1.0 ABSTRACT:

Biomarkers can be used to establish more homogeneous groups using the genetic makeup of the tumour to inform the selection of treatment for each individual patient. However, proper preclinical work and stringent validation are needed before taking forward biomarkers into confirmatory studies. Despite the challenges, incorporation of biomarkers into clinical trials could better target appropriate patients, and potentially be lifesaving. The authors conducted a systematic review to describe marker-based and adaptive design methodology for their integration in clinical trials, and to further describe the associated practical challenges. Studies published between 1990 to November 2015 were searched on PubMed. Titles, abstracts and full text articles were reviewed to identify relevant studies. Of the 4438 studies examined, 57 studies were included. The authors conclude that the proposed approaches may readily help researchers to design biomarker trials, but novel approaches are still needed.

2.0 INTRODUCTION:

Considerable challenges exist in the incorporation of biomarkers into clinical trials. This explains why they are mostly included as exploratory endpoints into current oncology clinical trials (1). Individual patient heterogeneity, both between primary and sites of metastasis as well as within metastatic lesions, is a major concern for successful treatment of advanced tumours (2). As patient biopsies often target a single piece of tissue at one time point only, and not at multiple ones longitudinally, tumour heterogeneity and alterations over time are not properly addressed, although they likely contribute to the evolution of drug resistance (2). Furthermore, biomarkers can represent molecular aberrations that can be driver or passenger events (2). Other issues include the percentage of cells and the method of obtaining a tumour sample, and in what sequence multi-combinatorial agents as well as their

dose levels should be used to target multiple aberrations (2). Despite this, biomarkers can be used to establish more homogeneous groups using the genetic makeup of the tumour to inform the selection of treatment for individual patients (3).

Biomarkers are classified into a few categories in the literature: prognostic, predictive, surrogate, screening or diagnostic, pharmacodynamic efficacy and resistance, and integral and integrated biomarkers (4-6). For the purposes of this article, we mainly focus on prognostic and predictive markers; with a brief overview of the others. A surrogate marker is a biomarker accepted by regulatory agencies as a substitute for a clinical endpoint and, when used as an early indicator of treatment efficacy, is potentially attractive in terms of cost-effectiveness (4); e.g. HIV load. Screening or diagnostic markers are used in the monitoring of disease including PSA levels in prostate cancer. Pharmacodynamic efficacy and resistance biomarkers are used to measure response and resistance to treatment, respectively (5). Finally, integral biomarkers determine patient incorporation and/or directs clinical trial procedures, while integrated biomarkers are not used to determine patient treatment (6). Prognostic markers provide an early indication of the clinical course of a patient independent of any specific intervention and may be considered in the clinical management of a patient; e.g. BRCA1/2 mutation-which can also be predictive of PARP inhibitors. These are prevalent in the literature, and guidelines for their evaluation are available with the gold standard being the REMARK criteria (7, 8). Predictive biomarkers are measured prior to an intervention and identify patients who are susceptible to a particular drug effect; however they are not necessarily prognostic of post-treatment clinical course (3), e.g. HER2 or KRAS (9). Predictive markers can only be properly validated in a prospectively designed randomized controlled trial testing for a marker-by-treatment interaction (10); but a very large sample size is often required (11). A biomarker can be both prognostic and predictive

such as Estrogen Receptor status and its prognostic association with relapse and its predictiveness of treatment benefit from tamoxifen (12).

It is critical that proper preclinical work and stringent validation be done before taking forward only the most promising biomarkers into confirmatory studies. The aim of this article is to provide an overview on the methods to incorporate biomarkers into clinical trials and to further describe the challenges.

3.0 MATERIALS AND METHODS:

Study selection followed the process described in the diagram in Figure 1. The design name, whether they are marker-based, adaptive, used in design in or testing during a trial, their description, advantages and disadvantages and trials using those designs were retrieved.

4.0 RESULTS:

Figure 1:

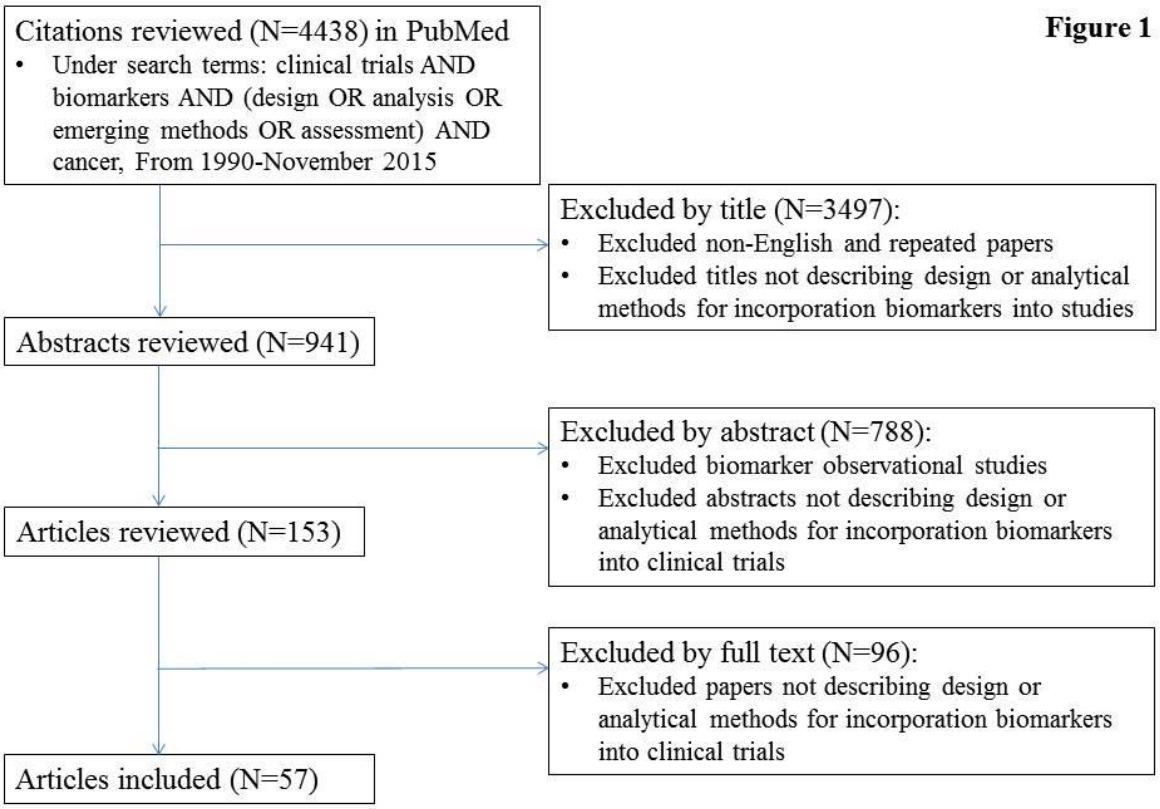

Figure 1

Fifty-seven articles were included in the review, and methods of incorporating biomarkers were identified (Table 1). Broadly, the methods fall into marker-based or adaptive, and being used as design or testing methods; and other novel designs.

