,79–81,84–93] (see Figure 4)</p>	<p>Useful when there is evidence that the novel treatment is more effective in the positive biomarker-defined subgroup than in the negative biomarker-defined subgroup but there is insufficient compelling data indicating that the experimental treatment does not benefit the biomarker-negative patients.</p>	<p>(A11) Ability to assess the treatment effect not only in the entire population but also in each biomarker-defined subgroup. Thus, this design can find the optimal treatment in the entire population and in each biomarker-defined subgroup.</p>	<p>(L7) In situations where there are several biomarkers and treatments this design may not be feasible as it involves randomization of patients between all possible treatment options and may require a large sample size.</p>
<p><b>Also called:</b> Marker-stratified design, Biomarker-stratified design, Stratified-Randomized design, Stratification design, Stratified design, Stratified Analysis design, Marker by treatment – interaction design, Marker-by-treatment interaction design, Treatment by marker interaction design, Treatment-by-marker interaction design, Marker × treatment interaction design, Treatment-marker interaction design, Biomarker-by-treatment interaction design, Non-targeted RCT (stratified by marker) design, Genomic Signature stratified designs, Signature-Stratified design, Randomization or analysis stratified by biomarker status design, marker-interaction design.</p>		<p>(A12) An ethical design even in situations where the biomarker is not useful as no treatment decisions are made based on biomarker status; all decisions are made randomly. Consequently, if the biomarker’s value is in doubt, this design may be preferred.</p>	<p>(L8) May not be feasible when the prevalence of the biomarker is low.</p>
<p><b>Examples of actual trials:</b> MARVEL (N023) [4,16,30,31,33,44,61,89], GALGB-30506 [15,61], RTOG0825 [45], EORTC 10994 p53 [12,66], IBCSG trial IX [18], MINDACT [18]</p>			<p>(L9) Might be expensive to test the entire population for its biomarker status.</p>
			<p>(L10) Measuring the biomarker up front may be logistically difficult.</p>
			<p>(L11) There is no guarantee of balanced groups for analysis.</p>
<p><b>Sequential Subgroup-Specific design</b> (11 papers) [13,14,19,22,53,57,58,60,69,91,94] (see Figure 5)</p>	<p>Recommended when prior evidence indicates that the biomarker-positive subpopulation benefits more from the novel treatment as compared to the biomarker-negative subpopulation.</p>	<p>(A13) Allows for the estimation of treatment effect in biomarker-positive and biomarker-negative subgroups.</p>	<p>(L12) Has less power when there is homogeneity of treatment across the different biomarker defined subgroups as compared to the overall/biomarker-positive designs.</p>

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<b>Also called:</b> sequential design, Fixed-sequence 2 design, hierarchical fixed sequence testing procedure		(A14) Preserves the overall type I error rates and allows for a smaller sample size than the parallel version mentioned below.	(L13) Need a much larger sample size than the overall/biomarker positive designs if we assume that the treatment effect is relatively homogeneous across the biomarker-defined subsets.
<b>Examples of actual trials:</b> PRIME [49], MARVEL [49]		(A15) Considered as the best direct evidence for clinical decision making as it tests the treatment effectiveness in both the biomarker-positive and biomarker-negative subset in a sequential way.	
		(A16) Do not require larger sample size than the overall/biomarker-positive designs when the prevalence of the biomarker-positive patients is small.	
<b>Parallel Subgroup-Specific design</b> (3 papers) [14,49,69] (see Figure 6)	Appropriate when the aim of the study is to give treatment recommendations for each biomarker-defined subgroup separately at the same time.	(A17) Same as (A13), (A16)	(L14) Same as (L12)
<b>Also called:</b> Phase III Biomarker-Stratified design			(L15) Allocates the overall level $\alpha$ between the two biomarker-defined subgroup tests which means that it will be more difficult to achieve statistical significance in the biomarker-positive subgroup.
<b>Examples of actual trials:</b> None identified <sup>a</sup>			
<b>Biomarker-positive and overall strategies with parallel assessment</b> (8 papers) [1,14,36,47,49,69,95,96] (see Figure 7)	Recommended when the aim of the study is to assess the treatment effect in both the entire population and in the biomarker-positive subset but not in the biomarker-negative population.	(A18) Can control the overall type I error $\alpha$ .	(L16) Can be overly conservative as in the SATURN trial because of the correlation between the test of treatment effect in the overall study population and in the biomarker subgroups.
<b>Also called:</b> Overall/biomarker-positive design with parallel assessment, prospective subset design, hybrid design		(A19) Can require smaller sample size as compared to the subgroup-specific designs, especially when we assume that the novel treatment equally benefits both biomarker-defined subgroups.	(L17) Cannot control the probability of rejecting the null hypothesis of no treatment effect in the biomarker-negative subset when the treatment benefit is restricted to biomarker-positive patients. Consequently, there is a high risk of inappropriately recommending the novel treatment for biomarker-negative patients due to the large treatment effect in biomarker-positive subset.

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<b>Examples of actual trials:</b> S0819 [14,49], SATURN [14,36,47,49,95,96], MONET1 [14,49], ARCHER [14,49], ZODIAC [49], MERIDIAN [49]			
<b>Biomarker-positive and overall strategies with sequential assessment</b> (11 papers) [13,14,30,44,49,69,80,84,85,88,94] (see Figure 8)	Might be useful in cases where the experimental treatment is expected to be effective in the overall population.	(A20) Same as (A18), (A19)	(L18) Can be problematic for determining whether the treatment is beneficial in the biomarker-negative subgroup.
<b>Also called:</b> Overall/biomarker-positive design with sequential assessment, sequential design, Fixed-sequence 2 design, hierarchical fixed sequence testing procedure			(L19) Same as (L17)
<b>Examples of actual trials:</b> Trial of letrozole plus lapatinib versus letrozole plus placebo in breast cancer, with the biomarker defined by human epidermal growth factor receptor 2 (HER2) [14], N0147 [30,49]			
<b>Biomarker-positive and overall strategies with fall-back analysis</b> (15 papers) [10,30,36,44,47,49,53,57,60,69,84,88,94,96,97] (see Figure 9)	Recommended when there is insufficient confidence in the predictive value of the biomarker and the novel treatment is assumed to probably benefit all patients.	(A21) Can assess the treatment effect in the biomarker-positive patients, if no benefit is detected in the overall population.	(L20) Same as (L17), (L18)
<b>Also called:</b> Biomarker-stratified design with fall-back analysis, fall-back design, prospective subset design, sequential design, other analysis plan design, Fallback design		(A22) Same as (A18), (A19)	
<b>Examples of actual trials:</b> None identified <sup>a</sup>			
<b>Marker Sequential test design</b> (4 papers) [14,49,69,94] (see Figure 10)	Recommended when biomarkers with strong credentials are available and we have convincing evidence that the novel treatment is more effective in biomarker-positive than in biomarker-negative patients.	(A23) Can provide clear evidence of treatment benefit in the biomarker-positive subgroup and in the biomarker-negative subgroup.	(L21) In situations where biomarker status is not available for some of the patients included in the study, this design can either exclude these patients or include them in the global test, however, further statistical adjustments might be required in that case.
<b>Also called:</b> MaST design, hybrid design	Appropriate when we can assume that the treatment will not be beneficial in the biomarker-negative subpopulation unless it is effective for the biomarker-positive subpopulation.	(A24) Enables sequential testing of the treatment effect in the entire study population and in the biomarker-defined subgroups to restrict testing of the treatment effect in the entire population when there is no significant result in the biomarker-positive subset, while controlling the appropriate type I error rates.	(L22) Does not decrease the sample size of the study as it was developed in order to increase the power compared to the sequential subgroup-specific design in situations where the novel treatment benefits equally both biomarker-negative and biomarker-positive patients.

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<p><b>Examples of actual trials:</b> ECOG E1910 [14,49]</p>		<p>(A25) Results in higher power as compared to the sequential subgroup-specific design in cases where the treatment effect is homogeneous across the biomarker-defined subgroups.</p>	
		<p>(A26) Preserves the power in situations where the treatment effect is restricted only to the biomarker-positive patients and at the same time it controls the relevant type I error rates.</p>	
		<p>(A27) Control the type I error rate for the biomarker-negative subgroup over all possible prevalence values.</p>	
		<p>(A28) The probability of erroneously concluding that the novel treatment is beneficial for the entire population when the global effect is driven by the biomarker-positive patients is minimized since the design only tests the treatment effect in the entire population when no significant effect is detected in the biomarker-positive subgroup.</p>	
<p><b>Hybrid designs</b> (14 papers) [1,13,15,29–31,36,46,48,55,66,84,88,98] (see Figure 11)</p>	<p>Can be used when there is prior evidence indicating that only a particular treatment is beneficial to a biomarker-defined subgroup which makes it unethical to randomize patients with that specific biomarker status to other treatment options.</p>	<p>(A29) The feasibility of a prognostic biomarker can be tested.</p>	<p>None found.</p>
<p><b>Also called:</b> Mixture design, Combination of trial designs, hybrid biomarker design</p>		<p>(A30) Allows for better risk assessment and improved individualized treatment since it assigns patients to treatments based on risk assessment scores instead of their biomarker status (biomarker-positive and biomarker-negative patients).</p>	
<p><b>Examples of actual trials:</b> TAILORx [15,48,55,58,63,66], EORTC MINDACT [15,48,55,66], ECOG 5202 study [30,46]</p>			
<p><b>Biomarker-strategy designs with biomarker assessment in the control arm</b> (21 papers) [15,25,26,32,33,36,45,61,62,64,79,82,85,86,92,93,99–103] (see Figure 12)</p>	<p>Useful when we want to test the hypothesis that the treatment effect based on the personalized approach is superior to that of the standard of care.</p>	<p>(A31) Biomarker can be validated without including all possible biomarker–treatment combinations [26] as in the non-biomarker-based arm all patients receive only the control treatment.</p>	<p>(L23) Unable to inform us whether the biomarker is predictive as these designs are able to answer the question about whether the biomarker-based strategy is more effective than standard treatment, irrespective of the biomarker status of the study population.</p>

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<p><b>Also called:</b> Marker strategy design, Biomarker-strategy design, Strategy design, Marker-based strategy design, Marker-based design, Random disclosure design, Customized strategy design, Parallel controlled pharmacogenetic study design, Marker-based strategy design I, Biomarker-guided design, Biomarker-based assignment of specific drug therapy design, Marker-based strategy I design, Biomarker-strategy design with a standard control, Marker strategy design for prognostic biomarkers</p>		<p>(A32) Have the option of testing the biomarker status of patients in the non-biomarker-strategy arm which can aid secondary analyses [26].</p>	<p>(L24) The evaluation of the true biomarker by treatment effect is not possible as the biomarker-positive patients receive only the experimental treatment and not the alternative treatment (control treatment). Consequently, this design cannot detect the case in which the control treatment might be more beneficial for the entire population.</p>
<p><b>Examples of actual trials:</b> GILT docetaxel [15], Randomized phase III trial conducted in Spain, dedicated to patients with advanced Non-Small Cell Lung Cancer (NSCLC) candidates for first-line chemotherapy [32,64,100], Study the effect of Magnetic Resonance Imaging (MRI) in patients with low back pain on patient outcome and to evaluate Doppler US of the umbilical artery in the management of women with intrauterine growth retardation (IUGR), Randomized controlled trial in recurrent platinum-resistant ovarian carcinoma [101]</p>		<p>(A33) Able to inform us whether the biomarker is prognostic.</p>	<p>(L25) In case that the number of biomarker-positive patients is very small, then the treatment received will be similar in biomarker-strategy arm and non-biomarker strategy arm. Consequently, the trial might give little information regarding the efficacy of the experimental treatment or it might not be able to detect it. As a result, this type of design should be used when there is an adequate number of biomarker-positive and biomarker-negative patients.</p>
		<p>(A34) Can be expanded to investigate several biomarkers and treatments [103]. Additionally, these designs can be attractive when evaluating multiple biomarkers or the predictive value of molecular profiling between several treatment options is to be assessed [45].</p>	<p>(L26) Unable to compare directly experimental treatment to control treatment as the aim is to compare not the treatments but the biomarker-strategies.</p>
		<p>(A35) Might be used more frequently in the future due to the wide variety of molecular biomarkers, complexity of gene expression arrays, and several treatments directed at similar targets [103].</p>	<p>(L27) Less efficient designs than biomarker-stratified designs [4,73] and a poor substitute for clinical trials which aim to compare the experimental treatment to control treatment, since it is possible for some patients in both the biomarker-based strategy arm and non-biomarker-based strategy arm to be assigned to the same treatment (due to the existence of biomarker-negative patients in both strategy arms the treatment effect can be diluted) [51]. Consequently, as a large overlap of patients receiving the same treatment might have occurred, the comparison of the two biomarker-strategy arms results in a hazard ratio which is forced towards unity, i.e., no treatment effect exists as the effect of experimental versus control treatment is diluted by the biomarker-based treatment selection. For this reason, a large sample size is needed to detect at least a small overall difference in outcomes between the two biomarker-strategy arms.</p>
			<p>(L28) Should be used only if you want to evaluate a complex biomarker-guided strategy with a variety of treatment options or biomarker categories [73].</p>

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
<b>Biomarker-strategy design without biomarker assessment in the control arm</b> (14 papers) [9,13,17,18,20,25,36,38,61,74,101,104–106] (see Figure 13)	In situations where it is not feasible or unethical to test the biomarker in the entire population.	(A36) Galanis et al., 2011 [45] stated that these designs can be attractive when evaluating multiple biomarkers or the predictive value of molecular profiling between several treatment options is to be assessed. Also, Freidlin and Korn, 2010 [73] claimed that these biomarker-strategy designs should be used only if researchers want to evaluate a complex biomarker-guided strategy with a variety of treatment options or biomarker categories.	(L29) Criticized for their potential cost increase due to the fact that patients without predicted responsive biomarker are double enrolled in the trial (biomarker-negative patients receive control treatment in both strategy arms).
<b>Also called:</b> Biomarker-strategy design with standard control, Direct-predictive biomarker-based, RCT of testing, Test-treatment, Parallel controlled pharmacogenetic diagnostic study, Marker strategy, Marker-based with no randomization in the non-marker-based arm, Classical, Marker-based strategy, Marker strategy design for prognostic biomarkers		(A37) Same as (A31), (A32), (A33)	(L30) Biomarker-positive and biomarker-negative subpopulations might be more imbalanced as compared with the first type of biomarker-strategy design due to the fact that the randomization to different treatment strategies is performed before the evaluation of the biomarker status (balancing the randomization is useful to ensure that all randomized patients have tissue available). This can happen especially when the number of patients is very small.
<b>Examples of actual trials:</b> A study, which evaluated the use of immediate computed tomography in patients with acute mild head injury [101,104].			(L31) Same as (L23), (L24), (L25), (L26), (L27)
<b>Biomarker-strategy design with treatment randomization in the control arm</b> (17 papers) [15,17,26,27,32,36,45,62,64,66,74,86,92,93,106–108] (see Figure 14)	In cases where we want to know whether the biomarker is not only prognostic but also predictive, these designs are preferable as compared to the two previously mentioned biomarker-strategy designs.	(A38) These designs have the ability to inform researchers about the potential superiority of the control treatment in the whole population or among a particular biomarker-defined subpopulation.	(L32) Generally require a larger sample size as compared to the marker-stratified designs.
<b>Also called:</b> Biomarker-strategy design with a randomized control, Modified marker-based strategy design (for predictive biomarkers), Biomarker-strategy design with randomized control, Marker-based design with randomization in the non-marker-based arm, Marker-based strategy design II, Marker-strategy design, Augmented strategy design, Trial design allowing the evaluation of both the treatment and the marker effect		(A39) Able to inform us whether the biomarker is prognostic or predictive.	(L33) Same as (L27)

Table 1. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Utility	Advantages	Limitations
Examples of actual trials: None identified <sup>a</sup>		(A40) Allow clarification of whether the results which indicate efficacy of the biomarker-directed approach to treatment are caused due to a true effect of the biomarker status or to an improved treatment irrespective of the biomarker status.	
		(A41) Same as (A36)	
<b>Reverse marker-based strategy</b> (4 papers) [86,92,93,109] (see Figure 15)	Enables testing the interaction hypothesis of treatment and biomarker in a more efficient way as compared to the first (i.e., Biomarker-strategy design with biomarker assessment in the control arm) and third biomarker-strategy subtype design (i.e., Biomarker-strategy design with randomization in the control arm and the marker stratified design)	(A42) Can estimate directly the marker-strategy response rate.	(L34) It has been claimed by Baker, 2014 [93] that other designs than the reverse marker-based strategy are more appropriate in order to investigate questions which include both treatment effect of biomarker-defined subgroups and the biomarker strategy treatment effect. These designs should allow the estimation of treatment effects within biomarker-defined subgroups as well as the estimation of the global treatment effect.
Also called: None found		(A43) Allows the estimation of the effect size of the experimental treatment compared to the control treatment for each biomarker-defined subset separately.	
Examples of actual trials: None identified <sup>a</sup>		(A44) There is no chance that the same treatment will be tailored to biomarker-positive patients who are randomized either to the biomarker-based strategy arm or the reverse marker strategy. Also, there is no possibility of the same treatment assignment to biomarker-negative patients who are randomly assigned to the two biomarker-based strategy arms.	
		(A45) It has been demonstrated by Eng, 2014 [92] that this new type of design is more than four times more efficient for testing the interaction between treatment and biomarker compared to Biomarker-strategy design with biomarker assessment in the control arm, Biomarker-strategy design with randomization in the control arm and the marker stratified design.	
<b>A specific randomized phase II trial design that can be used to guide decision making for further development of an experimental therapy.</b> (1 paper) [71] (see Figure 16)	Recommended when we want to conduct a Phase II randomized trial which allows decisions to be made about which type of Phase III biomarker-guided trial should be used.	(A46) Works well in providing recommendations for phase III trial design.	None found

<sup>a</sup> Although not found within the review, the design may be implemented in ongoing trials.

**Table 2.** Sample size formulae for biomarker-guided clinical trial designs.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Sample Size Formula	Definition
<u>Single arm designs</u>	Standard sample size formula can be used, more information can be found in the ‘methodology’ part of the ‘Single arm designs’ section in the main text.	
<u>Enrichment designs</u> [55,61,65,110–112]	Online tool for sample size calculation when using either binary or time-to-event endpoints is available on the following website: <a href="http://brb.nci.nih.gov/brb/samplesize/td.html">http://brb.nci.nih.gov/brb/samplesize/td.html</a> [113].	
	$E(D_{i, enrichment}) = \frac{nT\lambda_i}{2(\lambda_i + \varphi_i)} \left\{ 1 - \frac{e^{-(\lambda_i + \varphi_i)\tau}}{(\lambda_i + \varphi_i)T} \left[ 1 - e^{-(\lambda_i + \varphi_i)T} \right] \right\}$	<p><math>E(D_{i, enrichment})</math> is referred to the expected number of events per treatment arm (time-to-event outcome), <math>i</math> corresponds to either the experimental or the control treatment group, 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>\lambda</math> corresponds to the event hazard rate, <math>\varphi</math> is the loss to follow-up rate, <math>T</math> denotes the accrual time, patients enter the trial according to a Poisson process with rate <math>n</math> per year over the accrual period of <math>T</math> years, <math>\tau</math> corresponds to the follow-up period.</p>
	$D_{enrichment} = 4 \left[ \frac{(z_{\alpha/2} + z_{\beta})}{\log \theta_1} \right]^2$	<p><math>D_{enrichment}</math> is referred to the required total number of events (time-to-event outcome), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>z_{\alpha/2}</math>, <math>z_{\beta}</math> denote the upper <math>\alpha/2</math>- and upper <math>\beta</math>-points respectively of a standard normal distribution, <math>\alpha</math> and <math>\beta</math> denote the assumed type I error and type II error respectively, <math>\theta_1</math> denotes the assumed hazard ratio between the two treatment groups (control vs experimental) in the biomarker-positive subset.</p>
	$N_{enrichment/arm} = 2\bar{p}_Q (1 - \bar{p}_Q) \left[ \frac{(z_{\alpha/2} + z_{\beta})}{(p_A^Q - p_B)} \right]^2$	<p><math>N_{enrichment/arm}</math> is referred to the required number of patients per treatment arm (binary outcome), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>p_A^Q</math> and <math>p_B</math> are the response probabilities in the experimental and control groups respectively, <math>\bar{p}_Q = (p_A^Q + p_B)/2</math>.</p>
	$N_{enrichment/arm} = \frac{2\sigma^2 (z_{\alpha/2} + z_{\beta})^2}{(\mu_{A+} - \mu_{B+})^2}$	<p><math>N_{enrichment/arm}</math> is referred to the required total number of patients per treatment arm (continuous response endpoints), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>\sigma^2</math> denotes the anticipated common variance, <math>\mu_{A+}</math> and <math>\mu_{B+}</math> the mean responses for biomarker-positive patients in the experimental and control treatment arm respectively.</p>
	$N_{enrichment/arm} = 2\sigma^2 (z_{\alpha/2} + z_{\beta})^2 \{ \lambda_1 [(1 - \omega) \zeta + \omega] \}^{-2}$	<p><math>N_{enrichment/arm}</math> is referred to the required total number of patients per treatment arm (continuous response endpoints when accounting for error in the assaying of the study population), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>\omega</math> measures the accuracy of the assay and corresponds to the PPV (positive predictive value of the assay, i.e., the proportion of patients who are assigned biomarker positive status according to the assay who are truly biomarker positive), <math>\lambda_1</math> is the treatment effect in the biomarker-positive patients and <math>\zeta = \lambda_0/\lambda_1</math> (where <math>\lambda_0</math> is the treatment effect in the biomarker-negative patients).</p>