4.1. Overview of marker-based methods:

One of the most commonly used marker-based designs is the enrichment or targeted design (Figure 2a), which is appropriate when there is compelling preliminary evidence to suggest that treatment benefit or lack of toxicity is restricted to patients with a certain biomarker profile (13). An ideal biomarker for this design would need a well-established cut-off point and have an assay with a rapid turnaround time (4). A successful enrichment design is very efficient, increases the power of a study as compared to the unselected/all comers design, and may require only a small sample size if the treatment effect is large in the

biomarker positive subgroup, even if the biomarker positivity prevalence is low in the population of interest (14). Conversely, if the assay is imperfect, the treatment may actually have an effect in the negative subgroup or whole population which will remain unknown as only the positive subgroup is recruited (15, 16). Furthermore, this design may require a large population to be screened to identify the biomarker positive subgroup; moreover, it cannot determine whether the biomarker is predictive or not. A slight modification to the enrichment design is the hybrid or mixture design (Figure 2b) allowing the treatment effect of the intervention therapy in the biomarker positive subgroup to be compared with the treatment effect of the control arm in both the biomarker positive and negative population (16); this design would still require a well-established biomarker.

The vast majority of currently conducted trials collecting biological specimens for marker measurements use the Unselected or All Comers design (Figure 2c) as all patients meeting the eligibility criteria are entered into the trial independent of previous testing or the resulting status of the biomarker of interest. Furthermore, one does not need to be certain about the benefit of the marker in either the overall population or the biomarker defined subgroups as it provides the treatment effect in the overall population as a whole (13). Less established biomarkers needing further validation of their performance or having a slower assay turnaround times could be used in this design (4). However, the cost of measuring the biomarker in the whole population will be large if a high proportion of patients are not able to contribute biomarker measurements, hence the prevalence of the biomarker should be high (16).

The Marker-Based Strategy Design recruits eligible subjects regardless of their biomarker status, just like all-comer design and then randomly assigns the patients to either to have therapy determined by their marker status, in the biomarker directed arm, or to receive therapy independent of marker status (14) (Figure 2d). This is a cost-effective design in

comparison to the enrichment or biomarker stratified design in that the biomarker is only assessed in the biomarker directed arm and it is ethical in that there are no issues, including compliance, associated with withholding the biomarker status from the control-arm patients (14). As such this design can be used when multiple treatments are under investigation or when a treatment decision will be based on multiple markers (11); much like the enrichment design, biomarkers should also be well established. A major disadvantage with this design is the loss of power through the overlap in patients receiving the same treatment regimen in the biomarker directed and the control arm, as the biomarker is not assessed in the control arm (14). As randomisation has taken place before biomarker testing, the sub-populations may be imbalanced (11). As the intervention may be better than the control treatment for all patients regardless of biomarker status, a positive trial does not prove the utility of the biomarker (14). A modification to this design allows some clarification in whether any positive results are due to a true effect of marker status or to an improved regimen regardless of marker status (Figure 2e) (14).

The Interaction or Biomarker stratified design (Figure 2f) allows for the prognostic value of marker to be evaluated by comparing the outcomes of patients treated with the same regimen between the two marker groups (14). Stratifying on the biomarker upfront assures that only patients with adequate test results will enter the trial (11). This design is inefficient as the trial needs to be powered to detect either a difference in the effect of the treatment in biomarker-positive and-negative patients through an interaction, which requires a very large number of patients, or the effect in all patients, which is likely to be small due to a potentially small or even negative effect in biomarker-negative patients (14). This design also allows a test of interaction to be performed to determine whether a differential treatment effect exists in the two marker groups, assuming that the sample size is adequate for this test to be appropriately powered (11); it may not provide power for testing the treatment effect

separately in the two marker subgroups based on sample size (11). However, if the primary question requires separate testing in the marker subgroups, the study would need to be powered on this basis. Furthermore, this design cannot be used when multiple biomarkers or treatments are being evaluated due to cost as it is intended to evaluate one treatment or biomarker (11). If the biomarker or treatment is ineffective, this design can further be quite wasteful (11). An ideal biomarker for this design would not need to be as well-established in that equipoise would need to be present to justify randomly assigning these patients to therapy based on their biomarker status (14).

Standard superiority and futility interim monitoring can be used for most marker-based strategy and enrichment designs which simply focus on comparing the overall efficacy between the randomized arms (17); however these may be more difficult to use for biomarker-stratified designs due to the multiple hypotheses under study. To help clarify which design to use for further phase III testing of an intervention, Freidlin et al proposed a randomised phase II trial design whose results could help investigators decide whether or not to stop testing of that particular intervention, or to use a biomarker enrichment, biomarker stratified or standard phase III design (18) for the future phase III trial.