Table 2. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Sample Size Formula	Definition
<b>Marker Stratified designs</b> <a href="#">[31,53,60,92,111,112,114]</a>	Online tool for sample size calculation when using either binary or time-to-event endpoints is available on the following website: <a href="http://brb.nci.nih.gov/brb/samplesize/sdpap.html">http://brb.nci.nih.gov/brb/samplesize/sdpap.html</a> [115].	
	$D_{stratified} = 4 \frac{(z_{\alpha_1} + z_{\beta})^2}{[\log(\theta_1)]^2} + 4 \frac{(z_{\alpha_2} + z_{\beta})^2}{[\log(\theta_2)]^2}$	$D_{stratified}$ is referred to the required total number of events for the achievement of sufficient power in each biomarker-defined subgroup separately (time-to-event endpoint), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, $\theta_2$ corresponds to the hazard ratio of biomarker-negative subgroup, $a_1 = a_2 = a/2$ .
	$D_{stratified} = \frac{4(z_{\alpha/2} + z_{\beta})^2}{[k \log(\theta_1) + (1-k) \log(\theta_2)]^2}$	$D_{stratified}$ is referred to the required total number of events for the achievement of sufficient power in the overall population (time-to-event endpoint), $k$ is the proportion biomarker-positive patients, 1 : 1 ratio between the two treatment arms (experimental:control) is assumed.
	$N_{stratified} = \frac{4(z_{\alpha/2} + z_{\beta})^2}{\{[kPr_{(+)}(event) \log(\theta_1) + (1-k)Pr_{(-)}(event) \log(\theta_2)] / \sqrt{kPr_{(+)}(event) + (1-k)Pr_{(-)}(event)}\}^2}$	$N_{stratified}$ is referred to the required total number of patients for the achievement of sufficient power in the overall population (time-to-event endpoint), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, $Pr_{(+)}(event)$ , $Pr_{(-)}(event)$ are the probabilities of an event in biomarker-positive subset and biomarker-negative subset respectively.
	$\frac{D_{stratified}}{D_{enrichment}} = \frac{[\log(\theta_1)]^2}{[k \log(\theta_1) + (1-k) \log(\theta_2)]^2} = \frac{1}{\left[k + (1-k) \frac{\log(\theta_2)}{\log(\theta_1)}\right]^2}$	$\frac{D_{stratified}}{D_{enrichment}}$ is referred to the ratio of the required number of events between marker stratified and enrichment design (time-to-event endpoint).
	$\frac{N_{stratified}}{N_{enrichment}} \approx \frac{1}{\left[k + (1-k) \frac{\delta_{-}}{\delta_{+}}\right]^2}$	$\frac{N_{stratified}}{N_{enrichment}}$ is referred to the ratio of the required number of patients between marker stratified and enrichment design (binary outcome), $\delta_{-}$ , $\delta_{+}$ correspond to the treatment effectiveness in biomarker-negative and biomarker-positive subgroup respectively.
	$N_{stratified} = 2(z_{\alpha} + z_{1-\beta})^2 \left\{ \frac{r_{A+}(1-r_{A+}) + r_{B+}(1-r_{B+})}{(\beta_A + \beta_I)^2} + \frac{r_{A-}(1-r_{A-}) + r_{B-}(1-r_{B-})}{(\beta_A)^2} \right\}$	$N_{stratified}$ is referred to the required total number of patients (binary outcome), $\beta_0$ denotes a baseline effect, $\beta_A$ denotes the added effect of the experimental treatment, $\beta_+$ denotes the biomarker-positive effect and $\beta_I$ denotes the nonadditive effect, $\alpha$ corresponds to the target level, $1 - \beta$ corresponds to the power, $r_{A+}$ , $r_{B+}$ are the assumed response rates of biomarker-positive patients receiving the experimental and the control treatment respectively, $r_{A-}$ , $r_{B-}$ are the assumed response rates of biomarker-negative patients receiving the experimental and the control treatment respectively.

Table 2. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Sample Size Formula	Definition
<u>Sequential Subgroup-Specific design [57]</u>	$N_{\text{sequential subgroup-specific}}^+ = N_{\text{enrichment}}$	$N_{\text{sequential subgroup-specific}}^+$ is referred to the required number of biomarker-positive patients (binary outcome), $N_{\text{enrichment}}$ is the required number of biomarker-positive patients (binary outcome) in the enrichment design.
	$N_{\text{sequential subgroup-specific}} = \frac{N_{\text{enrichment}}}{k}$	$N_{\text{sequential subgroup-specific}}$ is referred to the required total number of patients (binary outcome), $N_{\text{enrichment}}$ is the required number of biomarker-positive patients (binary outcome) in the enrichment design.
	$N_{\text{sequential subgroup-specific}}^- = \frac{(1-k)N_{\text{enrichment}}}{k}$	$N_{\text{sequential subgroup-specific}}^-$ is referred to the required number of biomarker-negative patients (binary outcome), $N_{\text{enrichment}}$ is the required number of biomarker-positive patients (binary outcome) in the enrichment design.
	$D_{\text{sequential subgroup-specific}}^+ = D_{\text{enrichment}}$	$D_{\text{sequential subgroup-specific}}^+$ is referred to the required number of events for biomarker-positive patients (time-to-event outcome), $D_{\text{enrichment}}$ is the required number of events for biomarker-positive patients (time-to-event outcome).
	$D_{\text{sequential subgroup-specific}}^- = D_{\text{enrichment}} \left( \frac{\lambda_-}{\lambda_+} \right) \left( \frac{1-k}{k} \right)$	$D_{\text{sequential subgroup-specific}}^-$ is referred to the required number of events for biomarker-negative patients (time-to-event outcome), $D_{\text{enrichment}}$ is the required number of events for biomarker-positive patients (time-to-event outcome), $\lambda_-$ , $\lambda_+$ , are the event rates in biomarker-negative and biomarker-positive control subgroups.
<u>Parallel Subgroup-Specific design</u>	Same formula proposed for marker stratified designs could be considered to achieve sufficient power in each biomarker-defined subgroup simultaneously. However, in order to control the overall type I error rate of the design at the overall level of significance $\alpha$ it is required to allocate this overall $\alpha$ between the test for the biomarker-positive subgroup and the test for the biomarker-negative. Consequently, for biomarker-positive subgroup the reduced significance level $\alpha_1 = \alpha - \alpha_2$ can be used whereas the reduced significance level $\alpha_2 = \alpha - \alpha_1$ can be used for biomarker-negative subgroup.	
<u>Biomarker-positive and overall strategies with parallel assessment</u>	If there is significant confidence that the biomarker is predictive, the sample size estimation is aimed at having a sufficient number of biomarker-positive individuals to enable the treatment effect in the biomarker positive subgroup to be detected. Standard formula for sample size calculation of biomarker-positive subgroup proposed for the enrichment designs could be considered by using the reduced significance level $\alpha_1 = \alpha - \alpha_2$ . On the other hand, if there is no confidence in the predictive value of the biomarker, the sample size estimation is aimed at having a sufficient number of patients to detect a treatment effect in the overall study population; consequently, for the sample size calculation, the same formula proposed for marker stratified designs aiming to achieve sufficient power in the overall population could be applied by using the reduced significance level $\alpha_2 = \alpha - \alpha_1$ .	

Table 2. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Sample Size Formula	Definition
<u>Biomarker-positive and overall strategies with sequential assessment</u>	At the first stage, the standard formula for a traditional randomized trial which is the same with the formula proposed for enrichment designs can be applied for the biomarker-positive subgroup. At the second stage, the sample size formula proposed for marker stratified designs aiming to yield appropriate power for the entire population could be considered.	
<u>Biomarker-positive and overall strategies with fall-back analysis</u>	At the first stage, the sample size formula proposed for marker stratified designs aiming to yield appropriate power for the entire population could be considered by using the reduced significance level $a_1 = a - a_2$ . At the second stage, the formula proposed for enrichment designs could be applied for the biomarker-positive subgroup by using the reduced significance level $a_2 = a - a_1$ .	
<u>Marker Sequential test design (MaST)</u>	A standard sample size calculation (i.e., the same sample size calculation as for the enrichment designs) can be applied for the biomarker-positive subpopulation. However, in order to have sufficient number of biomarker-positive patients to detect treatment effectiveness in that particular biomarker-defined subset and consequently to reach the desired power, the sample size should be calculated by using the reduced significance level $a_1 [0, a]$ instead of the global significance level $\alpha$ which is used in the sample size formulae of the enrichment designs. The same formula could be considered for the sample size calculation of the biomarker-negative subgroup; however, the corresponding hazard ratio of that subgroup and the global significance level $\alpha$ should be used. For the sample size calculation of the entire population, the same formula proposed for marker stratified designs aiming to achieve sufficient power in the overall population could be considered by using the reduced significance level $a_2 = a - a_1$ .	
<u>Biomarker-strategy, design with biomarker assessment in the control arm [26,61,92]</u>	$D_{strategy\ 1} = 4 \left[ \frac{(z_{\alpha/2} + z_{\beta})}{k \log \theta_1} \right]^2$	$D_{strategy\ 1}$ is referred to the required total number of events (time-to-event outcome), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed.
	$N_{strategy\ 1} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 (\tau_m^2 + \tau_n^2)}{(v_m - v_n)^2}$	$N_{strategy\ 1}$ is referred to the required total sample size (continuous clinical endpoints), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, $z_{1-\alpha/2}$ , $z_{1-\beta}$ denote the lower $1 - \alpha/2$ - and lower $1 - \beta$ -points respectively of a standard normal distribution, $v_m$ and $v_n$ denote the mean response from the biomarker-based strategy arm and the non-biomarker-based strategy arm respectively, and $\tau_m^2$ , $\tau_n^2$ denote the variance of response for the biomarker-based strategy arm and non-biomarker-based strategy arm respectively.
	$N_{strategy\ 1/arm} = \frac{(z_{\alpha} + z_{1-\beta})^2 [g_1(1-g_1) + g_2(1-g_2)]}{\Delta_2^2}$	$N_{strategy\ 1/arm}$ is referred to the required total number of patients per arm (binary outcome), $g_1$ is the expected response rate in the biomarker-based strategy arm, $g_2$ is the expected response rate in the non biomarker-based strategy arm, $\Delta_2 = g_1 - g_2$ , $g_1, g_2$ can be found by calculating the formulae $kr_{A+} + (1 - k)r_{B-}$ and $r_B$ respectively, $r_B$ denotes the marginal effect of treatment B (control treatment).

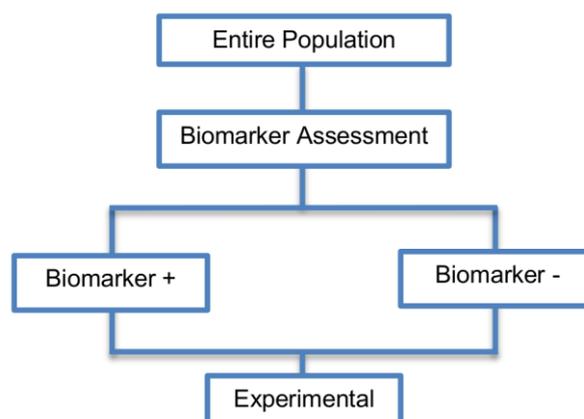
Table 2. Cont.

Types of Biomarker-Guided Non-Adaptive Trial Designs	Sample Size Formula	Definition
<b>Biomarker-strategy design without biomarker assessment in the control arm</b>	Same formulae as for the ‘Biomarker-strategy design with biomarker assessment in the control arm’ can be considered.	
<b>Biomarker-strategy design with treatment randomization in the control arm [26,31,92]</b>	$D_{strategy\ III} = \frac{4(z_{\alpha/2} + z_{\beta})^2}{\left\{ \log \left[ \frac{2km_{B+} + 2(1-k)m_{A-}}{k(m_{A+} + m_{B+}) + (1-k)(m_{A-} + m_{B-})} \right] \right\}^2}$	<p><math>D_{strategy\ III}</math> is referred to the required total number of events (time-to-event outcome), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>m_{A+}, m_{A-}, m_{B+}, m_{B-}</math>, denote the median survival for biomarker-positive and biomarker-negative patients receiving control and experimental treatments respectively.</p>
	$N_{strategy\ III} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 (\tau_m^2 + \tau_{nr}^2)}{(v_m - v_{nr})^2}$	<p><math>N_{strategy\ III}</math> is referred to the required total sample size (continuous clinical endpoints), 1 : 1 ratio between the two treatment arms (experimental:control) is assumed, <math>v_{nr}</math> denotes the mean response from the non-biomarker-based strategy arm, <math>\tau_{nr}^2</math> denotes the variance of response for the non-biomarker-based strategy arm respectively.</p>
	$N_{strategy\ III/arm} = \frac{(z_{\alpha} + z_{1-\beta})^2 [g_1(1-g_1) + g_3(1-g_3)]}{\Delta_3^2}$	<p><math>N_{strategy\ III/arm}</math> is referred to the required total number of patients per arm (binary outcome), <math>g_3</math> is the expected response rate in the non biomarker-based strategy arm and <math>\Delta_3 = g_1 - g_3</math>, the expected response rate <math>g_3</math> can be found by calculating the formula <math>r_A/2 + r_B/2</math>, <math>r_A</math> denotes the marginal effect of treatment A (experimental treatment).</p>
<b>Reverse marker-based strategy [92]</b>	$N_{strategy\ IV/arm} = \frac{(z_{\alpha} + z_{1-\beta})^2 [g_1(1-g_1) + g_4(1-g_4)]}{\Delta_4^2}$	<p><math>N_{strategy\ IV/arm}</math> is referred to the required total number of patients per arm (binary outcome), <math>g_4</math> is the expected response rate in the reverse biomarker-based strategy arm and <math>\Delta_4 = g_1 - g_4</math>, the expected response rate <math>g_4</math> can be found by calculating the formula <math>kr_{B+} + (1-k)r_{A-}</math>, <math>r_{B+}, r_{A-}</math> are the assumed response rates of biomarker-positive patients receiving the control treatment and biomarker-negative patients receiving the experimental treatment.</p>
<b>Randomized Phase II trial design with biomarkers [71]</b>	Online tool for sample size calculation is available on the following website: <a href="http://brb.nci.nih.gov/Data/FreidlinB/RP2BM">http://brb.nci.nih.gov/Data/FreidlinB/RP2BM</a> [116].	

### 2.1. Single Arm Designs

Single arm designs were referred to in seven papers (7%). In the context of biomarkers, these designs (Phase II designs) include the whole study population to which the same experimental treatment is prescribed, without taking into consideration biomarker status.

**Design:** In this design all patients are prescribed the experimental treatment and there is no comparison with a control treatment. These trial designs aid in the identification of association between biomarker status and the efficacy or safety of the experimental treatment. An illustration of this approach is shown in Figure 2.



**Figure 2.** Single arm designs.

**Utility:** These designs can be useful for the initial identification and/or validation of a biomarker and their aim is not to estimate the treatment effect in a definitive way but to identify whether the biomarker is sufficiently promising to proceed to a definitive Phase III biomarker-guided randomized controlled trial.

**Methodology:** In single arm designs first we assess the biomarker status of patients and then as all patients will be treated the same way we could compare the outcome of the biomarker-positive subgroup with the outcome of biomarker-negative subgroup. According to Tajik et al., 2012 [117], in terms of the required sample size, a standard formula can be used, however one should take into consideration the multiple testing issue that arise due to the exploration of several prognostic biomarkers (e.g., Bonferroni adjustment or normal exact method to protect against type I error  $\alpha$  for multiple tests are often considered [118]). Further information can be found in the paper of Zaslavsky and Scott, 2012 [118] who studied the sample size estimation in single arm clinical trials with multiple testing under frequentist and Bayesian framework.

**Statistical considerations:** The single arm approach can be considered as a simple statistical design as there is no need for randomization. However one limitation of this strategy is that there is no distinction between prognostic and predictive biomarkers i.e., as patients are not randomized to experimental and control treatment groups, it is not possible to determine whether an observed effect is attributable to the natural disease progression or to the treatment. Consequently, this study designs are unable to show the benefit of a biomarker with regard to the best choice of treatment.

### 2.2. Enrichment Designs

Enrichment designs are described in 71 papers (71%), either in Phase II or Phase III clinical trials, and involve randomizing only the biomarker-positive patients and comparing the experimental treatment versus the standard treatment only in this particular biomarker-defined subgroup.

**Design:** Figure 3 graphically represents the trial design. First, the entire population is screened in order to identify the biomarker status of each patient. Next, the random assignment of individuals

to different treatment arms is restricted only to the biomarker-positive subgroup. More precisely, biomarker-negative patients are excluded from the study and consequently, the assessment of the effectiveness of the experimental treatment is limited to the biomarker-positive subgroup. Thus, other patients apart from the biomarker-positive subpopulation can receive only the standard treatment (i.e., control treatment), but they are not included in the investigation during the trial design. The biomarker in this design is referred to as either the ‘selection’ or ‘enrichment’ biomarker.

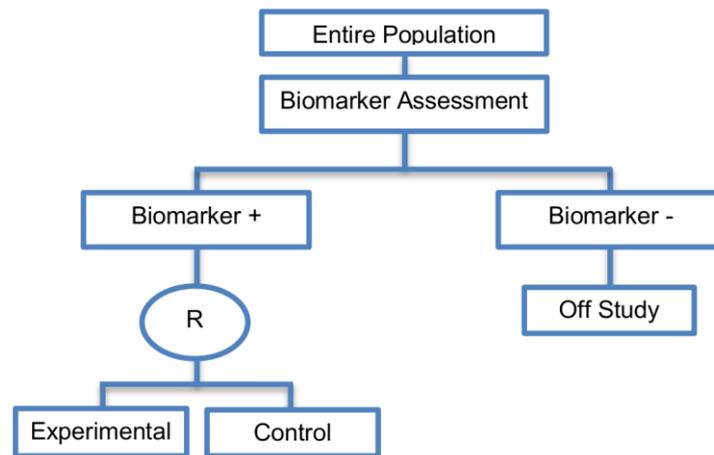


Figure 3. Enrichment designs. “R” refers to randomization of patients.

**Utility:** Enrichment designs are useful for clinical trials aiming to test the treatment effect in a specific biomarker-defined subpopulation where there is evidence to suggest that effectiveness is limited to those within that subgroup, but the candidate biomarker still requires prospective validation. This design is recommended when both the cut-off point for determination of biomarker status of patients and the analytical validity of the biomarker have been well established. A rapid turnaround time for assessing the biomarker status of a patient is also needed to avoid any delay in treatment initiation. This strategy is particularly useful where it is unethical to randomize the biomarker-negative population into different treatment arms, for example where there is prior evidence that the experimental treatment is not beneficial for biomarker-negative individuals, or is likely to cause them harm. However, when it remains unclear whether or not biomarker-negative individuals will benefit from the novel treatment, the enrichment design is not appropriate and alternative designs, which also assess effectiveness in the biomarker-negative individuals, should be considered (e.g., randomize-all designs).

**Methodology:** An online tool has been developed by Zhao and Simon [19,28,53,57,60] that allows sample size planning for the enrichment design both for binary and time-to-event (survival) outcomes, and is available at <http://brb.nci.nih.gov/brb/samplesize/td.html> [113]. For the purpose of estimating the sample size in the case of a survival outcome, data are simulated based on a marker stratified design (see next section for further information) in which both biomarker-positive and biomarker-negative subgroups are investigated in the study and formulae for the enrichment design described in the paper of Rubinstein et al., 1981 [110] are used. Furthermore, an exponential distribution of survival for the experimental and control treatment groups within both the biomarker-positive and biomarker-negative subpopulations is assumed. More precisely, Rubinstein et al. provide the formula of the expected number of events per treatment group allowing to include exponential loss to follow-up given the following assumptions: (i) patients enter the trial according to a Poisson process and patient entry times will be independent and identically distributed uniformly over  $[0, T]$  where  $T$  denotes the accrual time. Consequently, given the total number of patients  $N$ , the times from entry to the end of the trial will be independent and identically distributed uniformly over  $[\tau, T + \tau]$ , where  $\tau$  denotes the follow-up time and  $T + \tau$  the total duration of the study and (ii) 1:1 randomization

between experimental and control treatment group is considered. The expected number of events per treatment arm according to Rubinstein et al. is given by

$$E(D_{i, enrichment}) = \frac{nT\lambda_i}{2(\lambda_i + \varphi_i)} \left\{ 1 - \frac{e^{-(\lambda_i + \varphi_i)t}}{(\lambda_i + \varphi_i)T} [1 - e^{-(\lambda_i + \varphi_i)T}] \right\}, \tag{1}$$

where  $i$  corresponds to either the experimental or the control treatment group,  $\lambda$  corresponds to the event hazard rate,  $\varphi$  is the loss to follow-up rate and patients enter the trial according to a Poisson process with rate  $n$  per year over the accrual period of  $T$  years. However, the required total number of events in the two treatment groups (experimental and control treatment group) is given by

$$D_{enrichment} = 4 \left[ \frac{(z_{\alpha/2} + z_{\beta})}{\log \theta_1} \right]^2, \tag{2}$$

where  $\theta_1$  denotes the assumed hazard ratio between the two treatment groups (control vs. experimental) in the biomarker-positive subset and the constants  $z_{\alpha/2}$ ,  $z_{\beta}$  denote the upper  $\alpha/2$ - and upper  $\beta$ -points respectively of a standard normal distribution where  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively. Freidlin et al., 2010 [61] provided the aforementioned formula assuming that all random assignments use 1:1 randomization. As in a traditional randomized controlled trial, if the randomization is not equal, i.e., the ratio of allocation to treatment and control is  $R : 1$  rather than 1:1, the aforementioned formula for the required total number of events  $D_{enrichment}$  which assumes 1:1 randomization can be multiplied by  $(R + 1)^2/4R$  [119]. Consequently, the “4” in the formula of  $D_{enrichment}$  becomes  $(R + 1)^2/R$  and the corresponding formula for the total number of events becomes

$$D_{enrichment} = \frac{(R + 1)^2}{R} \left[ \frac{(z_{\alpha/2} + z_{\beta})}{\log \theta_1} \right]^2. \tag{3}$$

In a survival study, the calculation of the total sample size in terms of number of patients required in the two treatment groups (experimental and control treatment group) to be enrolled in order to yield the aforementioned total number of events depends on the probability of event over the duration of the study [120]. Consequently, the actual number of patients required in a survival study can be given by

$$N_{enrichment} = \frac{D_{enrichment}}{\Pr(event)}, \tag{4}$$

where  $\Pr(event)$  is the probability of observing an event in the two treatment groups in the study and  $D_{enrichment}$  is the required total number of events.  $\Pr(event)$  in a survival study can be given by

$$\Pr(event) = \pi_A Pr_A(event) + \pi_B Pr_B(event), \tag{5}$$

where

$$\pi_A = \frac{R}{R + 1} \text{ and } \pi_B = \frac{1}{R + 1}, \tag{6}$$

are the proportions of patients who are randomized to experimental and control treatment group respectively and  $Pr_A(event)$  and  $Pr_B(event)$  are the probability of events in experimental and control arm respectively [121]. Freedman, 1982 [122] provided an approximation of the probability of event for each treatment group assuming equal follow-up for all patients and thus simultaneous accrual for all patients whereas Schoenfeld, 1983 [123] provided a more exact approximation of the expected event rate as compared to Freedman’s approximation. More precisely, according to Freedman’s idea,

$$Pr_i(event) \approx 1 - S_i(\tau) \tag{7}$$

and according to Schoenfeld’s idea,

$$Pr_i(event) \approx 1 - \{S_i(\tau) + 4S_i(T/2 + \tau) + S_i(T + \tau)\}/6, \tag{8}$$

where  $i$  denotes the corresponding treatment group (either experimental or control),  $\tau$  denotes the follow-up time and  $T$  the accrual period,  $T/2 + \tau$  denotes the median follow-up time and  $T + \tau$  denotes the total duration of the study. Another approximation of the probability of event could be

$$Pr_i(event) \approx 1 - S_i(T/2 + \tau) \tag{9}$$

considering that the survival probability can be approximated as the probability that a patient survives past the median follow-up time (i.e.,  $T/2 + \tau$ ) [121].

The web-based interface is composed of two options. If the first option is chosen, the treatment effects for assay-negative and assay-positive patients must be specified in order to evaluate the relative efficiency of enrichment and untargeted design, i.e., marker stratified design (see next section for further information) in which apart from the biomarker-positive patients, biomarker-negative patients are also included; if the second option is chosen, it is possible to account for error in the assaying of the study population, thus, both the treatment effects for target-negative and target-positive patients must be specified as well as the assay’s sensitivity and specificity.

The sample size calculation using binary data is based on the formulas described by Simon and Maitournam [65,111,112] and again the two options offered when assuming a time-to-event outcome are available, i.e., options both with and without accounting for error in assaying the study population the biomarker status. When binary outcome is assumed and the allocation ratio is 1:1, the sample size of randomized patients required in each treatment arm (experimental and control) can be given as

$$N_{enrichment/arm} = 2\bar{p}_Q(1 - \bar{p}_Q) \left[ \frac{(z_{\alpha/2} + z_{\beta})}{(p_A^Q - p_B)} \right]^2, \tag{10}$$

where  $p_A^Q$  and  $p_B$  are the response probabilities in the experimental and control groups respectively,

$$\bar{p}_Q = \frac{p_A^Q + p_B}{2} \tag{11}$$

and  $z_{\alpha/2}$ ,  $z_{\beta}$  denote the upper  $\alpha/2$ - and upper  $\beta$ -points respectively of a standard normal distribution where  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively. The response probability in the experimental group can be found by

$$p_A^Q = p_B + \delta_+, \tag{12}$$

where  $\delta_+$  denotes the improvement in response probability for biomarker-positive patients. Consequently, the total sample size of randomized patients will be

$$N_{enrichment} = 2N_{enrichment/arm} \tag{13}$$

For continuous response endpoints the aforementioned formula  $N_{enrichment/arm}$  changes to

$$N_{enrichment/arm} = \frac{2\sigma^2(z_{\alpha/2} + z_{\beta})^2}{(\mu_{A+} - \mu_{B+})^2}, \tag{14}$$

where  $\sigma^2$  denotes the anticipated common variance,  $\mu_{A+}$  and  $\mu_{B+}$  the mean responses for biomarker-positive patients in the experimental and control treatment arm respectively. These formulae are the standard formulae used for a standard randomized trial.

In addition, if we want to account for error in the assaying of the study population, the number of patients to be randomized in each arm of the enrichment trial when using continuous response endpoints can be given by the following formula

$$N_{enrichment/arm} = 2\sigma^2(z_{\alpha/2} + z_{\beta})^2 \{ \lambda_1 [(1 - \omega)\zeta + \omega] \}^{-2} \tag{15}$$

where  $\omega$  measures the accuracy of the assay and corresponds to the PPV (positive predictive value of the assay, i.e., the proportion of patients who are assigned the biomarker-positive status according to the assay who are truly biomarker positive),  $\lambda_1$  is the treatment effect in the biomarker-positive patients and  $\zeta = \lambda_0/\lambda_1$  (where  $\lambda_0$  is the treatment effect in the biomarker-negative patients) [55].

Simon and Maitournam [65,111,112] considered that apart from the number of patients to be randomized, the number of patients needed to be screened should be also reported. Thus, they stated that the expected number of patients to be screened in the enrichment design is  $N_{enrichment}/k$  where  $k$  corresponds to the proportion of biomarker-positive patients. The online tool developed by Zhao and Simon provides both the number of patients to be screened and to be randomized.

**Statistical considerations:** Simon and Maitournam [65,111,112] undertook a simulation study, assuming a binary outcome, to compare power of the enrichment design with an untargeted design (i.e., marker stratified design, see next section for further information) in which all patients are randomized without measuring the biomarker. They concluded that the efficiency of the enrichment design relies both on the prevalence of the biomarker-positive patients and on the accuracy of the assay. Whilst in the situation where the assay cut-off point is not well established, there is a risk of severely compromising the power of the trial when using an enrichment design, if fewer than half of the entire study population are biomarker-positive and there is robust evidence that the experimental treatment does not benefit the biomarker-negative patients, the required number of randomized patients to allow sufficient power to detect a significant treatment effect is much smaller in the enrichment design than in the untargeted trial design. However, in the latter situation a greater number of individuals would need to be screened when using the enrichment design, and accruing the required number of biomarker positive patients could take a longer period of time. More precisely, Simon and Maitournam showed that an approximation of the ratio of the required number of patients to be randomized for the untargeted trial design as compared with the required number of patients randomized in the enrichment design when using binary outcome can be given by the following equation

$$\frac{N_{stratified}}{N_{enrichment}} \approx \frac{1}{[k + (1 - k)\frac{\delta_-}{\delta_+}]^2} = \left[ \frac{\delta_+}{k\delta_+ + (1 - k)\delta_-} \right]^2, \tag{16}$$

where  $k$  denotes the proportion of biomarker-positive patients,  $\delta_-$  and  $\delta_+$  correspond to the treatment effectiveness (i.e., improvement in response probability) in biomarker-negative and biomarker-positive subgroups respectively. Consequently, in the situation where it is known that the novel treatment does not benefit the biomarker-negative patients at all, the ratio of the number of patients needed for randomization in the untargeted design relative to the number of patients required for the enrichment design is approximately

$$\frac{N_{stratified}}{N_{enrichment}} \approx \frac{1}{k^2}, \tag{17}$$

as  $\delta_- = 0$ . For example, if half of patients are biomarker-positive ( $k = 0.5$ ) then a quarter of those needed to be randomized to the untargeted design trial would need to be randomized to the enrichment

design trial. In cases where the novel treatment is half as effective in biomarker-negative patients as in the biomarker-positive patients (i.e.,  $\delta_- / \delta_+ = 1/2$ ), the aforementioned ratio changes to

$$\frac{N_{stratified}}{N_{enrichment}} \approx \frac{4}{(k + 1)^2} \tag{18}$$

### 2.3. Randomize-All Designs

Randomize-all designs (also named as all-comers/untargeted/unselected/non-targeted/simple randomization designs) allow the inclusion of the entire population as eligible for randomization. Consequently, the whole study population who meet the eligibility criteria, is randomly assigned to the different treatment groups (experimental and control treatment group) regardless of biomarker status. This design allows assessment of treatment benefit for the entire population irrespective of biomarker status whilst at the same time allowing for treatment benefit to be tested in the two biomarker-defined subgroups separately.

Generally, they are useful when we are uncertain about the benefit of the experimental treatment in the overall population versus the biomarker-defined subgroups, the targeted treatment may benefit both biomarker-positive and biomarker-negative patients, the goal is to test the predictive ability of a biomarker, the assay reproducibility and accuracy is questionable, the turnaround time for biomarker assessment is long and the biomarker prevalence is high.

Randomize-all designs are composed of two main subtypes: the Marker-stratified designs and the Hybrid designs, which are discussed separately below.

#### 2.3.1. Marker Stratified Designs

These designs (prospective validation Phase III trials) were identified in 45 papers (45%) of our review.

**Design:** An illustration of the design is shown in Figure 4. Individuals are stratified into biomarker-positive and biomarker-negative subgroups according to the results of the biomarker assessment and then they are randomized either to the experimental or to the control treatment group. The biomarker status in the Marker-Stratified design acts as a stratification factor where stratification is used to ensure balance across treatment groups with regard to biomarkers. Only individuals with valid biomarker results enter the trial. Consequently, we have four treatment groups, i.e., biomarker-positive patients assigned to either the experimental treatment arm or the control treatment arm and biomarker-negative patients assigned to either the experimental treatment arm or the control treatment arm. Thus, we can assess the relationship between treatment effect and biomarker status.

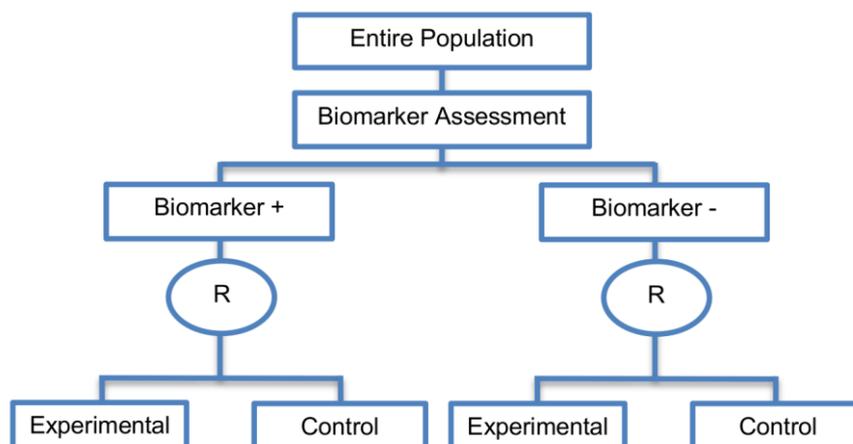


Figure 4. Marker Stratified designs. “R” refers to randomization of patients.

**Utility:** When there is enough evidence that the experimental treatment is more effective in the positive biomarker-defined subgroup than in the negative biomarker-defined subgroup but there is no sufficient compelling data that the experimental treatment is of no benefit in biomarker-negative individuals, the marker stratified design can be used.

**Methodology:** Biomarker status is used to stratify the randomization, rather than to restrict eligibility. Marker-stratified designs can be conducted using two different testing plans; the so-called marker-by-treatment interaction with separate tests and marker-by-treatment interaction with interaction test. Both of these approaches involve conducting two independent clinical trials.

Marker-by-treatment interaction using separate test was referred to in 15 papers (15%) of our review [4,11,12,15,29,42,45,53,57,60,80,82,84,87,88] and is also referred to as ‘separate randomization design’ and ‘separate by treatment interaction design’. This analysis plan is based on separate superiority tests in each biomarker-defined subgroup in order to detect the treatment efficacy in each subset. Two examples of actual trials which use this testing plan are the following: National Cancer Institute (NCI)-sponsored North Central Cancer Treatment Group Study N0975 [29] and the MARVEL trial [29].

The ‘marker-by-treatment interaction design using separate tests’ is a testing plan which determines whether the novel treatment is superior to the control treatment separately within each biomarker-defined subgroup. Consequently, the hypothesis to be tested, the calculation of the number of patients required for the trial, the estimation of the statistical power of the design and the randomization procedure of patients to different treatments are independent among the different subgroups [12]. The sample size of the trial should be calculated in such a way so as to yield adequate statistical power when testing whether the experimental treatment is superior to the control treatment separately in the two biomarker-defined subgroups. Hence, this approach is not widely used due to the required large sample size as essentially two separate trials are being conducted. Another limitation of this approach is that when multiple biomarker-defined subsets and treatments are to be investigated, it is difficult to implement in practice.

The ‘marker-by-treatment interaction using interaction test’ uses a test for interaction between the biomarker status and treatment assignment and was identified in 12 papers (12%) of our review [4,12,15,42,53,57,60,82,84,87,88,94]. A marker stratified design which uses this testing plan is also referred to in the literature as an ‘interaction design’ or ‘genomic signature stratified design’. First, a formal statistical test for interaction between biomarker status and treatment assignment is undertaken. If this interaction is not significant, then the study is continued by testing the different treatments overall at a two-sided significance level of 0.05, otherwise, the treatments are compared within each biomarker-defined subpopulation at a two-sided 0.05 significance level (i.e., the same as in the marker-by-treatment interaction design using separate tests). The sample size for this second testing plan is calculated with reference to the treatment effect in the entire study population. Therefore, it might not provide sufficient power for detecting the treatment effect in each biomarker defined-subset individually. More precisely, if the sample size is calculated for the overall analysis and the proportion of the biomarker-defined subpopulation which responds to the novel treatment is very small, the statistical power for the subgroup analysis may be inadequate. In addition, when several biomarker-defined subpopulations and treatments are to be investigated, this strategy is not easy to be implemented.

For the case of binary outcomes, Eng, 2014 [92] provided the formula for the required sample size to power the biomarker-positive and biomarker-negative patients separately. It is assumed that  $Y$  is a binary variable which corresponds to a patient’s response to their randomly tailored treatment and  $P(Y|Trt = i, M = j) = r_{ij}$  where  $i$  corresponds to either the experimental or control treatment and  $j$  corresponds to either the biomarker-positive patients or the biomarker-negative patients. Hence,

$$r_{ij} = \beta_0 + \beta_A I(Trt = A) + \beta_+ I(M = M^+) + \beta_I I(Trt = A, M = M^+), \quad (19)$$

where  $\beta_0$  denotes a baseline effect,  $\beta_A$  denotes the added effect of the experimental treatment,  $\beta_+$  denotes the biomarker-positive effect and  $\beta_I$  denotes the nonadditive effect. Consequently, the proposed formula for the required sample size can be given by

$$N_{stratified} = 2(z_\alpha + z_{1-\beta})^2 \left\{ \frac{r_{A+}(1-r_{A+}) + r_{B+}(1-r_{B+})}{(\beta_A + \beta_I)^2} + \frac{r_{A-}(1-r_{A-}) + r_{B-}(1-r_{B-})}{(\beta_A)^2} \right\}, \quad (20)$$

where  $\alpha$  correspond to the target level,  $1 - \beta$  corresponds to the power. Also,  $r_{A+}$ ,  $r_{B+}$  are the assumed response rates of biomarker-positive patients receiving the experimental and the control treatment respectively. Additionally,  $r_{A-}$ ,  $r_{B-}$  are the assumed response rates of biomarker-negative patients receiving the experimental and the control treatment respectively.

Mandrekar and Sargent, 2009 [31] provide a formula to calculate the required number of events when the trial has a survival outcome with 1:1 randomization to treatment arms, i.e.,

$$D_{stratified} = \frac{4(z_{\alpha/2} + z_\beta)^2}{\left[ \log\left(\frac{m_{A+}}{m_{B+}}\right) \right]^2} + \frac{4(z_{\alpha/2} + z_\beta)^2}{\left[ \log\left(\frac{m_{A-}}{m_{B-}}\right) \right]^2}, \quad (21)$$

where  $m_{A+}$ ,  $m_{A-}$ ,  $m_{B+}$ ,  $m_{B-}$ , indicate the median overall survival for biomarker-positive and biomarker-negative patients receiving control and experimental treatment, respectively and

$$\theta_1 = \frac{m_{A+}}{m_{B+}} = HR_{biom+}, \quad (22)$$

$$\theta_2 = \frac{m_{A-}}{m_{B-}} = HR_{biom-}, \quad (23)$$

correspond to the hazard ratios of biomarker-positive and biomarker-negative subgroups and  $z_{\alpha/2}$ ,  $z_\beta$  denote the upper  $\alpha/2$ - and upper  $\beta$ -points respectively of a standard normal distribution where  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively. More precisely, the total number of events is the sum of the required number of events for the biomarker-negative and biomarker-positive subpopulation. Freidlin et al., 2010 [61] stated that the required number of events in order to compare the experimental to the control treatment among the biomarker-positive patients for detecting a given effect size in this biomarker-positive subpopulation is identical to the number of events needed by an enrichment design (i.e.,  $D_{enrichment}$ ).

Another potential formula for the required total number of events when 1:1 randomization to treatment arms is assumed is given by

$$D_{stratified} = \frac{4(z_{\alpha/2} + z_\beta)^2}{[k \log(\theta_1) + (1-k) \log(\theta_2)]^2}. \quad (24)$$

Although the formula proposed by Mandrekar and Sargent, 2009 [31] achieves a specific power  $(1 - \beta)$  for each biomarker-defined subgroup separately, the aforementioned formula proposed in the book of Harrington, 2012 [114] aims to reach a power  $(1 - \beta)$  for the overall population. According to Harrington, 2012 the required total number of patients to be entered to a stratified trial can be given by

$$N_{stratified} = \frac{4(z_{\alpha/2} + z_\beta)^2}{\left\{ \frac{[kPr_{(+)}(event) \log(\theta_1) + (1-k)Pr_{(-)}(event) \log(\theta_2)]}{\sqrt{kPr_{(+)}(event) + (1-k)Pr_{(-)}(event)}} \right\}^2}, \quad (25)$$

where  $Pr_{(+)}(event)$ ,  $Pr_{(-)}(event)$  are the probability of an event in biomarker-positive subset and biomarker-negative subset respectively. If we divide the required total number of events for the

enrichment design by the aforementioned formula for the required total number of events for the stratified design, we can get the following approximation of the ratio

$$\frac{D_{stratified}}{D_{enrichment}} = \frac{[\log(\theta_1)]^2}{[k \log(\theta_1) + (1 - k) \log(\theta_2)]^2} = \frac{1}{\left[k + (1 - k) \frac{\log(\theta_2)}{\log(\theta_1)}\right]^2}. \quad (26)$$

Further, Zhao and Simon [19,28,53,57,60] have developed an online tool for the calculation of sample size for biomarker stratified randomized designs with binary or time-to-event endpoints which is available online at the following web site <http://brb.nci.nih.gov/brb/samplesize/sdpap.html> [115]. More precisely, the sample size for both binary and time-to-event endpoints can be performed with three different analysis plans; A, B and C. Before choosing one of these analysis plans in the web site, for binary endpoints we need to specify the probability of treatment response in the control arm as well as the proportion of biomarker-positive patients. For survival endpoints, the hazard ratio of biomarker-positive patients versus the biomarker-negative control patients which corresponds to the hazard ratio of prognostic effect as well as the proportion of biomarker-positive patients must be specified.

Analysis plan A is performed when there is confidence that an overall treatment effect exists. It determines the sample size on the basis of first of all comparing the experimental treatment to the control treatment in the entire randomized population at a reduced two-sided significance level  $\alpha < 0.05$ . If the overall test is not significant, then the experimental treatment is compared to the control treatment in the biomarker-positive patients using the type I error  $\alpha = 0.05$ . Analysis Plan A is similar to the ‘Biomarker-positive and overall strategies design’ with fall-back analysis described later in this paper; the difference lies in this in terms of the significance levels they have used. In order for the sample size to be estimated, the anticipated overall effect estimate, reduced two-sided significance level and power for the overall test need to be specified.

Analysis plan B is performed when there is confidence that there is a treatment effect in the biomarker-positive subpopulation. It determines the sample size on the basis of first of all comparing the experimental treatment to the control treatment in the biomarker-positive subgroup at a two-sided significance level of  $\alpha = 0.05$  level. If the treatment effect is found to be significant at this 0.05 level, then treatment effect is evaluated in the biomarker-negative subgroup again at a two-sided significance level of 0.05 level. This analysis plan is identical to the ‘Sequential subgroup specific design’ described later in this paper. In order for the sample size to be estimated, apart from the fixed significance level set to 0.05, the anticipated effect estimate in the biomarker-positive subpopulation and power need to be specified.