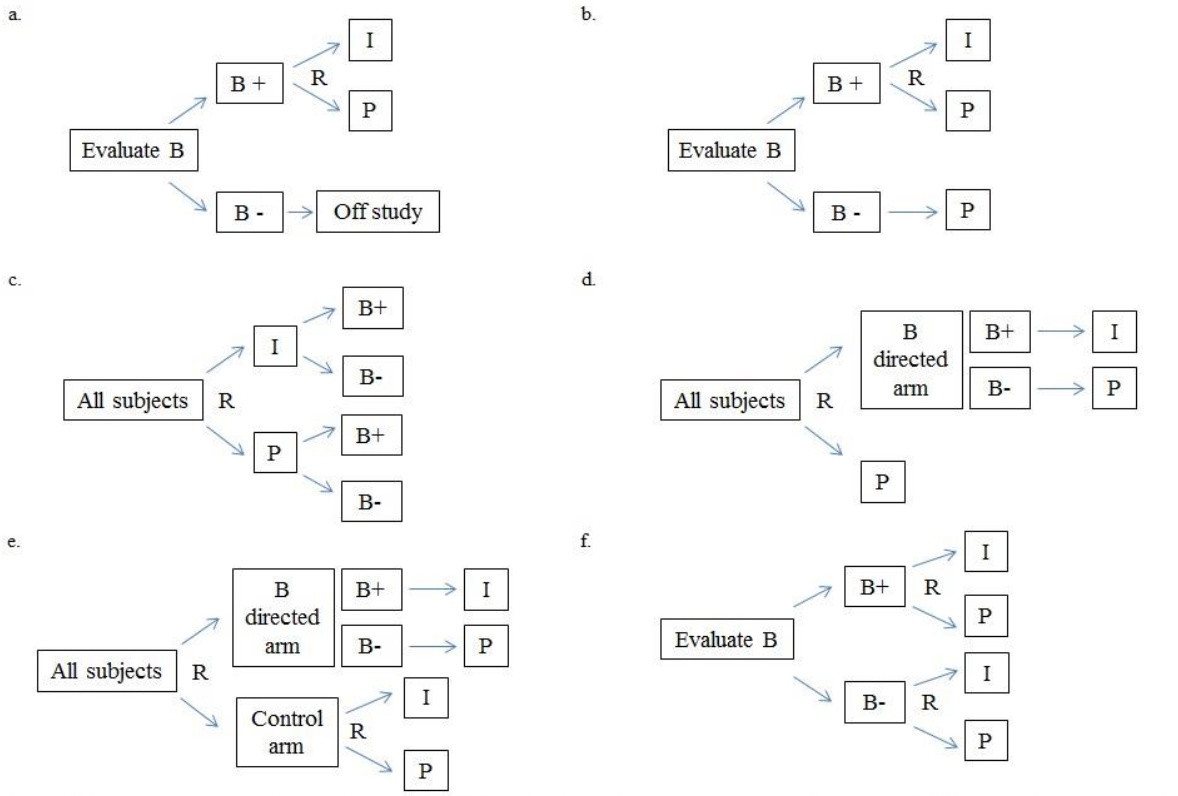

Figure 2 Legend: a. Enrichment design b. Hybrid/Mixture design c. Unselected/All comers design d. Marker based strategy with standard control e. Marker based strategy with randomized control f. Interaction/Biomarker stratified design
Abbreviations: B=Biomarker, R=Randomize, P=Placebo/control/standard of care arm, I=Intervention/experimental arm

Several statistical testing strategies can be used in the aforementioned designs if the required sample size to perform a subgroup-specific approach is not feasible. One such strategy is sequential testing which assumes that the new treatment is unlikely to be effective in the biomarker-negative patients unless it is effective in the biomarker positive patients (16, 19). This may make it difficult to determine whether the treatment is beneficial in the biomarker negative subgroup, as a large treatment effect in the biomarker positive patients could be driving an effect found in the whole population even if there is no treatment effect in the negative subgroup (19). The Marker sequential test design helps avoid the conclusion that the treatment benefits all patients when the overall effect is driven by the biomarker-positive patients and recommends treatment to either the whole population or to no patients regardless of biomarker status depending on testing results (19, 20). The goal of this testing is to stop subgroups for which the hypothesis has been answered, and allow reallocation of resources to

the open subgroups (20, 21). In general, crossing the efficacy boundary in the biomarker-positive subgroup results in stopping that subgroup, and crossing the inefficacy boundary results in the stopping of the entire study; however for the biomarker-negative subgroup, crossing either boundary results in the stopping of that subgroup only (20, 21).

Other marker-based methods are described in Table 1.

4.2. Overview of adaptive methods:

Adaptive designs are increasingly being used to incorporate biomarkers into clinical trials as they allow investigators to analyse the data mid-trial, associate those results with known biomarkers, and then modify the ongoing trial following the results, targeting those people most likely to benefit from their biomarker status (22, 23). Advantages and disadvantages of adaptive designs have been covered extensively in the literature (16, 24, 25). Ideal biomarkers for adaptive designs usually have a well-established cut-off point and have an assay with a rapid turnaround time (4) but there is some uncertainty about the benefit in overall population versus marker defined subgroups.

Bayesian adaptive randomization designs have been used to randomly assign patients to treatments based on the biomarker status. While equal randomisation can improve the efficiency of a trial by maximizing the statistical power, adaptive randomization offers a higher probability of assigning more patients to a more efficacious treatment, especially when the treatment difference is large or the relevant disease is rare (24). Several types of adaptive randomisation techniques have been proposed, including using short-term response information to facilitate adaptive randomization for survival clinical trials (26), covariate-adaptive randomization (24), response-adaptive randomization (24, 27) and outcome-based adaptive randomization (21, 28); however there may be potential bias if there are any time trends in the prognostic mix of the patients accruing to the trial (28). Patient accrual can be

modified using designs such as the adaptive accrual design (16), the biomarker-adaptive parallel Simon two-stage design (29) and the phase III design for the setting of a single binary biomarker stratification design (15) (Table 1).

Adaptive versions of the aforementioned marker-based designs also have been proposed such as the Bayesian adaptive marker-stratified design (27), the adaptive enrichment design (30, 31) and an adaptive version of testing approaches using utility functions (32). Furthermore, a Bayesian prediction model has been proposed to help predict whether a biomarker is truly associated to a clinical outcome using a meta-analytic approach (33). Finally, a Bayesian adaptive design has been suggested for simultaneously testing several predictive biomarkers and new experimental treatments in multi-arm phase II trials (34) (Table 1).