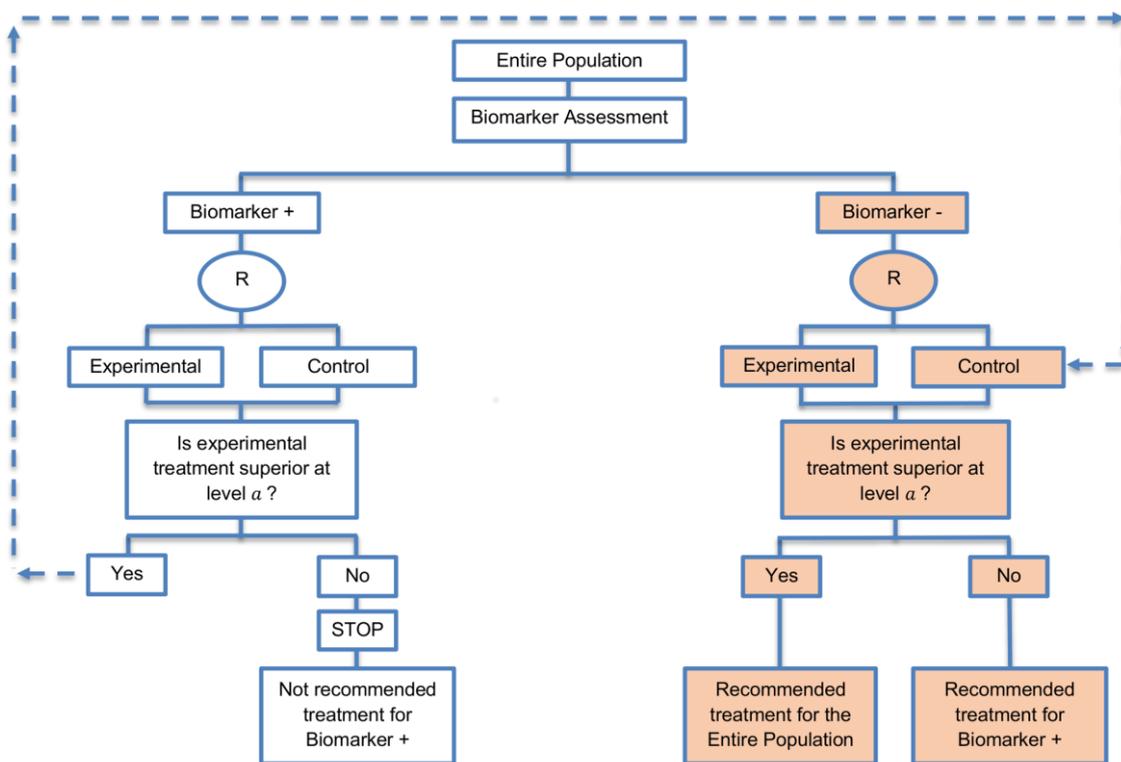
Analysis plan C first tests whether there is a statistically significant interaction between treatment and biomarker [60] at a significance level  $\alpha \leq 0.05$ . If the interaction is not significant, then the treatments are compared in the overall study population at a two-sided significance level 0.05. Otherwise, the treatments are compared within the two biomarker subgroups separately at a two-sided 0.05 significance level for each subgroup. Analysis Plan C follows either the ‘marker-by-treatment interaction process with interaction or the separate test process’ described above. In order for the sample size to be estimated, the anticipated treatment effect in the overall study population, the one-sided significance level for interaction test and the power for testing the treatment effect in the overall population need to be specified.

In marker stratified designs, three designs can be included which differ in terms of their statistical testing strategies, i.e., (i) Subgroup-specific designs (i.e., sequential subgroup-specific design, parallel subgroup-specific design); (ii) Biomarker-positive and overall strategies (i.e., biomarker-positive and overall strategies with parallel assessment, biomarker-positive and overall strategies with sequential assessment, biomarker-positive and overall strategies with fall-back analysis); (iii) Marker sequential test design (MaST) and they are discussed in the following sections.

**Statistical considerations:** Despite the fact that the marker stratified designs allow testing the treatment effect not only in the entire population but also in each biomarker-defined subpopulation, they might not be feasible when the prevalence of biomarker is low. Another limitation of such designs is that they might require a large sample size where several treatments and biomarkers are investigated in the study.

**Subgroup-Specific designs:** This strategy is an approach to analyze a biomarker-stratified trial. It is composed of two types; ‘Sequential Subgroup-Specific design’ and ‘Parallel Subgroup Specific design’. Both biomarker-positive and biomarker-negative subgroups can be tested in a sequential or in a parallel way. With the parallel way, we can assess simultaneously both biomarker-positive and biomarker-negative patients, whereas, with the sequential way we perform first the assessment of biomarker-positive patients and if the result is positive then we continue with the biomarker-negative patients.

**Sequential Subgroup-Specific design:** This approach was referred to in 11 papers (11%) of our review. Figure 5 graphically represents this approach.



**Figure 5.** Sequential Subgroup-Specific design. “R” refers to randomization of patients. Uncolored boxes are referred to the first stage of the trial and colored boxes are referred to the second stage of the trial.

**Design:** The sequential testing procedure uses the assumption that it is unlikely that the new treatment will be effective in the biomarker-negative patients unless it is effective in the biomarker-positive patients. First treatment effect is tested in the biomarker-positive subpopulation using the overall two-sided significance level  $\alpha = 0.05$  (Type I error); if this test is significant then treatment effect is tested in the biomarker-negative subgroup using the same level of significance  $\alpha$ .

**Utility:** Its use is recommended when there is compelling evidence that biomarker-positive individuals benefit more from the experimental treatment than the biomarker-negative patients. More precisely, it is appropriate when it is not expected for the novel treatment to be effective in biomarker-negative patients unless it is beneficial for the biomarker-positive patients.

**Methodology:** As this subgroup-specific design follows a sequential assessment and thus the design is composed of two stages, the sample size calculation is also staged. For binary outcome the required number of biomarker-positive patients is the same as for the enrichment design, i.e.,

$$N_{\text{Sequential subgroup-specific}}^+ = N_{\text{enrichment}} \tag{27}$$

As Simon, 2008 [60] stated, the total number of patients will be approximately

$$N_{\text{Sequential subgroup-specific}} = \frac{N_{\text{enrichment}}}{k} \tag{28}$$

where  $k$  is the proportion of biomarker-positive patients and the number of biomarker-negative patients will be approximately

$$N_{\text{Sequential subgroup-specific}}^- = \frac{(1 - k)N_{\text{enrichment}}}{k}. \tag{29}$$

For the conduct of this design, it is important to ensure that there is also an adequate number of biomarker-negative patients for analysis purposes. For time-to-event outcomes, the required number of events for biomarker-positive patients is the same with the required number of events in the enrichment design, i.e.,

$$D_{\text{Sequential subgroup-specific}}^+ = D_{\text{enrichment}}. \tag{30}$$

At the time that there are  $D_{\text{enrichment}}$  patients, the required number of events among biomarker-negative patients in terms of that among biomarker-positive patients ( $D_{\text{enrichment}}$ ) is given by

$$D_{\text{Sequential subgroup-specific}}^- = D_{\text{enrichment}} \left( \frac{\lambda_-}{\lambda_+} \right) \left( \frac{1 - k}{k} \right), \tag{31}$$

where  $\lambda_-$ ,  $\lambda_+$  are the event rates in biomarker-negative and biomarker-positive control subsets at the time when there are  $D_{\text{enrichment}}$  events in the biomarker-positive subgroup [60].

The significance levels  $\alpha$  can also be considered as one-sided significance levels in situations where our alternative hypothesis is not that there is just a treatment effect but that the treatment benefit in the experimental group is greater than that of the control group.

**Statistical considerations:** This strategy preserves the overall type I error rate  $\alpha$  but requires a smaller number of positive patients as compared to the second type of subgroup-specific design, the so-called parallel subgroup-specific design (see below). Furthermore, it enables the identification of treatment efficacy in the biomarker-positive and biomarker-negative subpopulations separately. However, it yields low power when there is homogeneity of treatment effect across the different biomarker-defined subpopulations. Furthermore, in case that test for treatment effect among biomarker-negative patients is not statistically significant, an ‘exploratory’ analysis on the biomarker-negative subgroup might be considered.

**Parallel Subgroup-Specific design:** This design was identified in three papers (3%) of our review.

**Design:** Parallel subgroup-specific design (Phase III), also referred to as a Phase III Biomarker-Stratified design evaluates treatment effects separately in the positive biomarker-defined subgroup and in the negative biomarker-defined subgroup simultaneously. A graphical illustration of this strategy is given in Figure 6.

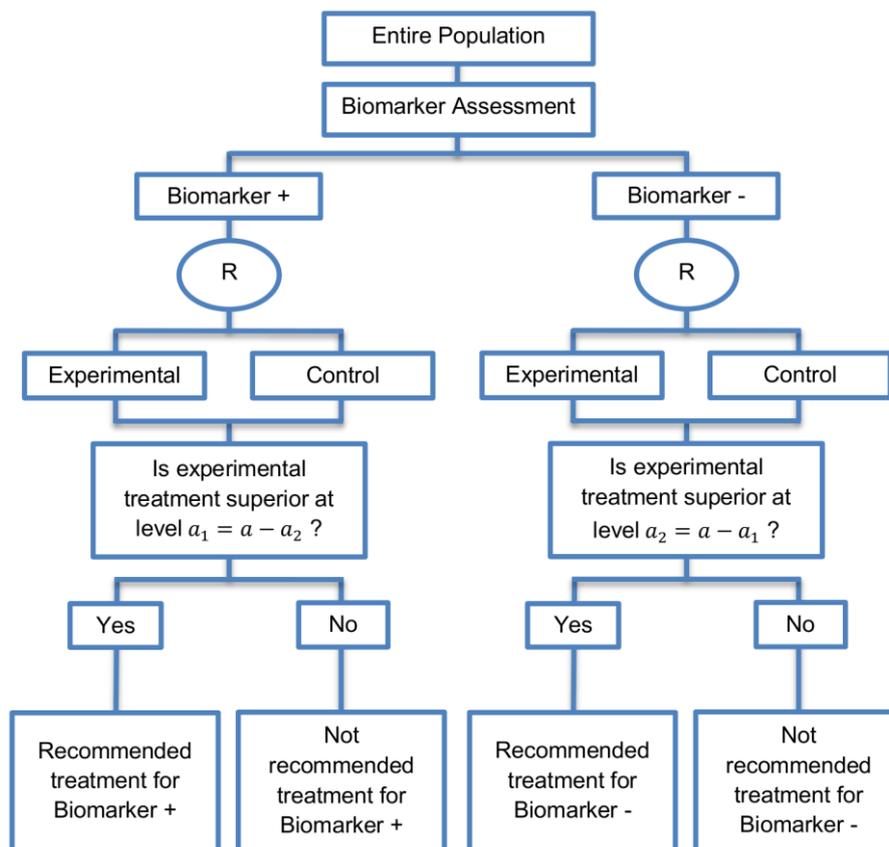


Figure 6. Parallel Subgroup-Specific design. “R” refers to randomization of patients.

**Utility:** It is appropriate when the aim of the study is to give treatment recommendations for each biomarker-defined subgroup separately at the same time.

**Methodology:** In order to control the overall type I error rate of the design at the overall level of significance  $\alpha$  (Type I error) it is required to allocate this overall  $\alpha$  between the test for the biomarker-positive subgroup and the test for the biomarker-negative subgroup using the Bonferroni correction method [124] for multiple testing; e.g., if we choose the value of 0.025 for the global significance level  $\alpha$ , then we could choose the values of  $a_1 = 0.010$  and  $a_2 = 0.015$  for testing the biomarker-negative and biomarker-positive subgroups respectively. This trial design is powered in such a way so as to detect the treatment effect in each biomarker-defined subgroup separately. A higher portion of the type I error rate can be given for the test within the biomarker-positive subgroup in order to maximize the power of the trial to identify the treatment effect in this subpopulation. However, even if there is a slight increase in the type I error probability spent on the test of one of the biomarker-defined subgroups, the power would probably not change much.

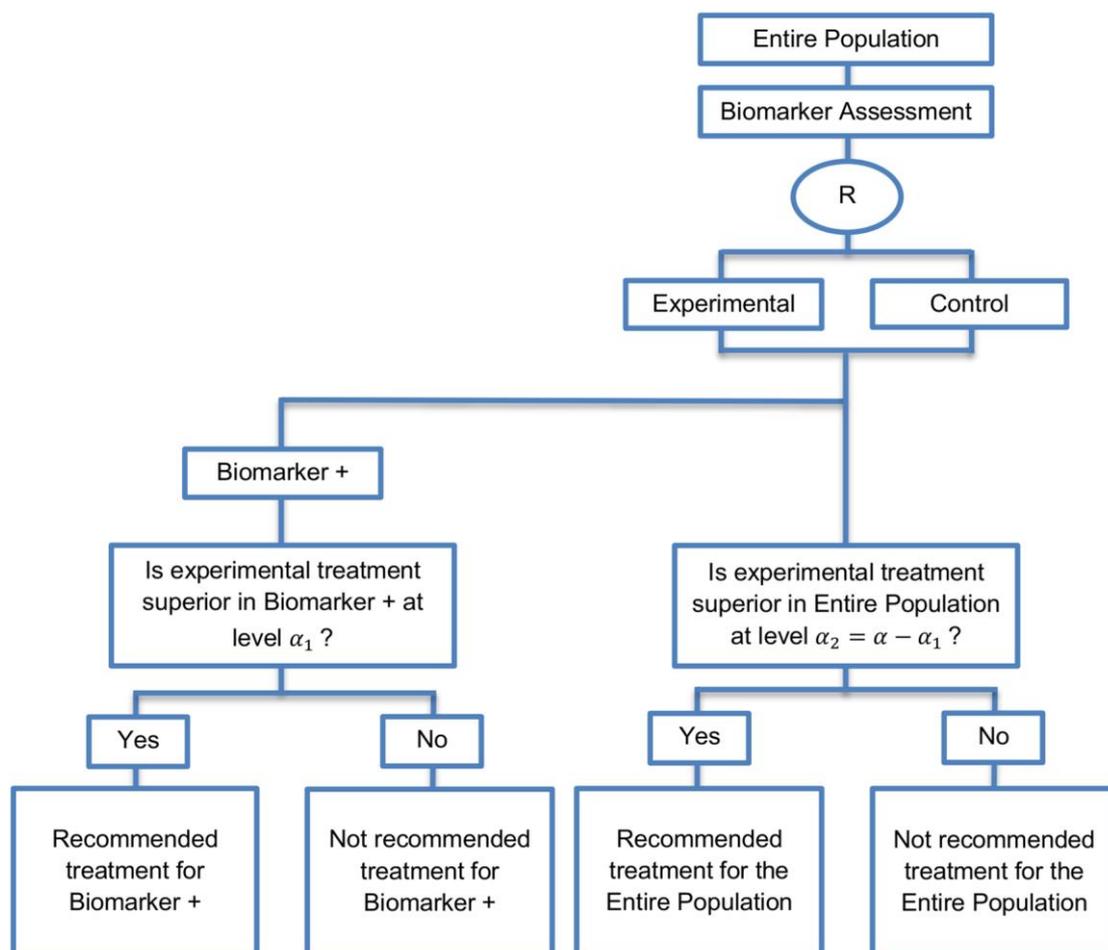
As in the sequential subgroup-specific design, the probability of rejecting either the null hypothesis of no treatment effect in the biomarker-positive subset or in the biomarker-negative effect under the global null hypothesis is less than or equal to the overall type I error rate  $\alpha$ . Additionally, the probability of rejecting the null hypothesis of no treatment effect in the biomarker-negative subpopulation when the treatment benefit is only restricted to biomarker-positive patients is less than or equal to  $\alpha$ . The significance levels  $\alpha$  can be considered as one-sided or two-sided significance levels.

**Statistical considerations:** With this approach, in case that the overall level of significance  $\alpha$  is equal in both subgroup-specific designs, it is more difficult to achieve statistical significance in the biomarker-positive subgroup as compared to the sequential subgroup-specific design due to the allocation of the overall significance level between the two biomarker-defined subgroup tests.

**Biomarker-positive and overall strategies:** This design provides an alternative strategy to analysing a biomarker-stratified design. It is an indirect way of evaluating both biomarker and treatment by testing the treatment effect in the entire study population and in the biomarker-positive subgroup separately. Three approaches are included in the biomarker-positive and overall strategies; the parallel assessment, the sequential assessment and the fall-back design (see below).

Despite the fact that the biomarker-positive subgroup and overall strategy design allows the treatment effect to be tested in the biomarker-positive subpopulation and provides good statistical power when the treatment effect is homogeneous across subgroups, this design is usually considered problematic and its use is not often recommended. More precisely, a major concern is that when the benefit of the novel treatment is limited to the biomarker-positive patients, it is possible that the design might lead to a wrong recommendation of treatment for the biomarker-negative patients. This might happen because when there is no treatment effect in the biomarker-negative subgroup, there might be an observed effect in the entire population due to the potentially large effect in the biomarker-positive patients. This concern is particularly pronounced in the sequential version of the design, which first tests the biomarker-positive subgroup and then, if it is positive, it tests the overall population.

**Biomarker-positive and overall strategies with parallel assessment:** This approach was identified in eight papers (8%) of our review. Figure 7 graphically represents this strategy. In the parallel version, we test both the overall population and biomarker-positive subgroup simultaneously.



**Figure 7.** Biomarker-positive and overall strategies with parallel assessment. “R” refers to randomization of patients.

**Design:** In this approach the treatment effect is tested in both the entire study population and in the biomarker-positive patients while controlling the type I error by allocating the overall significance level  $\alpha$  between the two tests. The significance level  $a$  can be considered as one-sided or two-sided.

**Utility:** The parallel version is recommended when the aim of the study is to assess the treatment effect in both the overall study population and in the biomarker-positive subgroup but not in the biomarker-negative subgroup.

**Methodology:** If there is significant confidence that the biomarker is predictive, the sample size estimation is aimed at having a sufficient number of biomarker-positive individuals to enable the treatment effect in the biomarker positive subgroup to be detected. On the other hand, if there is no confidence in the predictive value of the biomarker, the sample size estimation is aimed at having a sufficient number of patients to detect a treatment effect in the overall study population [14].

**Statistical considerations:** This design has the ability to control the probability of rejecting the null hypothesis of no treatment effect either in the biomarker-positive population or in the biomarker-negative population under the global null hypothesis of no treatment effect in the entire population at the overall significance level  $a$ . However, it cannot control the probability of rejecting the null hypothesis of no treatment effect in the biomarker-negative subset when the treatment benefit is restricted to biomarker-positive patients. Consequently, there is high risk of inappropriately recommending the experimental treatment for biomarker-negative patients.

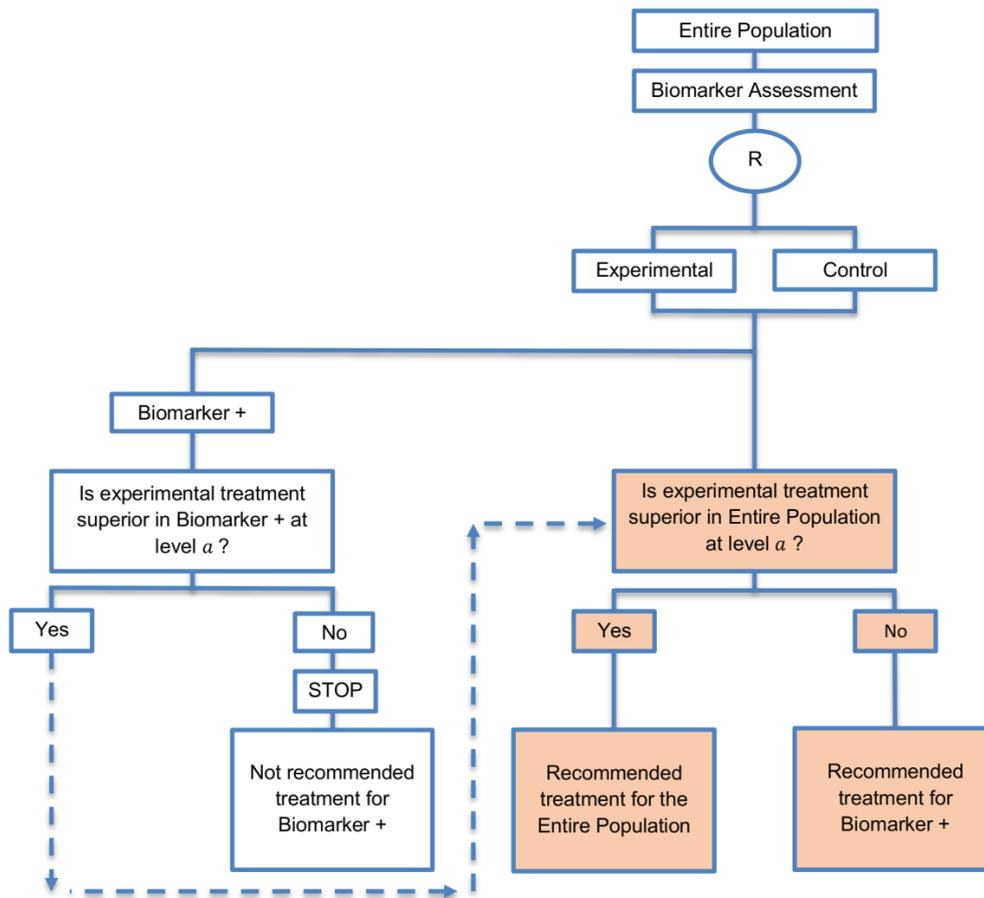
When the experimental treatment is compared to the control treatment within the overall population and the overall treatment effect is significant, then the test has high statistical power. If we are testing only the biomarker-positive subgroup and the treatment effect in this subgroup is significant, the statistical power is again high. This prospective subset analysis plan is based on testing both the overall study population and the biomarker-positive subgroup using significance levels, which are chosen in such a way that the overall significance level is equal or less than  $a$  (type I error). An easy way is to split  $a$  in such a way that the significance level for the entire population and the significance level for the biomarker-positive subset equals to overall significance level  $a$  (typically  $a = 0.05$ ). For example, the SATURN trial (NCT00556712) [96] which employs a prospective subset strategy used the value of 0.03 as level of significance to test the treatment effect in the entire population and the value of 0.02 to test the treatment effect in the biomarker-positive subset; therefore, the overall level of significance was preserved at 0.05. The approach can be overly conservative as in the SATURN trial because of the correlation between the global and subgroup test. Other approaches [98,125–128] have been proposed for adjusting the level of significance of both tests in a more accurate and less conservative way.

**Biomarker-positive and overall strategies with sequential assessment:** This approach was referred to in 11 papers (11%) of our review. A graphical illustration of this approach is shown in Figure 8.

**Design:** In this sequential version of the biomarker-positive and overall strategies, we first test the biomarker-positive subgroup using the significance level  $\alpha$ ; if the test is significant, then we test the treatment effect in the overall population using the same  $\alpha$  level. The significance levels  $a$  can be considered as one-sided or two-sided significance levels.

**Utility:** The sequential version might be useful in cases where the experimental treatment is expected to be effective in the overall study population.

**Methodology:** As this design comprises two sequential stages, it follows that the sample size calculation should also be staged. At the first stage, the standard formula for a traditional randomized trial can be used for the biomarker-positive subgroup using the significance level  $\alpha$  to estimate the treatment effect in that subset. More precisely, the formula used in the enrichment design for the required total number of events or the required number of patients can be used at the first stage of this design. At the second stage, the sample size must be adjusted in order to yield appropriate power for the entire population.



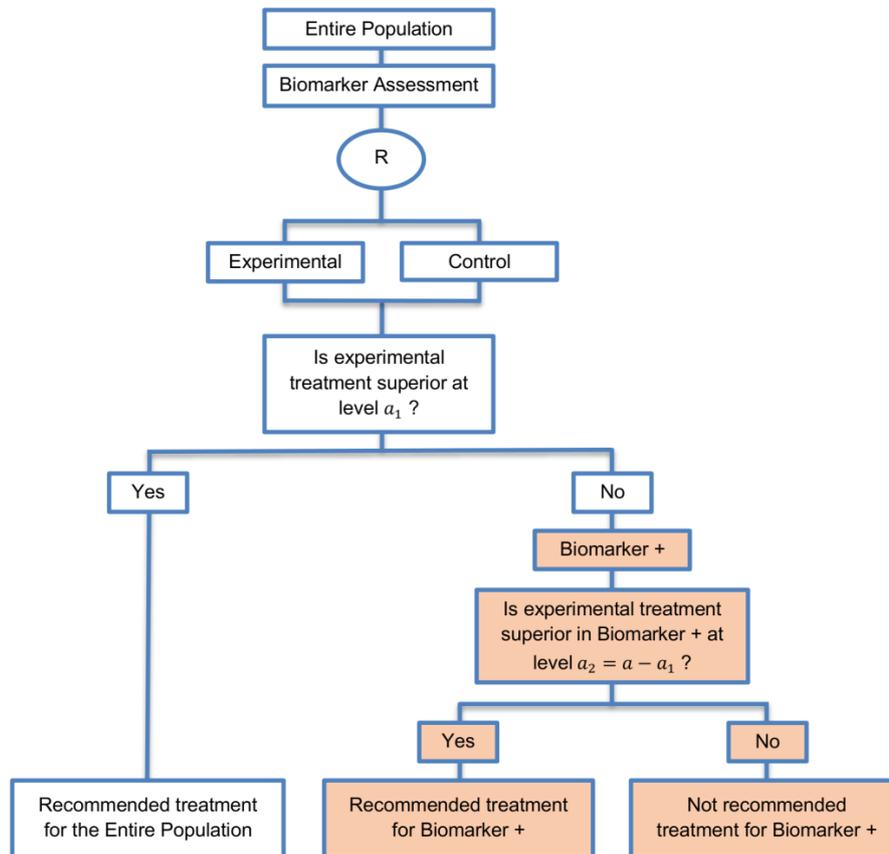
**Figure 8.** Biomarker-positive and overall strategies with sequential assessment. “R” refers to randomization of patients. Uncolored boxes are referred to the first stage of the trial and colored boxes are referred to the second stage of the trial.

**Statistical considerations:** As in the parallel version of this designs, this strategy does not allow for identification of treatment efficacy in the biomarker-negative subgroup and despite the fact that it can control the overall type I error  $\alpha$  it cannot control the probability of rejecting the null hypothesis of no treatment effect in the biomarker-negative subset when the treatment benefit is restricted to biomarker-positive patients. Consequently, for this design also there is high risk of inappropriately recommending the novel treatment for biomarker-negative patients.

**Biomarker-positive and overall strategies with fall-back analysis:** This strategy was identified in 15 papers (15%) of our review. It evaluates both the treatment effect in the overall study population and in the biomarker-positive subgroup sequentially. Figure 9 graphically represents this strategy.

**Design:** In the fall-back design, we first test the overall population using the reduced significance level  $a_1$  and if the test is significant, we consider that the novel treatment is effective in the overall population; however, if the result is not significant then we test the treatment effect in the biomarker-positive subgroup using the level of significance  $a_2 = a - a_1$ , where  $a$  is the overall significance level (Type I error rate). The significance levels  $a$  can be considered as one-sided or two-sided significance levels. The same analysis plan was used in the adaptive signature design which is further described in our methodological review regarding the biomarker-guided adaptive designs, Antoniou et al., 2016 [35]. More precisely, the difference between the adaptive signature design and the fall-back design is the following: in the adaptive signature design, in case that the first stage failures to show treatment effectiveness in the entire population, then the study population is divided in order to develop and validate a biomarker, using a split sample strategy, whereas in the biomarker-positive

and overall strategies design with fall-back analysis the biomarker assessment is conducted at the beginning of the trial. However, both of the designs test at the first stage the entire population at the significance level  $a_1$  and at the second stage the biomarker-positive patients at the significance level  $a_2 = a - a_1$ .



**Figure 9.** Biomarker-positive and overall strategies with fall-back analysis. “R” refers to randomization of patients. Uncolored boxes are referred to the first stage of the trial and colored boxes are referred to the second stage of the trial.

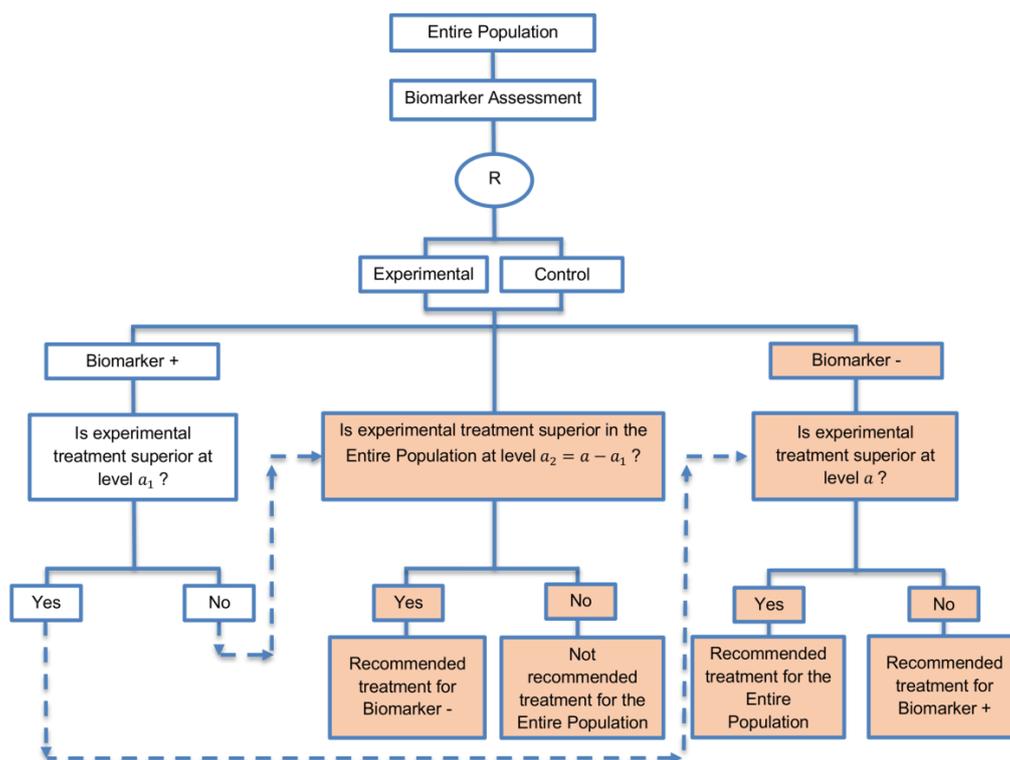
**Utility:** This approach is recommended when there is insufficient confidence in the predictive value of the biomarker and that the novel treatment is believed to be effective in all individuals (i.e., the rationale for the biomarker is weak). This design can be used in order to avoid the possibility of missing an important treatment effect in the biomarker-positive patients (with insufficient benefit in the biomarker-negative subgroup).

**Methodology:** The sample size should be set in such a way so as to yield adequate power for the overall test at the reduced significance level  $a_1$  and for the potential biomarker positive subgroup analysis at significance level  $a - a_1$  [60]. The fall-back version is identical to the parallel version of biomarker-positive and overall strategies in terms of sample sizes and study outcomes, however the difference between these approaches is that the fall-back strategy is useful in settings where a biomarker will be assessed only if the overall population benefit is not promising [14]. This strategy can test the treatment effectiveness in biomarker-positive patients even if there is no detected benefit of the novel treatment in the overall population. However, it does not evaluate clearly the treatment benefit in the biomarker-negative subpopulation.

**Statistical considerations:** As the two aforementioned biomarker-positive and overall designs, this strategy can again control the overall type I error  $\alpha$  but it cannot control the probability of rejecting the null hypothesis of no treatment effect in the biomarker-negative subgroup when the treatment

benefit is restricted to biomarker-positive patients. Consequently, there is high risk of inappropriately recommending the novel treatment for biomarker-negative patients. Song et al., 2007 [129] and George, 2008 [1] have discussed refinement of the significance levels associated with this design, which takes into account the correlation between the test for overall treatment effect and the test for the biomarker-positive treatment effect [60]. Additionally, a recent paper by Choai et al., 2015 [97] proposes a bias-corrected estimation method for treatment effects for the all-comers randomized clinical trials with a predictive biomarker which incorporate the fall-back analysis. For Choai et al., 2015 [97] the terminology “all-comers randomized clinical trials” is referred to the “Biomarker-positive and overall strategies with fall-back analysis”. More precisely, as this study design has an adaptive nature and is composed of two stages, a bias is possible to arise in the treatment effect estimation in the biomarker-positive subset when the first stage of the trial yields an overall result which is not significant and thus fails to demonstrate a treatment efficacy in the entire population. For this reason, Choai et al., 2015 [97], formulate a bias function using polynomials in order to take into account the possibility of failing to demonstrate overall treatment efficacy during the first stage of the trial.

**Marker Sequential test design (MaST):** This design was identified in four papers (4%) of our review and while controlling the appropriate type I error rates, it evaluates not only the biomarker-positive and biomarker-negative subgroups but also the entire population sequentially to limit the assessment of treatment effect in the overall population when it seems that the biomarker-positive subgroup does not benefit from the novel treatment. A graphical illustration of this approach is given in Figure 10.



**Figure 10.** Marker Sequential test design (MaST). “R” refers to randomization of patients. Uncolored boxes are referred to the first stage of the trial and colored boxes are referred to the second stage of the trial.

**Design:** In this design which owns an adaptive nature, first, the biomarker-positive subgroup is tested at a reduced level  $a_1$  in  $[0, a]$  and if the result is significant, then the biomarker-negative subgroup is tested at the global significance level  $\alpha$ . Otherwise, if the result is not significant, then the

overall population is tested at level  $a_2 = a - a_1$  in order to make a treatment recommendation for the biomarker-negative patients.

**Utility:** It is generally recommended when robust evidence is available regarding a biomarker and there is prior evidence showing that the novel treatment is more beneficial for the biomarker-positive patients as compared to the biomarker-negative patients. Additionally, it is appropriate when we can assume that the treatment will not be beneficial for the biomarker-negative subgroup unless it is effective for the biomarker-positive subgroup. Additionally, the marker sequential test design is considered as an alternative to the sequential subgroup-specific design when the aim is to consider the treatment effect not only in biomarker-positive but also in the biomarker-negative patients.

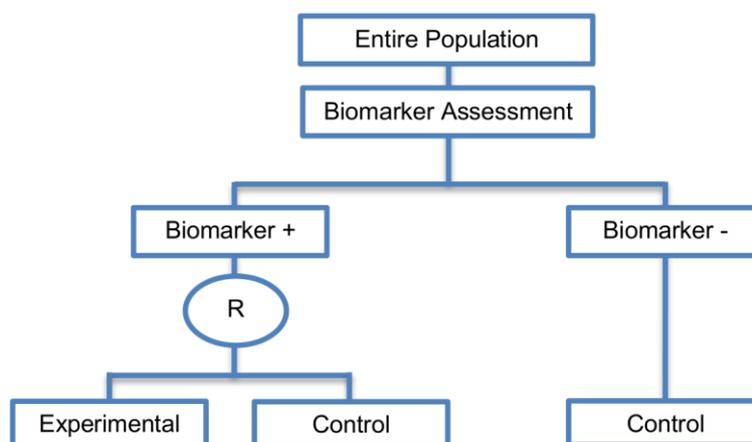
**Methodology:** Freidlin et al., 2014 [69] recommended using the value of 0.022 for the reduced significance level  $a_1$  in order to control the type I error rate for biomarker-negative patients at the global significance level  $\alpha = 0.025$  and the value of 0.04 for the reduced significance level  $a_1$  in order to control the type I error rate for biomarker-negative patients at the global significance level  $\alpha = 0.05$ .

Regarding the sample size for such a design where there is prior evidence indicating strong predictive ability of the biomarker, a standard sample size calculation (i.e., the same sample size calculation as for the enrichment designs) can be used for biomarker-positive subpopulation or alternatively, researchers can use the sample size calculation used for the sequential subgroup-specific design. However, in order to have sufficient number of biomarker-positive patients to detect treatment effectiveness in that particular biomarker-defined subset and consequently to reach the desired power, the sample size should be calculated using the reduced level  $a_1 [0, a]$  instead of the global significance level  $\alpha$  which is used in the sample size formulae of the enrichment and sequential subgroup-specific designs. This will result in a small increase in the number of patients as compared to the enrichment and sequential subgroup-specific designs. Otherwise, if the reduced significance level  $a_1$  is not used, this would yield minor loss of power.

**Statistical consideration:** Freidlin et al., 2014 [69] performed a comparison between the MaST and the sequential subgroup-specific design through a simulation study and concluded that the marker sequential design yields higher power in cases where the treatment effect is homogeneous across biomarker-defined subgroups. Additionally, with this approach, the power is preserved in situations where the experimental treatment is effective only for the biomarker-positive patients. Furthermore, in situations where biomarker status is not available for a portion of patients included in the trial, the marker sequential test design can either exclude these patients or include them in the global test, whereas, the proposed subgroup-specific designs do not consider inclusion of these patients in the analyses. If researchers decide to exclude patients with unavailable biomarker status from the study when using a MaST design, no statistical adjustment is required. On the other hand, if the inclusion of this study population is chosen, then this can result in inflation of the type I error rate for the biomarker-negative subpopulation above the global significance level  $\alpha$  due to the modification of correlation structure between the biomarker-defined subgroup tests and global test. In addition, while both MaST and subgroup-specific designs have the ability to control the probability of incorrectly rejecting the null hypothesis of no treatment effect in the biomarker-negative patients at the significance level  $\alpha$  when the experimental treatment does not work in either biomarker-defined subgroup, the sequential subgroup-specific approach typically has a smaller probability of incorrectly rejecting the null hypothesis of no treatment effect in the biomarker-negative subset (when the null hypothesis is true) as compared to the MaST design, especially under the global null hypothesis of no treatment effect in the entire population; the probability of incorrectly rejecting the null hypothesis of no treatment effect in the biomarker-negative patients depends on the choice of  $a_1$ . This conservativeness of sequential subgroup-specific design, which is due to its sequential nature, makes the MaST design advantageous [69].

### 2.3.2. Hybrid Designs

Hybrid designs (Phase III) were identified in 14 papers (14%) of our review and they can be included in the all-comers designs, where the entire population is firstly screened for biomarker status and all individuals enter the trial. A graphical illustration of this design is given in Figure 11.



**Figure 11.** Hybrid design. “R” refers to randomization of patients.

**Design:** In this approach, only the biomarker-positive patients are randomly assigned to either the experimental treatment group or to the control treatment group whereas the biomarker-negative patients receive the control treatment. These designs were first defined by Mandrekar and Sargent [30,31]. The difference compared with the enrichment designs is that the biomarker-negative patients are not excluded from the study.

**Utility:** Hybrid designs can be used when there is compelling prior evidence which shows detrimental effect of the experimental treatment for a specific biomarker-defined subgroup (i.e., biomarker-negative subgroup) or some indication of its possible excessive toxicity in that subgroup, thus making it unethical to randomize the patients within this population to the experimental treatment.

**Methodology:** Similar to the enrichment design, hybrid designs are powered to identify treatment effect only in the biomarker-defined subgroup which is randomly assigned to the experimental or control treatment groups. Consequently, the same formula used for the required number of patients or events for the enrichment designs can be used for hybrid designs. This design is a combination of an enrichment design where we randomize patients to either the experimental or the control treatment group and a single-arm design in biomarker-negative patients.

**Statistical considerations:** The strength of the hybrid design is that apart from the evaluation of the predictive ability of a biomarker, the feasibility of a prognostic biomarker can also be tested. It can be considered as an advantageous design of the enrichment designs when there is prior evidence showing not only that the control treatment works well for the biomarker-negative population but also a detrimental effect of the experimental treatment for that subgroup or possible excessive toxicity as we do not exclude these patients from the trial as it happens in the enrichment designs.

### 2.4. Biomarker-Strategy Designs

Generally, with biomarker-strategy designs, the study population is randomized to treatment strategies as opposed to treatments per se. More precisely, patients are randomized to either a biomarker-based treatment strategy arm where the biomarker is used in deciding on approach to treatment, or to an arm that does not use the biomarker to guide treatment. Consequently, biomarker-strategy designs make a comparison between two strategies—one which uses biomarker information to inform treatment approach and the other that does not.

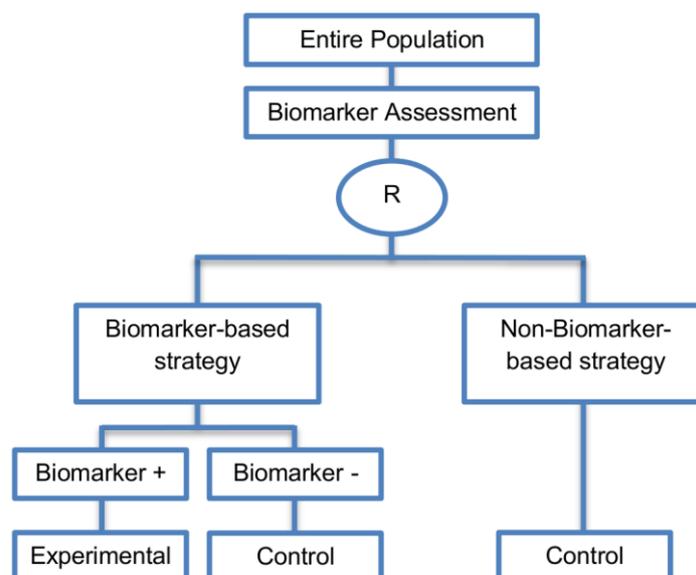
These designs are also known as biomarker-based strategy designs or signature-based strategy designs and they are composed of four subtypes; (i) biomarker-strategy designs with biomarker assessment in the control arm; (ii) biomarker-strategy designs without biomarker assessment in the control arm; (iii) biomarker-strategy designs with treatment randomization in the control arm and (iv) reverse marker-based strategy designs. Whilst patients randomized to the non-biomarker based strategy arm in the first two design subtypes are allocated the control treatment, in the third design subtype those patients undergo secondary randomization to either the control or experimental treatment. The fourth design subtype differs from the three aforementioned subtype designs as the non-biomarker based strategy arm is replaced by the reverse marker-strategy arm. The first and second types are similar with the difference being only in terms of ethical/feasibility issues regarding the acquisition of biomarker status at the beginning of the trial.

This approach is preferred when the study is planned for a confirmatory phase of a certain biomarker-based strategy allowing for comparison between the biomarker-based strategy and non-biomarker-based strategy.

#### 2.4.1. Biomarker-Strategy Design with Biomarker Assessment in the Control Arm

This approach is described in 21 (21%) papers of our review.

**Design:** First, the study population enrolled in the trial is tested for its marker status. Next, patients irrespective of their biomarker status are randomized either to the biomarker-based strategy arm (also referred to as personalized arm) or to the non-biomarker-based strategy arm. In the biomarker-based strategy arm, biomarker-positive patients receive the experimental treatment, whereas, biomarker-negative patients receive the control treatment. Patients who are randomized to the non-biomarker-based strategy arm receive the control treatment irrespective of their biomarker status. A graphical illustration of this design is given in Figure 12. This biomarker-strategy design can be extended to more than one experimental treatment. More precisely, this extension is referred to as Individual profile design in literature and was identified in two papers [36,72] (2%) of our review. This design includes different individual status, e.g., instead of biomarker-positive and biomarker-negative subgroups we can have patients who are positive for biomarker 1, biomarker 2, biomarker n, leading to the selection of personalized treatments, (patients who are positive for biomarker 1 are treated with the corresponding experimental treatment 1, etc.).



**Figure 12.** Biomarker-strategy design with biomarker assessment in the control arm. “R” refers to randomization of patients.

**Utility:** This approach is useful when we want to test the hypothesis that the treatment effect based on the biomarker-based strategy approach is superior to that of the standard of care.

**Methodology:** The clinical utility of a biomarker can be evaluated by comparing the two strategy groups. The predictive utility of the marker-based treatment strategy could be assessed by comparing the outcome of all patients in the biomarker-based strategy arm to all patients in the non-biomarker-based strategy arm. Patients in the marker-based strategy arm do not need to be limited to two treatments; in principle, a marker-based strategy involving many biomarkers and many possible treatments could be compared to standard of care treatment.