Finding an appropriate cut-off point as well as the process of biomarker validation is very difficult. Two adaptive designs have been proposed to help in this process. The first is the biomarker-adaptive threshold design (35) (Figure 3a) where the optimal cut-off point identifying the subgroup of patients with the greatest treatment effect is determined in phase 2 through a permutation analysis with confidence intervals further derived for the optimal threshold using bootstrap re-sampling (35). The second is the adaptive signature design (36) (Figure 3b) where if the phase 1 analysis is not significant, phase 2 begins using the remaining α ; either half of the study population is used to develop a signature and the other half to validate it through comparison of outcomes for the sensitive patients in the intervention and control arm being compared (36), or practically the entire study population is used in both the signature development and the validation steps in a cross-validation extension of this design (37). This approach was developed as the original method has limited power as only half of the patients are used in each of the development and validation portions of the design (36). In the cross-validation method, the trial population is first split

into a validation and a development subpopulation (37). For each development subpopulation, a predictive signature is developed which is then applied to find a subgroup of patients that are sensitive in the validation subpopulation (37). The process is then repeated for all the validation subpopulations so that each patient in the trial population appears in exactly one of the validation subpopulations and that by the end of the procedure, each patient is classified as being sensitive or not (37). A test statistic is used to assess the presence of a treatment effect in the sensitive patient subgroup with a permutation method to obtain a corresponding P value (37). A permuted data set is constructed and the entire process is repeated for the permuted data set with the corresponding test statistic calculated each time (37).

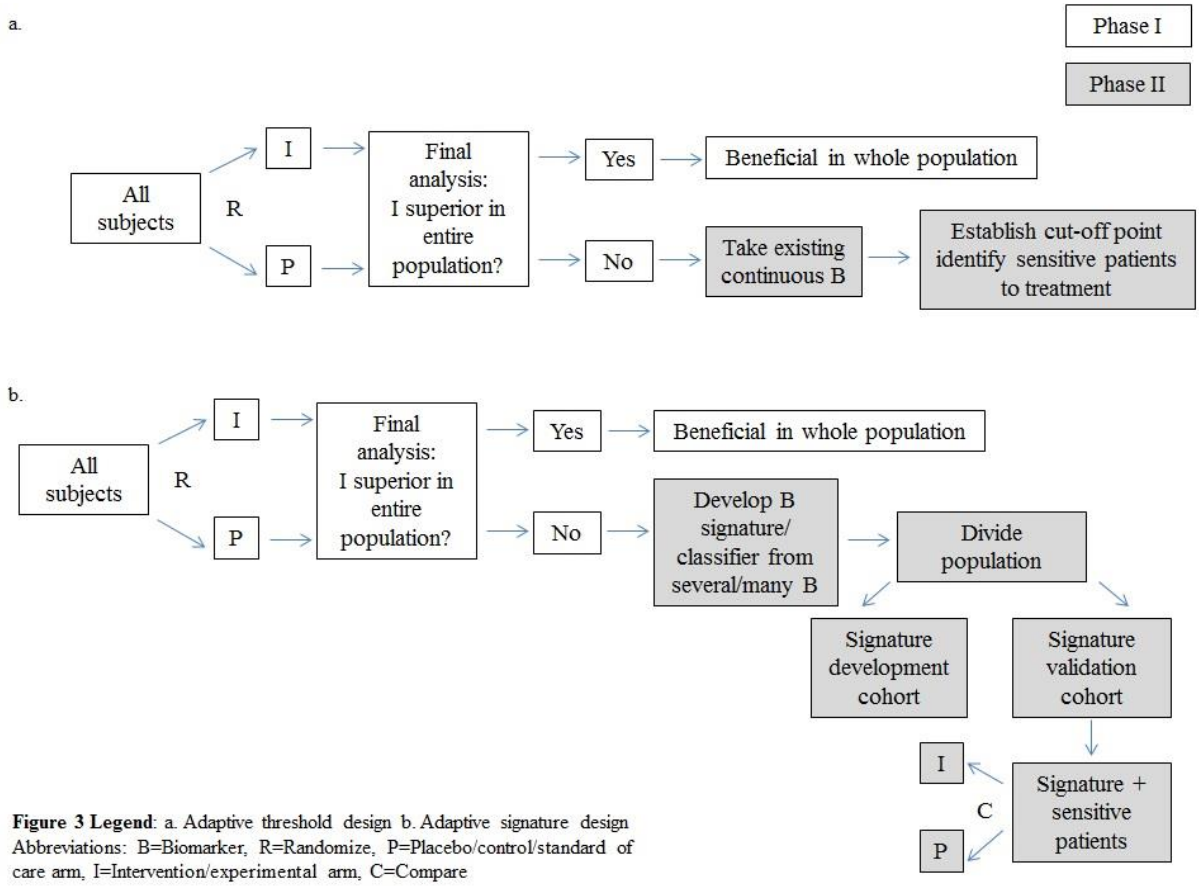

Figure 3 Legend: a. Adaptive threshold design b. Adaptive signature design Abbreviations: B=Biomarker, R=Randomize, P=Placebo/control/standard of care arm, I=Intervention/experimental arm, C=Compare

Table 1:

Name	Description	Key Advantages	Key Disadvantages	Trial example, if available (N/A=not available)	Reference Number
Marker-based design methods					
Enrichment/Targeted	Figure 2a	Efficient, Increases Power, Smaller sample size can be used when large treatment effect exists in biomarker positive subgroup	Unknown effect whole population or marker negative subgroup, large population needs to be screened to identify biomarker positive subgroup, strong link biomarker and treatment needed	CALGB-10603, Lung MATRIX	(13-15, 38)
Hybrid/Mixture	Figure 2b	Specimens and follow up information collected all patients for future testing	Large population needs to be screened, strong link biomarker and treatment needed	TAILORx, MINDACT	(16, 39)
Unselected/All comers	Figure 2c	Recruitment not dependent on previous testing or biomarker status	Cost may be high if high proportion of patients not able to contribute biomarker measurements, dilution treatment effect if only small subgroup benefits from treatment, use of testing methods may be difficult if ability to provide adequate biological specimen is not an eligibility criterion to participate in the trial	EGFR as a Marker for Erlotinib in Lung Cancer	(4, 16)
Marker Strategy Design without randomisation in non-marker directed arm	Figure 2d	Recruitment not dependent on biomarker status, Cost-effective as biomarker is only assessed in the biomarker directed arm, Ethical, Use with multiple markers and multiple treatments	Loss of power due to overlap in patients, dilution between arm difference with overlap in patients	Tumor Chemosensitivity Assay Ovarian Cancer Study	(14, 16)