According to Freidlin et al., 2010 [61], assuming a survival outcome, the required sample size in terms of number of events for this type of biomarker-strategy design in order to reach power  $(1 - \beta)$  at significance level  $\alpha$  (type I error) can be given by

$$D_{strategy\ I} = 4 \left[ \frac{(z_{\alpha/2} + z_{\beta})}{k \log \theta_1} \right]^2, \tag{32}$$

where  $k$  denotes the prevalence of biomarker-positive patients,  $\theta < 1$  denotes the assumed hazard ratio in the biomarker-positive subpopulation and  $z_{\alpha/2}$ ,  $z_{\beta}$  denote the upper  $\alpha/2$ - and upper  $\beta$ -points respectively of a standard normal distribution where  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively. According to Freidlin et al. 2010 [61], it is assumed that there is no treatment effect in the biomarker-negative subpopulation (corresponding to a hazard ratio of experimental treatment versus control treatment of 1) and that there is no prognostic effect of the biomarker under the control treatment. Consequently, the overall hazard ratio between experimental and control arms in biomarker-positive patients and biomarker-negative patients can be approximated by  $\exp[k \log \theta + (1 - k) \log 1] = \theta^k$  [61] and this is the reason why the formula which gives the required total number of events ( $D_{strategy}$ ) contains only the hazard ratio of biomarker-positive patients. Freidlin et al., 2010 [61] provided the aforementioned formula assuming that all random assignments use 1:1 randomization.

Additionally, Young et al., 2010 [26] determined the total sample size needed for this type of biomarker-strategy designs when using continuous clinical endpoints by

$$N_{strategy\ I} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 (\tau_m^2 + \tau_n^2)}{(v_m - v_n)^2}, \tag{33}$$

where  $z_{1-\alpha/2}$ ,  $z_{1-\beta}$  denote the lower  $1 - \alpha/2$ - and lower  $1 - \beta$ -points respectively of a standard normal distribution,  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively,  $v_m$  and  $v_n$  denote the mean response from the biomarker-based strategy arm and the non-biomarker-based strategy arm respectively, and  $\tau_m^2$ ,  $\tau_n^2$  denote the variance of response for the biomarker-based strategy arm and non-biomarker-based strategy arm respectively. Young et al., 2010 [26] also provided formulae for the aforementioned variances which depend on sensitivity and specificity of the assay, such that any error in the evaluation of biomarker in the biomarker-based strategy can be accounted for.

For the case of binary outcomes, Eng, 2014 [92] provided the formula for the required sample size for each arm in a test of proportions between the two randomization arms (biomarker-based strategy arm and non-biomarker-based strategy arm). This formula can be given by

$$N_{strategy\ I/arm} = \frac{(z_{\alpha} + z_{1-\beta})^2 [g_1(1 - g_1) + g_2(1 - g_2)]}{\Delta_2^2}, \tag{34}$$

where  $\alpha$  corresponds to the target level,  $1 - \beta$  corresponds to the power,  $g_1$  is the expected response rate in the biomarker-based strategy arm,  $g_2$  is the expected response rate in the non-biomarker-based strategy arm and  $\Delta_2 = g_1 - g_2$ . The expected response rates  $g_1, g_2$  can be found by calculating the formulae  $kr_{A+} + (1 - k)r_{B-}$  and  $r_B$  respectively, the prevalence of biomarker-positive patients

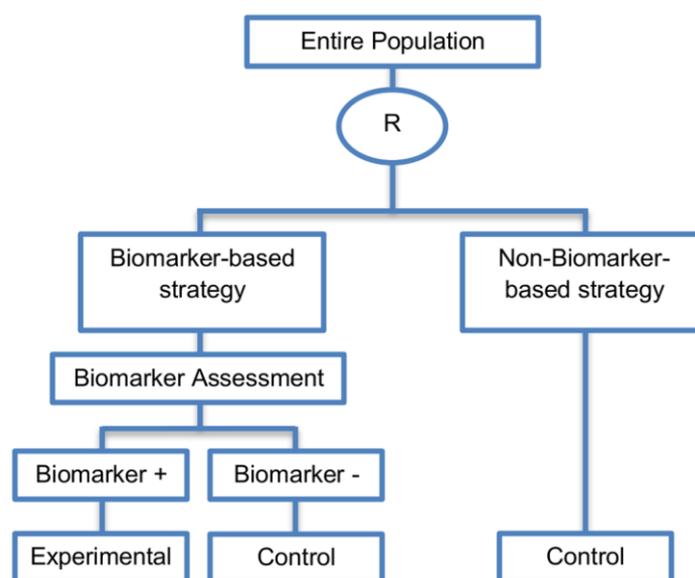
corresponds to  $k$  and  $r_{A+}, r_{B-}$  are the assumed response rates of biomarker-positive patients receiving the experimental treatment and biomarker-negative patients receiving the control treatment,  $r_B$  denotes the marginal effect of treatment B (control treatment).

**Statistical considerations:** This type of designs is able to inform researchers whether the biomarker is prognostic, since both biomarker positive and negative patients are exposed to the control treatment, but it cannot answer the question of whether the biomarker is predictive since only biomarker positive patients are exposed to the experimental treatment. Additionally, these designs have been criticized by many authors as less efficient than the marker-stratified designs since it is possible for some patients in both the biomarker-based strategy arm and non-biomarker-based strategy arm to be assigned to the same treatment (due to the existence of biomarker-negative patients in both strategy arms the treatment effect can be diluted) and they require a large sample size to detect an overall difference in outcomes between arms. Furthermore, these designs cannot compare experimental treatment to control treatment directly as they are designed to compare not the treatments but the biomarker-strategies. Another limitation of these designs is the uncertainty about whether the results which indicate efficacy of the biomarker-directed approach to treatment are caused due to a true effect of the biomarker or due to a treatment effect irrespective of the biomarker status.

#### 2.4.2. Biomarker-Strategy Design without Biomarker Assessment in the Control Arm

This strategy was identified in 14 papers (14%) of our review.

**Design:** In this approach, patients are again randomized between testing strategies (i.e., biomarker-based strategy and non-biomarker-based strategy) but it differs in terms of the timing of biomarker evaluation. More precisely, first, patients are randomized to either the biomarker-based strategy or to the non-biomarker-based strategy. Next, this design evaluates the biomarkers only in patients who are assigned to the biomarker-based strategy. Patients who are found to be biomarker-positive will receive the experimental treatment and patients who are biomarker-negative will receive the control treatment. On the other hand, the population which is randomized to the non-biomarker-based strategy will receive the control treatment. A graphical illustration of this design is given in Figure 13.



**Figure 13.** Biomarker-strategy design without biomarker assessment in the control arm. “R” refers to randomization of patients.

**Utility:** This design is useful in situations where it is either not feasible or ethical to test the biomarker in the entire population due to several logistical (e.g., specimens not submitted), technical (e.g., assay failure) or clinical reasons (e.g., tumor inaccessible); thus the biomarker status is obtained only in patients who are tailored to the biomarker-based strategy arm.

**Methodology:** The same mathematical formula for sample size calculation assuming a continuous clinical outcome proposed by Young et al. (2010) [26] and the formula assuming binary outcome proposed by Eng, 2014 [92] for the biomarker-strategy design with biomarker assessment in the control arm could be applied. Further, in terms of survival outcome, the same formula provided for the required number of events in the first version of biomarker-strategy designs (i.e., biomarker-strategy design with biomarker assessment in the control arm) could be considered.

**Statistical considerations:** These designs have the same advantages and limitations as the previously discussed biomarker-strategy design with biomarker assessment in the control arm, e.g., they have been criticized for their lack of efficiency due to the fact that biomarker negative patients are exposed to the control treatment in both arms of the trial. An additional limitation is that the biomarker-positive and biomarker-negative subpopulations might be more imbalanced as compared with the first type of biomarker-strategy design due to the fact that the randomization is performed before the evaluation of biomarker (balancing the randomization is useful to ensure that all randomized patients have tissue available).

#### 2.4.3. Biomarker-Strategy Design with Treatment Randomization in the Control Arm

Sargent and Allegra [108] proposed another version of Biomarker-strategy designs where there is a second randomization between experimental and control treatment in the non-biomarker guided strategy arm. This strategy is referred to in 17 papers (17%) of our review.

**Design:** A graphical illustration of this approach is given in Figure 14. The two previously described biomarker-strategy designs can answer the question about whether the biomarker-based strategy is more effective than standard treatment, irrespective of the biomarker status of the study population, whereas the biomarker-strategy design with treatment randomization in the control treatment is able to inform us about whether the biomarker-based strategy is better than not only the standard treatment but also better than the experimental treatment in the overall population. This is achieved by using a second randomization the ratio of which should be informed by the prevalence of the biomarker in question in the population as a whole to ensure balance between the study arms. Patients are first randomly assigned to either the biomarker-based strategy arm or to the non-biomarker-based strategy arm. Next, patients who are allocated to the non-biomarker-based strategy are again randomized either to the experimental treatment arm or to the standard treatment arm irrespective of their biomarker status. Patients who are allocated to the biomarker-based strategy and who are biomarker-positive are given the experimental treatment and patients who are biomarker-negative are given the control treatment. The clinical utility of the biomarker is evaluated by comparing treatment effect between the biomarker-based strategy arm and non-biomarker-based strategy arm. Such an approach can also identify whether a novel treatment is more effective in the entire population or in a biomarker-defined subgroup only, since both biomarker subgroups are exposed to both treatments.

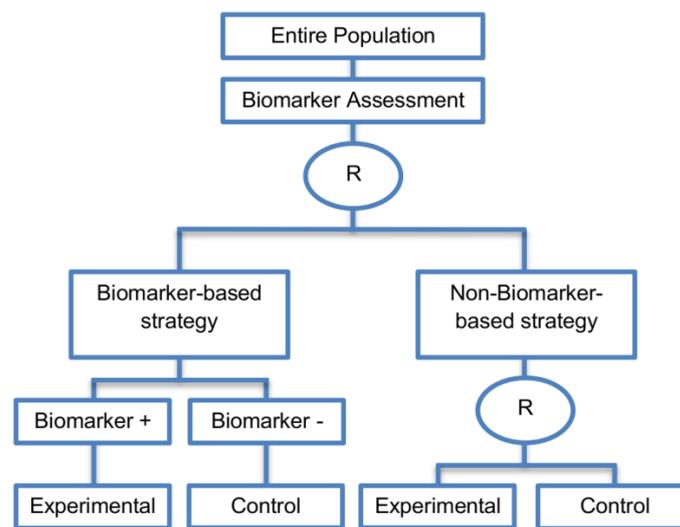
**Utility:** These designs are preferable as compared to the two previously discussed biomarker-strategy designs in cases where there is interest in whether the biomarker is not only prognostic but also predictive.

**Methodology:** Mandrekar and Sargent, 2009 [31] calculated the total required sample size in terms of number of events for the comparison of a survival outcome in the biomarker-based strategy

versus the non-biomarker-based strategy. According to them, the required total number of events when using 1:1 randomization to treatment arms is given by

$$D_{strategy\ III} = \frac{4(z_{\alpha/2} + z_{\beta})^2}{\left\{ \log \left[ \frac{2\kappa m_{B+} + 2(1-\kappa)m_{A-}}{\kappa(m_{A+} + m_{B+}) + (1-\kappa)(m_{A-} + m_{B-})} \right] \right\}^2}, \tag{35}$$

where  $\kappa$  denotes the prevalence of the biomarker-positive patients,  $m_{A+}, m_{A-}, m_{B+}, m_{B-}$ , denote the median survival for biomarker-positive and biomarker-negative patients receiving control and experimental treatments respectively. Also, the constants  $z_{\alpha/2}, z_{\beta}$  denote the upper  $\alpha/2$ - and upper  $\beta$ -points respectively of a standard normal distribution where  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively.



**Figure 14.** Biomarker-strategy design with treatment randomization in the control arm. “R” refers to randomization of patients.

Additionally, Young et al., 2010 [26], considering continuous clinical outcomes, calculated the total sample size by

$$N_{strategy\ III} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 (\tau_m^2 + \tau_{nr}^2)}{(v_m - v_{nr})^2}, \tag{36}$$

where  $z_{1-\alpha/2}, z_{1-\beta}$  denote the lower  $1 - \alpha/2$ - and lower  $1 - \beta$ -points respectively of a standard normal distribution,  $\alpha$  and  $\beta$  denote the assumed type I error and type II error respectively,  $v_m$  and  $v_{nr}$  denote the mean response from the biomarker-based strategy arm and the non-biomarker-based strategy arm, and  $\tau_m^2, \tau_{nr}^2$  denote the variance of response for the biomarker-based strategy arm and non-biomarker-based strategy arm respectively. The only differences in the mathematical formula for the total sample size  $n_t$  between this type of biomarker-strategy design and the first and second types mentioned above are the values of  $v_{nr}$  and  $\tau_{nr}^2$ , to reflect the fact that in the non-biomarker-based strategy arm patients are randomly assigned to either the experimental or control treatment. Again, the formulae can be adjusted to account for uncertainty in biomarker assessment.

For the case of binary outcomes, Eng, 2014 [92] provided the formula for the required sample size for each arm in a test of proportions between the two randomization arms (biomarker-based strategy arm and non-biomarker-based strategy arm). This formula can be given by

$$N_{strategy\ III/arm} = \frac{(z_{\alpha} + z_{1-\beta})^2 [g_1(1 - g_1) + g_3(1 - g_3)]}{\Delta_3^2} \tag{37}$$

where  $\alpha$  correspond to the target level,  $1 - \beta$  corresponds to the power,  $g_1$  is the expected response rate in the biomarker-based strategy arm,  $g_3$  is the expected response rate in the non biomarker-based strategy arm and  $\Delta_3 = g_1 - g_3$ . The expected response rates  $g_1, g_3$  can be found by calculating the formulae  $kr_{A+} + (1 - k)r_{B-}$  and  $r_A/2 + r_B/2$  respectively,  $r_A$  and  $r_B$  denote the marginal effect of treatment A (experimental treatment) and treatment B (control treatment) respectively.  $r_{A+}, r_{B-}$  are the assumed response rates of biomarker-positive patients receiving the experimental treatment and biomarker-negative patients receiving the control treatment. The prevalence of biomarker-positive patients corresponds to  $k$ .

**Statistical considerations:** Similar to both aforementioned biomarker-strategy designs, the biomarker-strategy design with treatment randomization in the control arm will need larger sample size as compared to the marker-stratified designs. However, one strength is that they allow clarification of whether the results which indicate efficacy of the biomarker-directed approach to treatment are caused due to a true effect of the biomarker or due to a treatment effect irrespective of the biomarker status which does not happen in the first two types of biomarker-strategy designs.

#### 2.4.4. Reverse Marker-Based Strategy Design

Eng, 2014 [92] proposed another version of biomarker-strategy designs where the non-biomarker-based strategy arm which is included in the three aforementioned subtypes of biomarker-strategy designs is replaced by the reverse marker-strategy arm. This strategy is referred to in four papers (4%) of our review.

**Design:** A graphical illustration of this approach is given in Figure 15. In this design patients are randomized either to the biomarker-based strategy arm or the reverse biomarker-based strategy arm. As in the previous three biomarker-strategy subtype designs, patients who are allocated to the biomarker-strategy arm receive the experimental treatment if they are biomarker-positive whereas biomarker-negative patients receive the control treatment. By contrast, patients who are randomly assigned to the reverse biomarker-based strategy arm receive control treatment if they are biomarker-positive, whereas biomarker-negative patients receive experimental treatment.

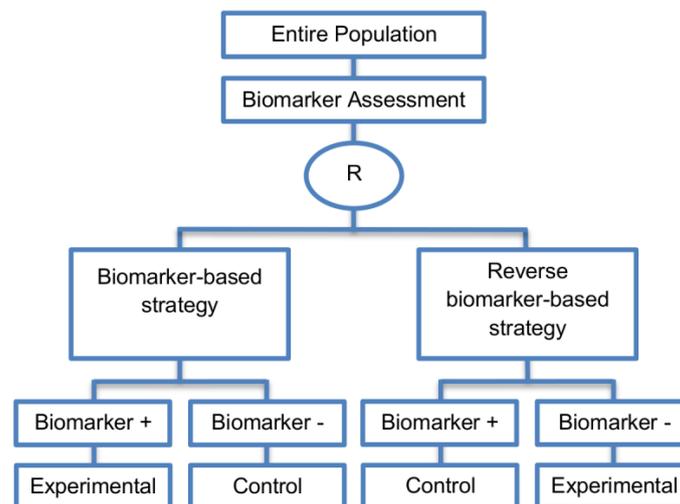


Figure 15. Reverse Marker-Based strategy design. “R” refers to randomization of patients.

**Utility:** Reverse marker-based strategy is a more efficient strategy as compared to the first and third biomarker-strategy subtype design for testing the interaction hypothesis of treatment and biomarker. This design should be used in cases where prior evidence indicates that both experimental and control treatment are effective in treating patients but the optimal strategy has not yet been identified.

**Methodology:** This subtype design is balanced (i.e., the randomization frequencies for each treatment are equal independent of the prevalence of the biomarker) and it is powered to evaluate the interaction between treatment and biomarker. For the case of binary outcomes, Eng, 2014 [92] provided the formula for the required sample size for each arm in a test of proportions between the two randomization arms (biomarker-based strategy arm and reverse biomarker-based strategy arm). This formula can be given by

$$N_{strategy\ IV/arm} = \frac{(z_{\alpha} + z_{1-\beta})^2 [g_1(1 - g_1) + g_4(1 - g_4)]}{\Delta_4^2} \tag{38}$$

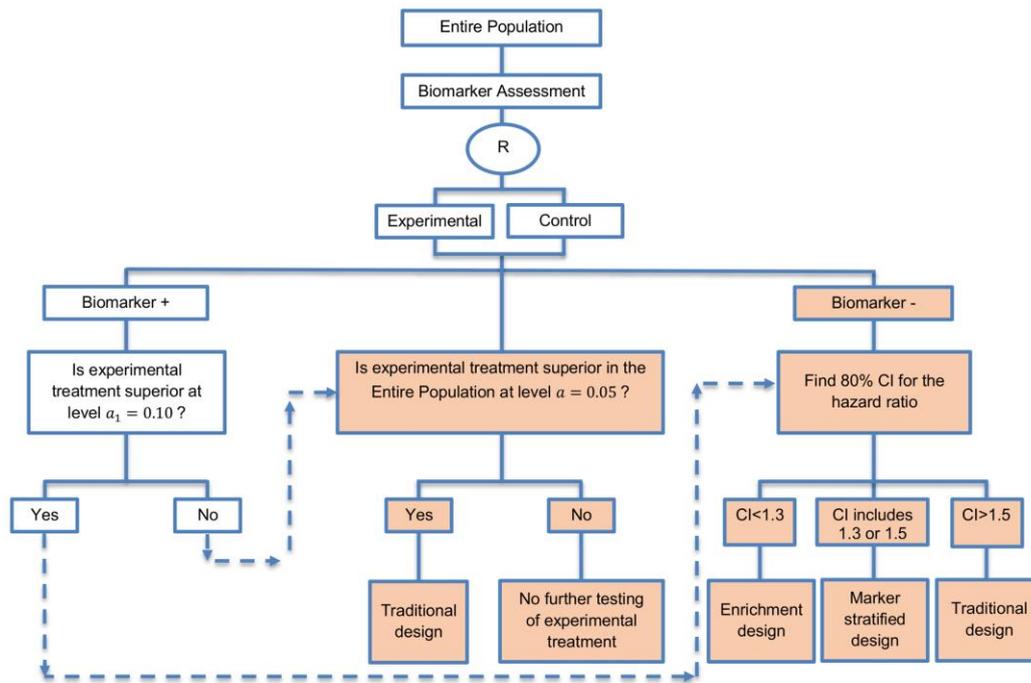
where  $\alpha$  correspond to the target level,  $1 - \beta$  corresponds to the power,  $g_1$  is the expected response rate in the biomarker-based strategy arm,  $g_4$  is the expected response rate in the reverse biomarker-based strategy arm and  $\Delta_4 = g_1 - g_4$ . The expected response rates  $g_1, g_4$  can be found by calculating the formulae  $kr_{A+} + (1 - k)r_{B-}$  and  $kr_{B+} + (1 - k)r_{A-}$  respectively,  $r_{A+}, r_{B-}$  are the assumed response rates of biomarker-positive patients receiving the experimental treatment and biomarker-negative patients receiving the control treatment and  $r_{A-}, r_{B+}$  are the assumed response rates of biomarker-negative patients receiving the experimental treatment and biomarker-positive patients receiving the control treatment The prevalence of biomarker-positive patients corresponds to  $k$ .