Modified Marker Strategy Design with randomisation in non-marker directed arm	Figure 2e	Recruitment not dependent on biomarker status, Cost – effective as biomarker is only assessed in the biomarker directed arm, Ethical, Use with multiple markers and multiple treatments	More costly than marker strategy design without randomisation, Potential dilution of the between-arm treatment difference	SHIVA	(4, 14)
Interaction/Biomarker stratified	Figure 2f	Allows prognostic value marker to be evaluated, gold standard for whether treatment is dependent on biomarker status	Costly, Cannot be used for multiple markers and treatments, Can be inefficient as needs to be powered to detect either a difference in effect of treatment in biomarker-positive and-negative patients or effect in all patients, Need large sample size for testing treatment effect separately in two marker subgroups	MARVEL	(14)
Randomised phase II trial design whose outcome could help investigators decide on a future phase III study design	In step 1, the null hypothesis is tested in the biomarker positive subgroup. Based on the results of step 1, step 2 will either test the null hypothesis in all randomly assigned patients or will test it in the biomarker negative subgroup. Based on the results of steps 1) and 2), the investigators will either decide to stop testing a particular intervention, or to use a biomarker enrichment, biomarker stratified or standard phase III design for a future phase III trial	Allows streamlining from phase II to III of drug development and can be incorporated into a phase II/III design strategy	Needs established and validated biomarker	N/A	(18)
Biomarker informed two-stage winner	After interim analysis, less promising arms of several treatment arms dropped, based on biomarker status, with only most promising arm continuing on to end of study	Assigns more patients to most promising treatment	Needs established and validated biomarker	N/A	(40)

Tandem two-step phase II trial design incorporating a prespecified pharmacogenomic predictor of response	All comers stage 1, if number of objective responses not high enough in all patients, pharmacogenomic predictor predicts patients likely to be responders and study continues accrual only in subgroup for stage 2	Recruitment not dependent on biomarker status until second stage	Needs established and validated biomarker of response	N/A	(41)
Joint inference on a subpopulation of "super-responders"	Joint inference on subpopulation of "super-responders", defined by baseline-expressed biomarkers, and full study population using two-way analysis of variance model	Study progresses only in patients more likely to benefit	Needs established and validated biomarker(s)	N/A	(42)
Run in phase III design using an intermediate measurement as a predictive biomarker	All patients tested after short run in period; if not statistically significant only marker positive subset tested. Dependent on results, all patients or only marker positive patients randomised to trial.	Study progresses only in patients more likely to benefit	Need intermediate (post treatment) measurement as predictive biomarker	N/A	(5)
Biomarker informed add-arm design for unimodal response	Interim decisions based on the measurements of a biomarker to identify inferior treatments in a multi arm study	Inferior study treatments are identified in study population	Requires strong surrogate biomarker for toxicity primary endpoint	N/A	(40)
Bridging Continual Reassessment Method (CRM)	Patients are divided into several subgroups with different maximum tolerated doses (MTD) to drug based on certain biomarkers through a mixture estimator to estimate dose-toxicity curve	Allows more appropriate MTDs to be employed in different subgroups of patients	Needs established and validated biomarker	N/A	(43)
Marker-based testing methods					

Sequential testing	Study hypothesis first tested in whole population using reduced α and if significant treatment considered effective in all; if not, treatment effect evaluated in biomarker defined subset, using remaining α. If evidence treatment works best in biomarker defined subgroup, testing first done in subgroup followed by whole population.	Sequentially controls and preserves study-wise type I errors	Needs established and validated biomarker, difficult determine whether treatment beneficial in biomarker negative subgroup	N/A	(16, 19)
Marker sequential test	First tests biomarker-positive patients at reduced α, followed by biomarker-negative patients with the remaining α, if test significant. If test not significant, then overall population is tested with remaining α.	Allows overall type-I error level to be controlled	Needs established and validated biomarker	N/A	(19, 20)
Parallel testing strategy 1	Separate testing done simultaneously in biomarker positive and negative subgroups	Help determine effect in both positive and negative subgroups	Requires high sample size as necessitates allocating overall α between two subgroup tests making significance hard to achieve	N/A	(19)
Parallel testing strategy 2	Tests both overall and biomarker-positive populations simultaneously with strength of the predictive value of biomarker determining whether required sample size driven by biomarker-positive or whole population hypotheses	Helps determine a more appropriate sample size	Needs predictive biomarker	N/A	(19)
Adaptive design methods					
Covariate-adaptive randomization	Uses overall covariate distribution among treatment groups to determine treatment allocation for next enrolling patient	Allows prognostic factors to be balanced among treatment arms	Needs very accurate covariate information	N/A	(24)

Response-adaptive randomization	Uses previous patient's response to treatment in interim data to determine next enrolling patient's allocation, so that if successful, next patient will be assigned to same treatment; otherwise to the alternative treatment, potentially incorporating marker status	Higher probability of assigning more patients to more efficacious treatment	Necessitates response to be assessed in relatively short time period	I-SPY2	(24, 27, 44)
Outcome-based adaptive randomization	Uses outcome data accumulated in trial to randomly assign patients to treatments based on biomarker status	Higher probability of assigning more patients to more efficacious treatment	Requires short-term reliable outcome and may result in bias	I-SPY2, BATTLE	(21, 44, 45)
Adaptive accrual	Accrual ensues in both marker-defined subgroups until interim analysis where if treatment effect in one group does not reach futility boundary, accrual stopped to subgroup with only other subgroup continuing accrual until total planned sample size reached	Allows more appropriate patients to be targeted for treatment	Need to ensure subgroups are well defined by good validated biomarker	N/A	(16)
Biomarker-adaptive parallel Simon two-stage	Conducts two parallel studies in biomarker negative and positive subgroups, design continues enrolling unselected patients in stage II if number of responses to drug in biomarker-negative group in stage I meets/exceeds a cut-off. Otherwise only biomarker positive patients enrolled	Allows more appropriate patients to be targeted for treatment	Need prospectively defined cut-off	N/A	(29)
Phase III design for setting of a single binary biomarker stratification	Futility monitoring performed in biomarker negative patients at interim analysis based on joint prior distribution for treatment effects in both positive and negative subgroups, accrual can then be continued or halted in subgroup	Allows more appropriate patients to be targeted for treatment	Needs established and validated biomarker	N/A	(15)

Bayesian adaptive marker-stratified	Both response-adaptive randomization according to the patients' biomarker profiles as well as an interim analysis with early stopping rules are used	Higher probability of assigning more patients to more efficacious treatment	Needs established and validated biomarker	N/A	(27)
Adaptive enrichment	All eligible subjects recruited in first stage, followed by interim analysis where study design may be switched to all-comer design or allows termination of the biomarker negative cohort depending on the interim analysis results, in Stage 2	Allows sample size, end points, randomization ratio and eligibility criteria to be adjusted	Study population drift, loss of study power, loss of integrity of original trial, Needs established and validated biomarker	N/A	(30, 31)
Adaptive testing methods					
Adaptive version of the testing approaches	Treatment effects of potential marker-based subpopulations evaluated through utility functions at interim analysis for stage two trial testing, using a Bayesian approach assuming prior distribution on efficacy parameters. Patients first recruited from whole population and at interim analysis, trial adapted to continue only in subpopulation.	Allows more appropriate patients to be targeted for treatment	Prevalence in population should be well known, needs established and validated biomarker	N/A	(32)
Bayesian prediction model to help predict whether biomarker truly associated to clinical outcome	Bayesian meta-analytic method for building prediction model between biomarker and the clinical endpoint. Used to predict rate ratio of clinical endpoint from an early biomarker. Proposed prediction model evaluated using extensive simulations	Requires only previous trial-level summary data in model building, no patient-level data necessary	Sample size estimation involving biomarker may pose difficulties, go/no go decision rules based on biomarker need to be determined	N/A	(33)
Bayesian adaptive design for simultaneously testing several predictive biomarkers and new experimental treatments in multi-arm phase II trials	Uses Bayesian adaptive randomisation procedure where patients recruited before first interim analysis are initially randomised equally between control and all experimental treatments linked to their tumour biomarkers. At each interim analysis,	Time and cost efficient way of matching and testing novel predictive biomarkers and new interventions	Managing the randomisation and changing the allocation ratios after an interim analysis can be challenging, needs established and validated biomarkers	N/A	(34)

	Bayesian logistic regression model fitted to model the probability of treatment success. Posterior probabilities from the model that each experimental treatment is superior to the control for each biomarker profile can be used to update the allocation probabilities to each arm				
Biomarker-adaptive threshold	Figure 3a	Allows for parallel evaluation of new intervention in all patients at prespecified α, as well as establishment/validation of biomarker cut-off point, if the phase 1 analysis is not significant, using the remaining α	Need pre-existing biomarker	N/A	(35)
Adaptive signature	Figure 3b	Allows development and testing of a biomarker signature based on high-dimensional data.	Limited power unless K-fold cross-validation procedure used	N/A	(36, 37)

4.3. Other emerging design methods:

Some methods of incorporating biomarkers into clinical trials do not easily fit into the marker-based/adaptive methods classification. One of these is the Longitudinal Cohort Study With or Without Nested Clinical Trials (46), where tumour profiling of patients in an accredited diagnostic laboratory allows an individual centre to participate and enrol patients in multiple nested clinical trials. The Histology-Based Clinical Trial Design, also called an umbrella or platform trial (46), allows a number of agents to be matched to specific molecular characteristics in one tumour type using a prespecified set of rules in a standing trial structure (15). The effectiveness of treatment assignments based on molecular profiling results are then compared to a control arm (2). A key advantage is that the design, conduct, and analysis of each sub-study are independent of the other sub-studies; however this model requires a large sample size and it is resource intensive (2). This method is used by the LUNG-MAP (47), BATTLE (45), Lung MATRIX (38) and FOCUS4 (48) studies, among others. An alternative to the umbrella trial is the Histology-Independent, Aberration-Specific Clinical Trial Design, otherwise known as a bucket or basket trial (46) which is designed to discover the effects of a targeted agent against a specific molecular characteristic across different tumour types. The key disadvantages of this model are the potential of false negative conclusions if the trial has insufficient representation of patients with tumour types having the molecular characteristic of interest, and that it evaluates only one drug-molecular characteristic pair at a time (2). The NCI MATCH trial uses this method (49).

5.0 DISCUSSION:

Incorporation of a biomarker into the design of clinical trial poses numerous challenges including that it must statistically be based on the prevalence and distribution of the marker in

the patient population and the chosen clinical endpoint so that the required sample size to test the hypothesis can be estimated (50); as well as estimating the unavailability rate of biomarker measurements within those calculations (14). Furthermore, the optimal patient population for the study needs to be considered, as well as whether the strength of the marker effect is sufficient to separate patients into meaningful outcome groups without overlap (50). Regulatory agencies now require that biomarker cut-off points splitting patients into high and low risk groups be defined and validated for use in patient populations (51). Splitting a continuous measure into a dichotomous group reduces the power to detect a real association with outcome, but if optimized cut-off values are used, then they should be determined using a training data set with an independent testing data set to validate the cut-off point (51). Using a cut-off point as reported from another study or defined based on the distribution of marker level among patients without use of clinical outcome data is also unbiased (51). Clinically the biomarker needs to be considered in terms of the toxicity of the proposed therapy and should be able to be assimilated into routine clinical practice in a cost effective way, must possess a highly significant predictive value, and be independent of the known clinicopathologic predictors of prognosis (50). Additionally, it needs to have therapeutic implications readily interpretable by a clinician and have been validated in independently confirmed phase III studies (51). Biologically, this requires that the marker can be assessed reproducibly in numerous clinical samples by a well-characterized, standardised, accurate and quality controlled assay system (50). For such an assay to be developed, this necessitates that the biology of the marker be well understood and that the marker assay results be statistically associated to high-quality patient data (51). To ensure reproducibility, preclinical work should include a statement about the test's quality controls, what the assay is designed to measure, optimal assay conditions, the specific kit or critical reagents, details of the scoring system, selection of a uniform threshold for binary interpretation of results, a statement

regarding the reproducibility or precision, sensitivity, specificity, and a reference to the clinical validation of the assay, such as comparing results using the same samples in different laboratories (50). As assays can change due to new platform availability and as new promising biomarkers can be discovered during the course of an ongoing trial, a flexible protocol should be adopted to incorporate emerging changes along with good specimen storage so that biopsies can be re-tested with the new assay and results compared to the old assay (52). Also as bioinformatics software and assay platform regularly change, it is important to specify the version used during the course of a clinical trial.