**Statistical considerations:** This design enables the evaluation of the interaction between the biomarker and different treatments and can estimate directly the marker-strategy response rate. Additionally, this subtype design allows the estimation of the effect size of the experimental treatment compared to the control treatment for each biomarker-defined subgroup separately. Also, there is no chance that the same treatment will be tailored to biomarker-positive patients who are randomized either to the biomarker-based strategy arm or the reverse marker strategy (i.e., biomarker-positive patients in the biomarker-based strategy will be given only the experimental treatment and biomarker-positive patients in the reverse marker strategy arm will be given only the control treatment). Also, there is no possibility of the same treatment assignment to biomarker-negative patients who are randomly assigned to the two biomarker-based strategy arms (i.e., biomarker-negative patients in the marker-based strategy arm will be treated with the control treatment, whereas biomarker-negative patients in the reverse marker strategy arm will be treated with the experimental treatment). According to Eng, 2014 [92] who compared the reverse marker-based strategy design with the first (i.e., biomarker-strategy design with biomarker assessment in the control arm) and third (i.e., biomarker-strategy design with treatment randomization in the control arm) subtype of biomarker-strategy designs in the case of binary outcomes, the effect size in order to make a comparison of the different treatment strategy arms would be larger than in the first and third subtype designs. Furthermore, it has been shown by Eng, 2014 that in situations where a randomly chosen treatment has a better than 7% response rate, the reverse marker-based strategy design works better as compared to the third biomarker-strategy subtype (i.e., Biomarker-strategy design with treatment randomization in the control arm). It has also been demonstrated that this novel design is more than four times more efficient in order to test the interaction between treatment and biomarker compared to Biomarker-strategy design with biomarker assessment in the control arm, Biomarker-strategy design with randomization in the control arm and the marker stratified design. Eng, 2014 demonstrated the benefits of the Reverse Marker-Based strategy design with the aim to assess the interaction between treatment and biomarker. However, Baker, 2014 [93] stated that other designs than the Reverse Marker-Based strategy design would be more appropriate in order to investigate questions which include treatment effect of biomarker-defined subgroups and biomarker-based strategy arms.

2.5. Other Designs

A Randomized Phase II Trial Design with Biomarker Proposed by Freidlin et al., 2012

Freidlin et al., 2012 [71] proposed a biomarker-guided Phase II clinical trial design in which when it is completed, it recommends which type of Phase III trial should be used. These recommendations for a Phase III trial are the following: (i) enrichment design; (ii) marker-stratified design; (iii) a traditional trial design without a biomarker; or (iv) drop consideration of the experimental treatment. A graphical illustration of this design is given in Figure 16.



**Figure 16.** Randomized Phase II trial design with biomarkers. “R” refers to randomization of patients. CI refers to the confidence interval. Uncolored boxes are referred to the first stage of the trial and colored boxes are referred to the second stage of the trial.

**Design:** For this type of randomized Phase II trial, it is assumed that the experimental treatment will be more beneficial among biomarker-positive patients than biomarker-negative patients without ruling out the efficacy of the novel treatment in biomarker-negative patients. The intermediate endpoint of progression-free survival (PFS) is used which is able not only to give the results earlier but also to target larger treatment effects as compared to overall survival (OS) endpoint.

The design starts by comparing the experimental treatment with the control treatment in the biomarker-positive subgroup using a one-sided level of significance  $\alpha_1 = 0.10$ . The null hypothesis is that the progression-free survival for biomarker-positive patients is the same for both experimental and control treatment arm ( $HR_{0, biom+} \leq 1$  vs.  $HR_{1, biom+} > 1$ ). Next, if the null hypothesis is rejected, which means that the experimental treatment is better than the control treatment in the biomarker-positive subgroup we continue with the calculation of an 80% two-sided confidence interval (CI) for the hazard ratio (control vs experimental) in the biomarker-negative subpopulation. Three decisions are made according to the values of the CI: (i) if the entire CI is less than 1.3 then we can continue with a Phase III enrichment design; (ii) if the CI includes the values 1.3 or 1.5 then we can continue with a Phase III marker-stratified design and (iii) if the entire CI is greater than 1.5 then it seems that the biomarker is not useful as the novel treatment benefits only the biomarker-negative patients, thus, the biomarker should be dropped and a traditional randomized Phase III design should be conducted. Otherwise, if the null hypothesis is not rejected at the one-sided significance  $\alpha_1 = 0.10$

(meaning that that the experimental treatment is not better than the experimental treatment in the biomarker-positive subgroup), then we continue with the comparison of treatments in the overall study population at one-sided level of significance  $\alpha = 0.05$ . If the null hypothesis of no treatment effect in the entire population is rejected, then the authors recommend to drop the biomarker and to continue with a traditional randomized Phase III trial due to the fact that the biomarker seems to be useless. On the other hand, if the null hypothesis is not rejected, the experimental treatment should not be tested further as it does not seem to be effective.

**Utility:** This design should be used when we want to conduct a Phase II randomized trial which allows decisions to be made about which type of Phase III biomarker-guided trial to proceed with. It is appropriate when there is prior evidence that the novel treatment benefits mostly the biomarker-positive patients without ruling out treatment effect in biomarker-negative patients.

**Methodology:** Freidlin et al., 2012 [71] have provided an online tool for calculating the sample size which can be found on the following website <http://brb.nci.nih.gov/Data/FreidlinB/RP2BM> [116]. In order for a sample size to be estimated, the following information is required: (i) the significance levels for testing the treatment effect in the biomarker-positive subgroup and in the entire population; (ii) cut-offs and confidence intervals for the hazard ratio in the biomarker-negative subgroup; (iii) the prevalence of biomarker-positive patients; (iv) the median progression-free survival in each treatment arm in each biomarker-defined subgroup and (v) the accrual parameters. Regarding the accrual parameters, the author specifies the minimum sample size for biomarker-positive patients for which the accrual continues until this number is reached, the maximum number of over-accrual in biomarker-positive subgroup for which the accrual to the entire population stops after this number is reached and the maximum accrual number in biomarker-negative patients for which the accrual to this biomarker-defined subgroup stops when this number is reached.

**Statistical considerations:** In real life, it might not be possible to obtain the biomarker status for the entire population. If the biomarker status is unknown for some patients, then these individuals could be included in the analysis of the overall population. More precisely, in case that the proportions of patients with unknown biomarker status is low, the randomization of them to either the experimental or the control treatment could be considered in the second stage of this Phase II trial where we test the treatment effectiveness in the entire population. Another statistical consideration is that researchers should take into account the adjustment for inflation in Phase III type I error as the chosen Phase III trial design depends on the performance of the aforementioned randomized Phase II trial. Additionally, the authors suggest generally that in cases where it seems that the control treatment has been shown more beneficial, an aggressive interim inefficacy/futility should be used, i.e., when the estimated hazard ratio of control treatment versus the experimental treatment is equal or less than one when half of the required number of events have been observed, then the accrual should stop to that biomarker-defined subgroup.

### 3. Discussion

A number of biomarker-guided trial designs have been proposed in the past decade, including both biomarker-guided adaptive and non-adaptive trial designs. We have undertaken a comprehensive review of the literature using an in-depth search strategy to report on the biomarker-guided designs proposed to date, with a view to providing the research community with clarity in definition, methodology and terminology of the various trial designs. The review is split in two parts due to its size; the first part of the review is focused on adaptive designs which are extensively discussed in our published paper "Biomarker-Guided Adaptive Trial Designs in Phase II and Phase III: a Methodological Review", Antoniou et al., 2016 [35], whereas, herein we focus on non-adaptive designs which incorporate biomarkers.

The review has demonstrated ambiguity and confusion regarding the biomarker-guided non-adaptive designs proposed by different authors. In this review, we focus on 5 main types of such designs including their subtypes and variations. Knowledge on how to implement and analyse these

designs are essential in testing the effectiveness of a biomarker-guided approach to treatment; hence, a comprehensive review giving this knowledge is essential for the research community. In our in-depth study, we provide researchers with analytical information of these study designs not only in terms of their utility, advantages and limitations but also in terms of their methodology. In addition, a graphical illustration for each biomarker-guided design is given. A guidance document by Tajik et al., 2012 [117] regarding the evaluation of putative biomarkers in randomized clinical trials came to our knowledge by personal communication as we were not able to identify it during our literature search.

The non-adaptive designs do not allow modifications of important aspects of the trial such as refinement of the existing study population, treatment assignment, study endpoints, study duration, etc. In non-adaptive designs, all these factors are defined before the initiation of the study and they are kept fixed during the course of the clinical trial. However, there is a great potential of failure when implementing such conventional designs due to potential wrong design assumptions of the key aspects of the study that might be made before the conduct of the trial. Hence, an adaptive design clinical study which allows on-going adaptations based on accumulating study data from interim analysis might hold advantageous position as compared to the non-adaptive trial design due to its flexibility. However, before implementing an adaptive design a lot of issues should be taken into careful consideration by research teams in order to prove that there are good reasons for conducting such designs. Regulatory and logistical issues, requirement of additional efforts for the achievement of the design, potential difficulties, possible increased cost and time, statistical challenges including the potential increase of the chance of a false conclusion that the treatment is effective (inflation of Type I error) and whether the adaptation process has led to positive study results that are difficult to interpret irrespective of having control of Type I error should be considered [130]. A recent paper by Dimairo et al., 2015 [131] refers to a number of obstacles and barriers when implementing adaptive designs in practice. Several key stakeholders in clinical trials research have been interviewed (i.e., UK Clinical Trials Units directors, funding board and panel members, statisticians, regulators, chief investigators, data monitoring committee members and health economists) expressing difficulties of adaptive designs. Lack of appropriate knowledge and familiarity of these designs in the scientific community, insufficient time and funding structure, additional work required due to the complexity of such designs and the needed statistical expertise and appropriate software are some of the highlighted difficulties mentioned in the paper of Dimairo et al., 2015 [131]. In addition, this study includes the characterisation of potential benefits of an adaptive design to patients, clinical trials as well as funders.

The different designs proposed so far for biomarker-guided designs, both non-adaptive designs which remain an appealing approach to a great extent mainly due to their simplicity and adaptive designs which are more flexible need to be further explored by the research community, as the proper choice and use of such designs can result in a great increase in the efficiency of a trial and expedite the development of novel treatments.

The characteristics and methodology of the five main designs and their subtypes are discussed in the current paper, whilst information on their variations are summarized in File S1–S4. Additional references for these variations and the literature review search strategy are provided in [132,133].

**Supplementary Materials:** The following are available online at [www.mdpi.com/2075-4426/7/1/1/s1](http://www.mdpi.com/2075-4426/7/1/1/s1), File S1–S4: Extensions of Biomarker-guided non-adaptive trial designs, Keywords S1: Literature review search strategies for both biomarker-guided clinical trial designs and for traditional trial designs.

**Acknowledgments:** The authors express their gratitude to the two anonymous reviewers for their suggestions.

**Author Contributions:** M.A., R.K.-D. and A.L.J. conceived and designed the experiments; M.A. performed the experiments; M.A. analyzed the data; M.A., R.K.-D. and A.L.J. contributed reagents/materials/analysis tools; M.A., R.K.-D. and A.L.J. wrote the paper. First author (M.A.) screened the available titles and abstracts, and second and third authors (R.K.-D., A.L.J.) were consulted where it was questionable whether a paper should be included or not. A data extraction form was designed to collect all necessary information, and the summary of the extracted data was reviewed by the second and third authors (R.K.-D., A.L.J.).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. George, S.L. Statistical issues in translational cancer research. *Clin. Cancer Res.* **2008**, *14*, 5954–5958. [[CrossRef](#)] [[PubMed](#)]
2. Chabner, B. Advances and challenges in the use of biomarkers in clinical trials. *Clin. Adv. Hematol. Oncol.* **2008**, *6*, 42–43. [[PubMed](#)]
3. Group, B.D.W. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **2001**, *69*, 89–95.
4. Shi, Q.; Mandrekar, S.J.; Sargent, D.J. Predictive biomarkers in colorectal cancer: Usage, validation, and design in clinical trials. *Scand. J. Gastroenterol.* **2012**, *47*, 356–362. [[CrossRef](#)] [[PubMed](#)]
5. Pihlstrom, B.L.; Barnett, M.L. Design, operation, and interpretation of clinical trials. *J. Dent. Res.* **2010**, *89*, 759–772. [[CrossRef](#)] [[PubMed](#)]
6. Rigatto, C.; Barrett, B.J. Biomarkers and surrogates in clinical studies. *Methods Mol. Biol.* **2009**, *473*, 137–154. [[PubMed](#)]
7. Mandrekar, S.J.; An, M.-W.; Sargent, D.J. A review of phase II trial designs for initial marker validation. *Contemp. Clin. Trials* **2013**, *36*, 597–604. [[CrossRef](#)] [[PubMed](#)]
8. Karuri, S.W.; Simon, R. A two-stage bayesian design for co-development of new drugs and companion diagnostics. *Stat. Med.* **2012**, *31*, 901–914. [[CrossRef](#)] [[PubMed](#)]
9. Matsui, S. Genomic biomarkers for personalized medicine: Development and validation in clinical studies. *Comput. Math. Methods Med.* **2013**, *2013*, 865980. [[CrossRef](#)] [[PubMed](#)]
10. Buyse, M.; Michiels, S. Omics-based clinical trial designs. *Curr. Opin. Oncol.* **2013**, *25*, 289–295. [[CrossRef](#)] [[PubMed](#)]
11. Wu, W.; Shi, Q.; Sargent, D.J. Statistical considerations for the next generation of clinical trials. *Semin. Oncol.* **2011**, *38*, 598–604. [[CrossRef](#)] [[PubMed](#)]
12. Sargent, D.J.; Conley, B.A.; Allegra, C.; Collette, L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J. Clin. Oncol.* **2005**, *23*, 2020–2027. [[CrossRef](#)] [[PubMed](#)]
13. Chen, J.J.; Lu, T.-P.; Chen, D.-T.; Wang, S.-J. Biomarker adaptive designs in clinical trials. *Transl. Cancer Res.* **2014**, *3*, 279–292.
14. Freidlin, B.; Sun, Z.; Gray, R.; Korn, E.L. Phase III clinical trials that integrate treatment and biomarker evaluation. *J. Clin. Oncol.* **2013**, *31*, 3158–3161. [[CrossRef](#)] [[PubMed](#)]
15. Gosho, M.; Nagashima, K.; Sato, Y. Study designs and statistical analyses for biomarker research. *Sensors* **2012**, *12*, 8966–8986. [[CrossRef](#)] [[PubMed](#)]
16. Ming-Wen An, S.J.M.; Daniel, J.S. Biomarkers-guided targeted drugs: New clinical trials design and practice necessity. *Adv. Personal. Cancer Manag.* **2011**, 30–41.
17. Buyse, M. Towards validation of statistically reliable biomarkers. *Eur. J. Cancer Suppl.* **2007**, *5*, 89–95. [[CrossRef](#)]
18. Lee, C.K.; Lord, S.J.; Coates, A.S.; Simes, R.J. Molecular biomarkers to individualise treatment: Assessing the evidence. *Med. J. Aust.* **2009**, *190*, 631–636. [[PubMed](#)]
19. Simon, R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Personal. Med.* **2010**, *7*, 33–47. [[CrossRef](#)] [[PubMed](#)]
20. Fraser, G.A.M.; Meyer, R.M. Biomarkers and the design of clinical trials in cancer. *Biomark. Med.* **2007**, *1*, 387–397. [[CrossRef](#)] [[PubMed](#)]
21. Mandrekar, S.J.; Sargent, D.J. Design of clinical trials for biomarker research in oncology. *Clin. Investig.* **2011**, *1*, 1629–1636. [[CrossRef](#)] [[PubMed](#)]
22. Simon, R. Advances in clinical trial designs for predictive biomarker discovery and validation. *Curr. Breast Cancer Rep.* **2009**, *1*, 216–221. [[CrossRef](#)]
23. Polley, M.-Y.C.; Freidlin, B.; Korn, E.L.; Conley, B.A.; Abrams, J.S.; McShane, L.M. Statistical and practical considerations for clinical evaluation of predictive biomarkers. *J. Natl. Cancer Inst.* **2013**, *105*, 1677–1683. [[CrossRef](#)] [[PubMed](#)]
24. Bradley, E. Incorporating biomarkers into clinical trial designs: Points to consider. *Nat. Biotechnol.* **2012**, *30*, 596–599. [[CrossRef](#)] [[PubMed](#)]
25. Beckman, R.A.; Clark, J.; Chen, C. Integrating predictive biomarkers and classifiers into oncology clinical development programmes. *Nat. Rev. Drug Discov.* **2011**, *10*, 735–748. [[CrossRef](#)] [[PubMed](#)]

26. Young, K.Y.; Laird, A.; Zhou, X.H. The efficiency of clinical trial designs for predictive biomarker validation. *Clin. Trials* **2010**, *7*, 557–566. [[CrossRef](#)] [[PubMed](#)]
27. Lee, J.J.; Xuemin, G.; Suyu, L. Bayesian adaptive randomization designs for targeted agent development. *Clin. Trials* **2010**, *7*, 584–596. [[CrossRef](#)] [[PubMed](#)]
28. Simon, R. Clinical trials for predictive medicine: New challenges and paradigms. *Clin. Trials* **2010**, *7*, 516–524. [[CrossRef](#)] [[PubMed](#)]
29. Buyse, M.; Sargent, D.J.; Grothey, A.; Matheson, A.; de Gramont, A. Biomarkers and surrogate end points—The challenge of statistical validation. *Nat. Rev. Clin. Oncol.* **2010**, *7*, 309–317. [[CrossRef](#)] [[PubMed](#)]
30. Mandrekar, S.J.; Sargent, D.J. Clinical trial designs for predictive biomarker validation: Theoretical considerations and practical challenges. *J. Clin. Oncol.* **2009**, *27*, 4027–4034. [[CrossRef](#)] [[PubMed](#)]
31. Mandrekar, S.J.; Sargent, D.J. Clinical trial designs for predictive biomarker validation: One size does not fit all. *J. Biopharm. Stat.* **2009**, *19*, 530–542. [[CrossRef](#)] [[PubMed](#)]
32. Hoering, A.; Leblanc, M.; Crowley, J.J. Randomized phase III clinical trial designs for targeted agents. *Clin. Cancer Res.* **2008**, *14*, 4358–4367. [[CrossRef](#)] [[PubMed](#)]
33. Kelloff, G.J.; Sigman, C.C. Cancer biomarkers: Selecting the right drug for the right patient. *Nat. Rev. Drug Discov.* **2012**, *11*, 201–214. [[CrossRef](#)] [[PubMed](#)]
34. Chow, S.-C. Adaptive clinical trial design. *Annu. Rev. Med.* **2014**, *65*, 405–415. [[CrossRef](#)] [[PubMed](#)]
35. Antoniou, M.; Jorgensen, A.L.; Kolamunnage-Dona, R. Biomarker-guided adaptive trial designs in phase II and phase III: A methodological review. *PLoS ONE* **2016**, *11*, e0149803. [[CrossRef](#)] [[PubMed](#)]
36. Tajik, P.; Zwinderman, A.H.; Mol, B.W.; Bossuyt, P.M. Trial designs for personalizing cancer care: A systematic review and classification. *Clin. Cancer Res.* **2013**, *19*, 4578–4588. [[CrossRef](#)] [[PubMed](#)]
37. Lader, E.W.; Cannon, C.P.; Ohman, E.M.; Newby, L.K.; Sulmasy, D.P.; Barst, R.J.; Fair, J.M.; Flather, M.; Freedman, J.E.; Frye, R.L.; et al. The clinician as investigator: Participating in clinical trials in the practice setting: Appendix 1: Fundamentals of study design. *Circulation* **2004**, *109*, e302–e304. [[CrossRef](#)] [[PubMed](#)]
38. Stingl Kirchheiner, J.C.; Brockmüller, J. Why, when, and how should pharmacogenetics be applied in clinical studies? Current and future approaches to study designs. *Clin. Pharm. Ther.* **2011**, *89*, 198–209. [[CrossRef](#)] [[PubMed](#)]
39. Sambucini, V. A bayesian predictive two-stage design for phase II clinical trials. *Stat. Med.* **2008**, *27*, 1199–1224. [[CrossRef](#)] [[PubMed](#)]
40. Ang, M.-K.; Tan, S.-B.; Lim, W.-T. Phase II clinical trials in oncology: Are we hitting the target? *Expert Rev. Anticancer Ther.* **2010**, *10*, 427–438. [[CrossRef](#)] [[PubMed](#)]
41. Farley, J.; Rose, P.G.; Farley, J.; Rose, P.G. Trial design for evaluation of novel targeted therapies. *Gynecol. Oncol.* **2010**, *116*, 173. [[CrossRef](#)] [[PubMed](#)]
42. Kaplan, R.; Maughan, T.; Crook, A.; Fisher, D.; Wilson, R.; Brown, L.; Parmar, M. Evaluating many treatments and biomarkers in oncology: A new design. *J. Clin. Oncol.* **2013**, *31*, 4562–4568. [[CrossRef](#)] [[PubMed](#)]
43. Hodgson, D.R.; Wellings, R.; Harbron, C. Practical perspectives of personalized healthcare in oncology. *New Biotechnol.* **2012**, *29*, 656–664. [[CrossRef](#)] [[PubMed](#)]
44. Mandrekar, S.J.; Sargent, D.J. Predictive biomarker validation in practice: Lessons from real trials. *Clin. Trials* **2010**, *7*, 567–573. [[CrossRef](#)] [[PubMed](#)]
45. Galanis, E.; Wu, W.; Sarkaria, J.; Chang, S.M.; Colman, H.; Sargent, D.; Reardon, D.A. Incorporation of biomarker assessment in novel clinical trial designs: Personalizing brain tumor treatments. *Curr. Oncol. Rep.* **2011**, *13*, 42–49. [[CrossRef](#)] [[PubMed](#)]
46. Van Schaeuybroeck, S.; Allen, W.L.; Turkington, R.C.; Johnston, P.G. Implementing prognostic and predictive biomarkers in CRC clinical trials. *Nat. Rev. Clin. Oncol.* **2011**, *8*, 222–232. [[CrossRef](#)] [[PubMed](#)]
47. Buyse, M.; Michiels, S.; Sargent, D.J.; Grothey, A.; Matheson, A.; de Gramont, A. Integrating biomarkers in clinical trials. *Expert rev. Mol. Diagn.* **2011**, *11*, 171–182. [[CrossRef](#)] [[PubMed](#)]
48. Sparano, J.A.; Paik, S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J. Clin. Oncol.* **2008**, *26*, 721–728. [[CrossRef](#)] [[PubMed](#)]
49. Freidlin, B.; Korn, E.L. Biomarker enrichment strategies: Matching trial design to biomarker credentials. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 81–90. [[CrossRef](#)] [[PubMed](#)]
50. Simon, R.; Polley, E. Clinical trials for precision oncology using next-generation sequencing. *Personal. Med.* **2013**, *10*, 485–495. [[CrossRef](#)]