As tissue biopsies are invasive and tumours have heterogeneity, there is a potentially small amount of available tumour specimen and given that not all tumours can be biopsied due to poor accessibility, tissue type is another challenge. The distribution of the marker in normal and abnormal tissues (50) needs to be determined in a sufficient numbers of samples. Serial biopsies of a patient's tumour provide a dynamic view of the individual patient's disease course and response to treatment (53); however these are difficult to obtain. Size of the tumour is a potential issue when tissue has been collected retrospectively as larger tumours are more likely to end up in frozen tumour banks; allowing generalizability to only those. The limited and differential availability of tumour tissue itself may be a function of the disease, hospital diagnostic practices, pathology laboratory preservation, storage protocols, study population, and has implications for study generalizability and power (54). Appropriate, non-invasive and readily accessible tissues, such as liquid biopsy samples including circulating blood cells, circulating tumour cells (CTC) or buccal swabs, could serve as potential surrogates for tumour tissue (55). These need to be inexpensive to measure, making CTCs problematic due to the small numbers that can be detected with existing methods (53). However CTCs provide an integrated picture of all clones present in the patient (53). Imaging is another attractive alternative as it is noninvasive and therefore better

suited to serial measurements (6). However imaging can be quite costly as it needs to be performed on one subject at a time and requires sophisticated equipment and imaging probes (6). Conversely, imaging offers the ability to assess target expression across all sites of disease, to assess sites challenging to biopsy and assay, such as bone, and can act as an early response marker as biochemical and molecular changes will likely precede changes in tumour size, potentially reducing trial costs by reducing trial duration (6). Adding biomarker to trials is very costly in general especially in the case of a low prevalence marker in the population of interest (50).

Ethical concerns are another challenge as knowledge of the biomarker status may affect compliance to the randomized treatment and the biomarker status may still correlate with clinical features of the patient that might influence treatment preference, even if the biomarker status is withheld (14). Reliable conclusions about a biomarker are more likely to arise from large, collaborative (including both research organisations and pharmaceutical companies), phase III studies as opposed to many undersized studies using a variety of statistical methodology and clinical inclusion criteria (51). The European Organisation for Research and Treatment of Cancer and the National Clinical Trials Network are paving the way through the creation collaborative groups to conduct large-scale studies (46, 49).

Only a few biomarkers have been validated sufficiently to be in clinical use including KRAS (56), estrogen and progesterone receptors, c-erbB-2/HER-2/NEU (57), CA-125, prostate specific antigen, and human chorionic gonadotropin (55). Robust preclinical studies should first be performed focusing on reliably identifying the drug target as well as developing a validated assay for the biomarker (1). The biomarker should then be added to phase I trials to better characterise both its assay and its performance, followed by its incorporation in phase II trials for hypothesis testing and finally, its inclusion in a phase III trial for proper clinical validation (1). It is critical that proper preclinical work and stringent

validation be done before taking forward only the most promising biomarkers into confirmatory studies. Limitations of this review include the exclusion of Non-English papers and of those papers with no abstract available.

6.0 CONCLUSIONS:

Although the incorporation of biomarkers into clinical trials poses numerous challenges, lifesaving treatments can be better targeted to appropriate patients so their inclusion is of paramount importance. The future of biomarkers in clinical trials is bright as novel designs can help greatly to simplify their incorporation. The marker-based designs described above are already in common usage; however designs such as the biomarker-adaptive threshold design and the adaptive signature design can help to greatly mitigate many of the aforementioned challenges and should be in greater use. Further approaches are needed to answer methodological issues that have not been addressed within the methods presented here.

7.0 REFERENCES:

1. McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15:1898-905.
2. Kummar S, Williams PM, Lih CJ, Polley EC, Chen AP, Rubinstein LV, et al. Application of molecular profiling in clinical trials for advanced metastatic cancers. Journal of the National Cancer Institute. 2015;107.
3. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27:4027-34.
4. Mandrekar SJ, An MW, Sargent DJ. A review of phase II trial designs for initial marker validation. Contemporary clinical trials. 2013;36:597-604.
5. Hong F, Simon R. Run-in phase III trial design with pharmacodynamics predictive biomarkers. Journal of the National Cancer Institute. 2013;105:1628-33.
6. Mankoff DA, Farwell MD, Clark AS, Pryma DA. How Imaging Can Impact Clinical Trial Design: Molecular Imaging as a Biomarker for Targeted Cancer Therapy. Cancer journal (Sudbury, Mass). 2015;21:218-24.

7. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumor MARKer prognostic studies (REMARK). Nature clinical practice Urology. 2005;2:416-22.
8. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. Journal of the National Cancer Institute. 1991;83:154-5.
9. Khambata-Ford S, Garrett CR, Meropol NJ, Basik M, Harbison CT, Wu S, et al. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25:3230-7.
10. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. Breast cancer research and treatment. 1998;52:289-303.
11. Polley MY, Freidlin B, Korn EL, Conley BA, Abrams JS, McShane LM. Statistical and practical considerations for clinical evaluation of predictive biomarkers. Journal of the National Cancer Institute. 2013;105:1677-83.
12. Hayes DF, Bast RC, Desch CE, Fritsche H, Jr., Kemeny NE, Jessup JM, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. Journal of the National Cancer Institute. 1996;88:1456-66.
13. Rothmann MD, Zhang JJ, Lu L, Fleming TR. Testing in a Prespecified Subgroup and the Intent-to-Treat Population. Drug information journal. 2012;46:175-9.
14. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. Journal of the National Cancer Institute. 2010;102:152-60.
15. Simon R. Biomarker based clinical trial design. Chinese clinical oncology. 2014;3:39.
16. Lin JA, He P. Reinventing clinical trials: a review of innovative biomarker trial designs in cancer therapies. British medical bulletin. 2015;114:17-27.
17. Liu A, Liu C, Li Q, Yu KF, Yuan VW. A threshold sample-enrichment approach in a clinical trial with heterogeneous subpopulations. Clinical trials (London, England). 2010;7:537-45.
18. Freidlin B, McShane LM, Polley MY, Korn EL. Randomized phase II trial designs with biomarkers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30:3304-9.
19. Freidlin B, Sun Z, Gray R, Korn EL. Phase III clinical trials that integrate treatment and biomarker evaluation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31:3158-61.
20. Freidlin B, Korn EL, Gray R. Marker Sequential Test (MaST) design. Clinical trials (London, England). 2014;11:19-27.
21. Freidlin B, Korn EL. Biomarker-adaptive clinical trial designs. Pharmacogenomics. 2010;11:1679-82.
22. Time to adapt. Nature. 2010;464:1245-6.
23. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, et al. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. Statistics in medicine. 2009;28:1445-63.
24. Zang Y, Lee JJ. Adaptive clinical trial designs in oncology. Chinese clinical oncology. 2014;3:49.
25. Antoniou M, Jorgensen AL, Kolamunnage-Dona R. Biomarker-Guided Adaptive Trial Designs in Phase II and Phase III: A Methodological Review. PloS one. 2016;11:e0149803.
26. Huang X, Ning J, Li Y, Estey E, Issa JP, Berry DA. Using short-term response information to facilitate adaptive randomization for survival clinical trials. Statistics in medicine. 2009;28:1680-9.
27. Lee JJ, Xuemin G, Suyu L. Bayesian adaptive randomization designs for targeted agent development. Clinical trials (London, England). 2010;7:584-96.
28. Korn EL, Freidlin B. Outcome--adaptive randomization: is it useful? Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29:771-6.