51. Baker, S.G.; Kramer, B.S.; Sargent, D.J.; Bonetti, M. Biomarkers, subgroup evaluation, and clinical trial design. *Discov. Med.* **2012**, *13*, 187–192. [[PubMed](#)]
52. Buch, M.H.; Pavitt, S.; Parmar, M.; Emery, P. Creative trial design in RA: Optimizing patient outcomes. *Nat. Rev. Rheumatol.* **2013**, *9*, 183–194. [[CrossRef](#)] [[PubMed](#)]
53. Simon, R. Clinical trials for predictive medicine. *Stat. Med.* **2012**, *31*, 3031–3040. [[CrossRef](#)] [[PubMed](#)]
54. Scher, H.I.; Nasso, S.F.; Rubin, E.H.; Simon, R. Adaptive clinical trial designs for simultaneous testing of matched diagnostics and therapeutics. *Clin. Cancer Res.* **2011**, *17*, 6634–6640. [[CrossRef](#)] [[PubMed](#)]
55. Sato, Y.; Laird, N.M.; Yoshida, T. Biostatistic tools in pharmacogenomics—Advances, challenges, potential. *Curr. Pharm. Des.* **2010**, *16*, 2232–2240. [[CrossRef](#)] [[PubMed](#)]
56. Mandrekar, S.J.; Sargent, D.J.; Mandrekar, S.; Sargent, D. Genomic advances and their impact on clinical trial design. *Genome Med.* **2009**, *1*, 69. [[CrossRef](#)] [[PubMed](#)]
57. Simon, R. Designs and adaptive analysis plans for pivotal clinical trials of therapeutics and companion diagnostics. *Expert Opin. Med. Diagn.* **2008**, *2*, 721–729. [[CrossRef](#)] [[PubMed](#)]
58. Dobbin, K.K. Statistical design and evaluation of biomarker studies. *Methods Mol. Biol.* **2014**, *1102*, 667–677. [[PubMed](#)]
59. Ananthkrishnan, R.; Menon, S. Design of oncology clinical trials: A review. *Crit. Rev. Oncol./Hematol.* **2013**, *88*, 144–153. [[CrossRef](#)] [[PubMed](#)]
60. Simon, R. The use of genomics in clinical trial design. *Clin. Cancer Res.* **2008**, *14*, 5984–5993. [[CrossRef](#)] [[PubMed](#)]
61. Freidlin, B.; McShane, L.M.; Korn, E.L. Randomized clinical trials with biomarkers: Design issues. *J. Natl. Cancer Inst.* **2010**, *102*, 152–160. [[CrossRef](#)] [[PubMed](#)]
62. Johnson, D.R.; Galanis, E. Incorporation of prognostic and predictive factors into glioma clinical trials. *Curr. Oncol. Rep.* **2013**, *15*, 56–63. [[CrossRef](#)] [[PubMed](#)]
63. Sparano, J. TAILORx: Trial assigning individualized options for treatment (Rx). *Clin. Breast Cancer* **2006**, *7*, 347–350. [[CrossRef](#)] [[PubMed](#)]
64. Di Maio, M.; Gallo, C.; De Maio, E.; Morabito, A.; Piccirillo, M.C.; Gridelli, C.; Perrone, F. Methodological aspects of lung cancer clinical trials in the era of targeted agents. *Lung Cancer* **2010**, *67*, 127–135. [[CrossRef](#)] [[PubMed](#)]
65. Maitournam, A.; Simon, R. On the efficiency of targeted clinical trials. *Stat. Med.* **2005**, *24*, 329–339. [[CrossRef](#)] [[PubMed](#)]
66. Collette, L.; Bogaerts, J.; Suci, S.; Fortpied, C.; Gorlia, T.; Coens, C.; Mauer, M.; Hasan, B.; Collette, S.; Ouali, M.; et al. Statistical methodology for personalized medicine: New developments at EORTC headquarters since the turn of the 21st century. *Eur. J. Cancer Suppl.* **2012**, *10*, 13. [[CrossRef](#)]
67. Mandrekar, S.J.; Sargent, D.J. All-comers versus enrichment design strategy in phase II trials. *J. Thorac. Oncol.* **2011**, *6*, 658–660. [[CrossRef](#)] [[PubMed](#)]
68. Simon, R. Development and validation of biomarker classifiers for treatment selection. *J. Stat. Plan. Inference* **2008**, *138*, 308–320. [[CrossRef](#)] [[PubMed](#)]
69. Freidlin, B.; Korn, E.L.; Gray, R. Marker sequential test (MaST) design. *Clin. Trials* **2014**, *11*, 19–27. [[CrossRef](#)] [[PubMed](#)]
70. Wason, J.; Marshall, A.; Dunn, J.; Stein, R.C.; Stallard, N. Adaptive designs for clinical trials assessing biomarker-guided treatment strategies. *Br. J. Cancer* **2014**, *110*, 1950–1957. [[CrossRef](#)] [[PubMed](#)]
71. Freidlin, B.; McShane, L.M.; Polley, M.-Y.C.; Korn, E.L. Randomized phase II trial designs with biomarkers. *J. Clin. Oncol.* **2012**, *30*, 3304–3309. [[CrossRef](#)] [[PubMed](#)]
72. Ziegler, A.; Koch, A.; Krockenberger, K.; Grosshennig, A. Personalized medicine using DNA biomarkers: A review. *Hum. Genet.* **2012**, *131*, 1627–1638. [[CrossRef](#)] [[PubMed](#)]
73. Freidlin, B.; Korn, E.L. Biomarker-adaptive clinical trial designs. *Pharmacogenomics* **2010**, *11*, 1679–1682. [[CrossRef](#)] [[PubMed](#)]
74. Eickhoff, J.C.; Kim, K.; Beach, J.; Kolesar, J.M.; Gee, J.R. A bayesian adaptive design with biomarkers for targeted therapies. *Clin. Trials* **2010**, *7*, 546–556. [[CrossRef](#)] [[PubMed](#)]
75. Ferraldeschi, R.; Attard, G.; de Bono, J.S. Novel strategies to test biological hypotheses in early drug development for advanced prostate cancer. *Clin. Chem.* **2013**, *59*, 75–84. [[CrossRef](#)] [[PubMed](#)]
76. Coyle, V.M.; Johnston, P.G. Genomic markers for decision making: What is preventing us from using markers? *Nat. Rev. Clin. Oncol.* **2010**, *7*, 90–97. [[CrossRef](#)] [[PubMed](#)]

77. Chen, C.F.; Lin, J.R.; Liu, J.P. Statistical inference on censored data for targeted clinical trials under enrichment design. *Pharm. Stat.* **2013**, *12*, 165–173. [[CrossRef](#)] [[PubMed](#)]
78. Liu, J.P.; Lin, J.R. Statistical methods for targeted clinical trials under enrichment design. *J. Formos. Med. Assoc.* **2008**, *107*, 35–42. [[CrossRef](#)]
79. Scheibler, F.; Zumbé, P.; Janssen, I.; Viebahn, M.; Schröer-Günther, M.; Grosselfinger, R.; Hausner, E.; Sauerland, S.; Lange, S. Randomized controlled trials on pet: A systematic review of topics, design, and quality. *J. Nucl. Med.* **2012**, *53*, 1016–1025. [[CrossRef](#)] [[PubMed](#)]
80. An, M.-W.; Mandrekar, S.J.; Sargent, D.J. A 2-stage phase II design with direct assignment option in stage II for initial marker validation. *Clin. Cancer Res.* **2012**, *18*, 4225–4233. [[CrossRef](#)] [[PubMed](#)]
81. Zheng, G.; Wu, C.O.; Yang, S.; Waclawiw, M.A.; DeMets, D.L.; Geller, N.L. NHLBI clinical trials workshop: An executive summary. *Stat. Med.* **2012**, *31*, 2938. [[CrossRef](#)] [[PubMed](#)]
82. Bria, E.; Di Maio, M.; Carlini, P.; Cuppone, F.; Giannarelli, D.; Cognetti, F.; Milella, M. Targeting targeted agents: Open issues for clinical trial design. *J. Exp. Clin. Cancer Res.* **2009**, *28*, 66. [[CrossRef](#)] [[PubMed](#)]
83. French, B.; Joo, J.; Geller, N.L.; Kimmell, S.E.; Rosenberg, Y.; Anderson, J.L.; Gage, B.F.; Johnson, J.A.; Ellenberg, J.H. Statistical design of personalized medicine interventions: The Clarification of Optimal Anticoagulation through Genetics (Coag) trial. *Trials* **2010**, *11*, 108. [[CrossRef](#)] [[PubMed](#)]
84. Lin, J.-A.; He, P. Reinventing clinical trials: A review of innovative biomarker trial designs in cancer therapies. *Br. Med. Bull.* **2015**, *114*, 17–27. [[CrossRef](#)] [[PubMed](#)]
85. Renfro, L.A.; Mallick, H.; An, M.-W.; Sargent, D.J.; Mandrekar, S.J. Clinical trial designs incorporating predictive biomarkers. *Cancer Treat. Rev.* **2016**, *43*, 74–82. [[CrossRef](#)] [[PubMed](#)]
86. Ondra, T.; Dmitrienko, A.; Friede, T.; Graf, A.; Miller, F.; Stallard, N.; Posch, M. Methods for identification and confirmation of targeted subgroups in clinical trials: A systematic review. *J. Biopharm. Stat.* **2016**, *26*, 99–119. [[CrossRef](#)] [[PubMed](#)]
87. Simon, R.; Wang, S.J. Use of genomic signatures in therapeutics development in oncology and other diseases. *Pharmacogenom. J.* **2006**, *6*, 166–173. [[CrossRef](#)] [[PubMed](#)]
88. European Medicines Agency. Reflection Paper on Methodological Issues Associated with Pharmacogenomic Biomarkers in Relation to Clinical Development and Patient Selection. Available online: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2011/07/WC500108672.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf) (accessed on 10 October 2015).
89. Lai, T.L.; Liao, O.Y.-W.; Kim, D.W. Group sequential designs for developing and testing biomarker-guided personalized therapies in comparative effectiveness research. *Contemp. Clin. Trials* **2013**, *36*, 651–663. [[CrossRef](#)] [[PubMed](#)]
90. Foley, R.N. Analysis of randomized controlled clinical trials. *Methods Mol. Biol.* **2009**, *473*, 113–126. [[PubMed](#)]
91. Tajik, P.; Bossuyt, P.M. Genomic markers to tailor treatments: Waiting or initiating? *Hum. Genet.* **2011**, *130*, 15–18. [[CrossRef](#)] [[PubMed](#)]
92. Eng, K.H. Randomized reverse marker strategy design for prospective biomarker validation. *Stat. Med.* **2014**, *33*, 3089–3099. [[CrossRef](#)] [[PubMed](#)]
93. Baker, S.G. Biomarker evaluation in randomized trials: Addressing different research questions. *Stat. Med.* **2014**, *33*, 4139–4140. [[CrossRef](#)] [[PubMed](#)]
94. Matsui, S.; Choai, Y.; Nonaka, T. Comparison of statistical analysis plans in randomize-all phase III trials with a predictive biomarker. *Clin. Cancer Res.* **2014**, *20*, 2820–2830. [[CrossRef](#)] [[PubMed](#)]
95. Cappuzzo, F.; Ciuleanu, T.; Stelmakh, L.; Cicenias, S.; Szczésna, A.; Juhász, E.; Esteban, E.; Molinier, O.; Brugger, W.; Melezínek, I.; et al. Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: A multicentre, randomised, placebo-controlled phase 3 study. *Lancet Oncol.* **2010**, *11*, 521. [[CrossRef](#)]
96. Hoffmann-La Roche. A Randomized, Double-Blind Study to Evaluate the Effect of Tarceva or Placebo Following Platinum-Based CT on Overall Survival and Disease Progression in Patients with Advanced, Recurrent or Metastatic NSCLS Who Have Not Experienced Disease Progression or Unacceptable Toxicity during Chemotherapy. Available online: <https://clinicaltrials.gov/ct2/show/NCT00556712?term=NCT00556712&rank=1> (accessed on 10 October 2015).
97. Choai, Y.; Matsui, S. Estimation of treatment effects in all-comers randomized clinical trials with a predictive marker: Estimating treatment effects in marker-based randomized trials. *Biometrics* **2015**, *71*, 25. [[CrossRef](#)] [[PubMed](#)]

98. Wang, S.-J.; O'Neill, R.T.; Hung, H.M.J. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm. Stat.* **2007**, *6*, 227–244. [[CrossRef](#)] [[PubMed](#)]
99. Cree, I.A.; Kurbacher, C.M.; Lamont, A.; Hindley, A.C.; Love, S. A prospective randomized controlled trial of tumour chemosensitivity assay directed chemotherapy versus physician's choice in patients with recurrent platinum-resistant ovarian cancer. *Anti-Cancer Drugs* **2007**, *18*, 1093. [[CrossRef](#)] [[PubMed](#)]
100. Cobo, M.; Isla, D.; Massuti, B.; Montes, A.; Sanchez, J.M.; Provencio, M.; Viñolas, N.; Paz-Ares, L.; Lopez-Vivanco, G.; Muñoz, M.A.; et al. Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: A phase III trial in non-small-cell lung cancer. *J. Clin. Oncol.* **2007**, *25*, 2747. [[CrossRef](#)] [[PubMed](#)]
101. Lijmer, J.G.; Bossuyt, P.M.M. Various randomized designs can be used to evaluate medical tests. *J. Clin. Epidemiol.* **2009**, *62*, 364. [[CrossRef](#)] [[PubMed](#)]
102. Wang, S.-J. Biomarker as a classifier in pharmacogenomics clinical trials: A tribute to 30th anniversary of PSI. *Pharm. Stat.* **2007**, *6*, 283–296. [[CrossRef](#)] [[PubMed](#)]
103. Cho, D.; McDermott, D.; Atkins, M. Designing clinical trials for kidney cancer based on newly developed prognostic and predictive tools. *Curr. Urol. Rep.* **2006**, *7*, 8–15. [[CrossRef](#)] [[PubMed](#)]
104. Af Geijerstam, J.L.; Oredsson, S.; Britton, M. Medical outcome after immediate computed tomography or admission for observation in patients with mild head injury: Randomised controlled trial. *Br. Med. J.* **2006**, *333*, 465. [[CrossRef](#)] [[PubMed](#)]
105. Ferrante di Ruffano, L.; Davenport, C.; Eisinga, A.; Hyde, C.; Deeks, J.J. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J. Clin. Epidemiol.* **2012**, *65*, 282. [[CrossRef](#)] [[PubMed](#)]
106. Mandrekar, S.J.; Grothey, A.; Goetz, M.P.; Sargent, D.J. Clinical trial designs for prospective validation of biomarkers. *Am. J. Pharmacogenom.* **2005**, *5*, 317–325. [[CrossRef](#)]
107. Therasse, P.; Carbonnelle, S.; Bogaerts, J. Clinical trials design and treatment tailoring: General principles applied to breast cancer research. *Crit. Rev. Oncol. Hematol.* **2006**, *59*, 98–105. [[CrossRef](#)] [[PubMed](#)]
108. Sargent, D.; Allegra, C. Issues in clinical trial design for tumor marker studies. *Semin. Oncol.* **2002**, *29*, 222–230. [[CrossRef](#)] [[PubMed](#)]
109. Tanniou, J.; van der Tweel, I.; Teerenstra, S.; Roes, K.C.B. Subgroup analyses in confirmatory clinical trials: Time to be specific about their purposes. *BMC Med. Res. Methodol.* **2016**, *16*, 20. [[CrossRef](#)] [[PubMed](#)]
110. Rubinstein, L.V.; Gail, M.H.; Santner, T.J. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *J. Chronic Dis.* **1981**, *34*, 469–479. [[CrossRef](#)]
111. Simon, R.; Maitournam, A. Evaluating the efficiency of targeted designs for randomized clinical trials: Supplement and correction. *Clin. Cancer Res.* **2006**, *12*, 3229. [[CrossRef](#)] [[PubMed](#)]
112. Simon, R.; Maitournam, A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin. Cancer Res.* **2004**, *10*, 6759. [[CrossRef](#)] [[PubMed](#)]
113. Biomarker targeted randomized design. Available online: <http://brb.nci.nih.gov/brb/samplesize/td.html> (accessed on 15 September 2016).
114. Harrington, R.A. Applied Bioinformatics and Biostatistics in Cancer Research. In *Designs for Clinical Trials: Perspectives on Current Issues*; Springer Science+Business Media, LLC: New York, NY, USA, 2012.
115. Biomarker stratified randomized design. Available online: <https://brb.nci.nih.gov/brb/samplesize/sdpap.html> (accessed on 15 September 2016).
116. Freidlin, B. Randomized phase ii trial designs with biomarkers. Available online: <http://brb.nci.nih.gov/Data/FreidlinB/RP2BM> (accessed on 15 September 2016).
117. Tajik, P.; Zwinderman, A.H.; Mol, B.W.; Bossuyt, P.M. Evaluating putative predictive biomarkers in randomized clinical trials. Available online: [http://www.zonmw.nl/fileadmin/documenten/DO\\_Farmacotherapie\\_Dure\\_Weesgeneesmiddelen/HTA\\_pharmacotherapy\\_predictive\\_markers\\_guidance\\_document.pdf](http://www.zonmw.nl/fileadmin/documenten/DO_Farmacotherapie_Dure_Weesgeneesmiddelen/HTA_pharmacotherapy_predictive_markers_guidance_document.pdf) (accessed on 15 September 2016).
118. Zaslavsky, B.G.; Scott, J. Sample size estimation in single-arm clinical trials with multiple testing under frequentist and bayesian approaches. *J. Biopharm. Stat.* **2012**, *22*, 819–835. [[CrossRef](#)] [[PubMed](#)]
119. Wittes, J. Sample size calculations for randomized controlled trials. *Epidemiol. Rev.* **2002**, *24*, 39–53. [[CrossRef](#)] [[PubMed](#)]
120. Collette, L. Chapman & Hall/CRC Texts in Statistical Science Series. In *Modelling Survival Data in Medical Research*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2003.

121. Kleinbaum, D.G.; Klein, M. *Statistics for Biology and Health. Survival Analysis: A Self-Learning Text*, 3rd ed.; Springer: New York, NY, USA, 2012.
122. Freedman, L.S. Tables of the number of patients required in clinical trials using the logrank test. *Stat. Med.* **1982**, *1*, 121–129. [[CrossRef](#)] [[PubMed](#)]
123. Schoenfeld, D.A. Sample-size formula for the proportional-hazards regression model. *Biometrics* **1983**, *39*, 499–503. [[CrossRef](#)] [[PubMed](#)]
124. Bland, J.M.; Altman, D.G. Multiple significance tests: The bonferroni method. *Br. Med. J.* **1995**, *310*, 170. [[CrossRef](#)]
125. Jiang, W.; Freidlin, B.; Simon, R. Biomarker-adaptive threshold design: A procedure for evaluating treatment with possible biomarker-defined subset effect. *J. Natl. Cancer Inst.* **2007**, *99*, 1036–1043. [[CrossRef](#)] [[PubMed](#)]
126. Wang, S.-J.; Hung, H.M.J.; O'Neill, R.T. Adaptive patient enrichment designs in therapeutic trials. *Biom. J.* **2009**, *51*, 358–374. [[CrossRef](#)] [[PubMed](#)]
127. Alosch, M.; Huque, M.F.; Alosch, M.; Huque, M.F. A flexible strategy for testing subgroups and overall population. *Stat. Med.* **2009**, *28*, 3. [[CrossRef](#)] [[PubMed](#)]
128. Spiessens, B.; Debois, M. Adjusted significance levels for subgroup analyses in clinical trials. *Contemp. Clin. Trials* **2010**, *31*, 647. [[CrossRef](#)] [[PubMed](#)]
129. Song, Y.; Chi, G.Y.H.; Song, Y.; Chi, G.Y.H. A method for testing a prespecified subgroup in clinical trials. *Stat. Med.* **2007**, *26*, 3535. [[CrossRef](#)] [[PubMed](#)]
130. Chang, M. Chapman & Hall/CRC Biostatistics Series. In *Adaptive Design Theory and Implementation Using SAS and R*, 2nd ed.; CRC Press: London, UK, 2014.
131. Dimairo, M.; Boote, J.; Julious, S.A.; Nicholl, J.P.; Todd, S. Missing steps in a staircase: A qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials. *Trials* **2015**, *16*, 430. [[CrossRef](#)] [[PubMed](#)]
132. Spira, A.; Edmiston, K.H. Clinical trial design in the age of molecular profiling. *Methods Mol. Biol.* **2012**, *823*, 19–34. [[PubMed](#)]
133. Medline Plus basic course manual 2012. Available online: [http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0ahUKEwjS7\\_OmodvJAhWGVhQKHZr0AZMQFggdMAA&url=http%3A%2F%2Fbma.org.uk%2F-%2Fmedia%2Ffiles%2Fpdfs%2Fabout%2520the%2520bma%2Flibrary%2Fmedline%2520plus%2520basic%2520course%2520manual%25202012.pdf&usg=AFQjCNGFxcWiS11CJsroeeIETAWjW0neUA](http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0ahUKEwjS7_OmodvJAhWGVhQKHZr0AZMQFggdMAA&url=http%3A%2F%2Fbma.org.uk%2F-%2Fmedia%2Ffiles%2Fpdfs%2Fabout%2520the%2520bma%2Flibrary%2Fmedline%2520plus%2520basic%2520course%2520manual%25202012.pdf&usg=AFQjCNGFxcWiS11CJsroeeIETAWjW0neUA) (accessed on 15 September 2016).



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).