29. Jones CL, Holmgren E. An adaptive Simon Two-Stage Design for Phase 2 studies of targeted therapies. Contemporary clinical trials. 2007;28:654-61.
30. Simon N, Simon R. Adaptive enrichment designs for clinical trials. Biostatistics (Oxford, England). 2013;14:613-25.
31. Wang SJ, O'Neill RT, Hung HM. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical statistics. 2007;6:227-44.
32. Graf AC, Posch M, Koenig F. Adaptive designs for subpopulation analysis optimizing utility functions. Biometrical journal Biometrische Zeitschrift. 2015;57:76-89.
33. Jiang Z, Song Y, Shou Q, Xia J, Wang W. A Bayesian prediction model between a biomarker and the clinical endpoint for dichotomous variables. Trials. 2014;15:500.
34. Wason JM, Abraham JE, Baird RD, Gournaris I, Vallier AL, Brenton JD, et al. A Bayesian adaptive design for biomarker trials with linked treatments. British journal of cancer. 2015;113:699-705.
35. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. Journal of the National Cancer Institute. 2007;99:1036-43.
36. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. Clinical cancer research : an official journal of the American Association for Cancer Research. 2005;11:7872-8.
37. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16:691-8.
38. Middleton G, Crack LR, Popat S, Swanton C, Hollingsworth SJ, Buller R, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2015;26:2464-9.
39. Viale G, Slaets L, Bogaerts J, Rutgers E, van't Veer L, Piccart-Gebhart MJ, et al. High concordance of protein (by IHC), gene (by FISH; HER2 only), and microarray readout (by TargetPrint) of ER, PgR, and HER2: results from the EORTC 10041/BIG 03-04 MINDACT trial. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2014;25:816-23.
40. Wang J, Chang M, Menon S. Biomarker Informed Add-arm Design for Unimodal Response. Journal of biopharmaceutical statistics. 2015.
41. Pusztai L, Anderson K, Hess KR. Pharmacogenomic predictor discovery in phase II clinical trials for breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13:6080-6.
42. Glimm E, Di Scala L. An approach to confirmatory testing of subpopulations in clinical trials. Biometrical journal Biometrische Zeitschrift. 2015;57:897-913.
43. Liu S, Pan H, Xia J, Huang Q, Yuan Y. Bridging continual reassessment method for phase I clinical trials in different ethnic populations. Statistics in medicine. 2015;34:1681-94.
44. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clinical pharmacology and therapeutics. 2009;86:97-100.
45. Kim ES, Herbst RS, Wistuba, II, Lee JJ, Blumenschein GR, Jr., Tsao A, et al. The BATTLE trial: personalizing therapy for lung cancer. Cancer discovery. 2011;1:44-53.
46. Lacombe D, Tejpar S, Salgado R, Cardoso F, Golfinopoulos V, Aust D, et al. European perspective for effective cancer drug development. Nature reviews Clinical oncology. 2014;11:492-8.
47. Steuer CE, Papadimitrakopoulou V, Herbst RS, Redman MW, Hirsch FR, Mack PC, et al. Innovative Clinical Trials: The LUNG-MAP Study. Clinical pharmacology and therapeutics. 2015;97:488-91.
48. Kaplan R, Maughan T, Crook A, Fisher D, Wilson R, Brown L, et al. Evaluating many treatments and biomarkers in oncology: a new design. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31:4562-8.

49. Abrams J, Conley B, Mooney M, Zwiebel J, Chen A, Welch JJ, et al. National Cancer Institute's Precision Medicine Initiatives for the new National Clinical Trials Network. American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting. 2014:71-6.
50. Hammond ME, Taube SE. Issues and barriers to development of clinically useful tumor markers: a development pathway proposal. Seminars in oncology. 2002;29:213-21.
51. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. British journal of cancer. 1994;69:979-85.
52. Biankin AV, Piantadosi S, Hollingsworth SJ. Patient-centric trials for therapeutic development in precision oncology. Nature. 2015;526:361-70.
53. D'Arcangelo M, Margetts J, Greystoke A. The use of circulating biomarkers in early clinical trials in patients with cancer. Biomarkers in medicine. 2015;9:1011-23.
54. Hoppin JA, Tolbert PE, Taylor JA, Schroeder JC, Holly EA. Potential for selection bias with tumor tissue retrieval in molecular epidemiology studies. Annals of epidemiology. 2002;12:1-6.
55. Schilsky RL, Taube SE. Tumor markers as clinical cancer tests--are we there yet? Seminars in oncology. 2002;29:211-2.
56. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26:1626-34.
57. Bast RC, Jr., Ravdin P, Hayes DF, Bates S, Fritsche H, Jr., Jessup JM, et al. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2001;19:1865-78.

8.0 FIGURE LEGENDS:

Figure 1. Diagram

Figure 2. Common biomarker-based designs: a. Enrichment design b. Hybrid/Mixture design c. Unselected/All comers design d. Marker-based strategy with standard control e. Marker-based strategy with randomized control f. Interaction/Biomarker stratified design. Abbreviations: B=Biomarker, R=Randomize, P=Placebo/control/standard of care arm, I=Intervention/experimental arm.

Figure 3. Common Adaptive designs: a. Adaptive threshold design b. Adaptive signature design. Abbreviations: B=Biomarker, R=Randomize, P=Placebo/control/standard of care arm, I=Intervention/experimental arm, C=Compare.