CADTH

CADTH METHODS AND GUIDELINES

# Adaptive and Novel Trial Designs: An Overview of Key Methodologies and Issues in Critical Appraisal


| | |
|---|---|
| Service Line: | CADTH Methods and Guidelines |
| Version: | 1.0 |
| Publication Date: | January 2018 |
| Report Length: | 34 Pages |

CADTH

**Cite As:** Adaptive and novel trial designs: an overview of key methodologies and issues in critical appraisal. Ottawa: CADTH; 2018 Jan.



**About CADTH:** CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

**Funding:** CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

CADTH

## Table of Contents



### Tables

## Figure

## Abbreviations

| | |
|---|---|
| **ATD** | adaptive trial design |
| **PhRMA** | Pharmaceutical Research and Manufacturers of America |

# Introduction

The CADTH *Guidelines for the Economic Evaluation of Health Technologies: Canada* provides guidance on conducting economic evaluations in Canada.[1] In 2014, CADTH began updating this document and identified several technical areas in which additional information would assist in the update. CADTH commissioned work from experts in the field to produce a series of technical reports, one of which focused on adaptive trial designs (ATDs). This report is based on some of the information provided in a draft technical report outlining some common forms of adaptive design.[2] This report is not intended to be used as a guidance document on how adaptive or novel clinical trials should be designed or conducted.

In recent years, there has been increasing interest in incorporating novel methodological features into confirmatory clinical trials, including methods that fall under the category of ATDs. An ATD is broadly defined as a trial in which an alteration in the clinical trial's procedures or statistical analysis can be implemented, on the basis of interim analysis, once the trial has been initiated.[3] Adaptations can be incorporated into different types of clinical trial designs and at any stage, from the early learning or exploratory phase to late confirmatory phases. Adaptations frequently involve alterations to the sample size, the randomization procedures, or the inclusion criteria.

Adaptive features are appealing for several reasons. Depending on the specific adaptation, the efficiency of the trial may be improved by reducing the sample size requirement or by shortening the trial duration.[4] Adaptive features also have the potential to make trials more informative. For example, adaptive features can provide more detailed information on the relationship between dose and response, or improve the ability to demonstrate a treatment effect in a clinical trial if such an effect exists.[4] However, there are concerns about the use of adaptive features in clinical trials related to the potential for adaptations to bias estimates of treatment effect and increase the risk of spurious findings if the statistical analysis does not appropriately consider the impact of the adaptation on the type I error rate and adjust it accordingly.

Other nonadaptive novel features of trial designs can also be employed for various reasons. Enrichment strategies, for example, can be used to select a subset of the population in which a treatment effect may be more readily demonstrated.[5] Although this approach has some advantages, limiting the inclusion to a specific subgroup can potentially affect the generalizability of the estimated treatment effects to a more broadly defined population.

This report, *Adaptive and Novel Trial Designs: An Overview of Key Methodologies and Issues in Critical Appraisal*, describes the major types of adaptive trials, highlights key novel trial designs, and discusses their potential limitations from a clinical methodological perspective. Where available, guidance from regulatory bodies (such as FDA) and the viewpoint of the Pharmaceutical Research and Manufacturers of America (PhRMA) with respect to adaptive and novel clinical trial designs will also be presented to supplement the discussion of limitations and critical appraisal points. Implications that the features of ATDs may have when applying the trial results to an economic model or evaluation are then considered, as this report is intended to help inform the update to the CADTH *Guidelines for the Economic Evaluation of Health Technologies*. As such, the focus of this report will be on the adaptive designs used in confirmatory phase studies, including adaptive randomization, adaptive enrichment, group sequential analysis, and sample size re-estimation. The combined phase II and phase III seamless design will also be included, as the confirmatory phase III portion is relevant as well.

## Definition of an Adaptive Trial Design

Currently, there is no clear consensus on the definition of an ATD. The PhRMA working group proposed the following definition: "A clinical trial design that uses accumulating data to decide how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial."[6] The working group also specified that "changes are made *by design*, and not on an *ad hoc* basis; therefore, adaptation is a design feature aimed to enhance the trial, not a remedy for inadequate planning."[6]

European Medicines Agency (EMA) defines a study as 'adaptive' if the "statistical methodology allows modification of a design element (for example, sample size, randomization ratio, number of treatment arms) and an interim analysis with full control of the type I error."[7] The EMA further suggests that "adaptive designs should not be seen as means to alleviate the burden of rigorous planning of a clinical trial. Instead, adaptive designs would be best utilized as a tool for planning clinical trials in areas where it is necessary to cope with difficult experimental situations.[7] The EMA highlights two key features of adaptive designs, one being the need to control the false-positive error rate, and the other being the need to change a clinical trial to accommodate challenging experimental issues. The definition emphasizes that such changes to a trial must be accounted for in the planning stages (i.e., specified a priori) and must include statistical considerations to control the false-positive (type I) error rate.

## Potential Benefits and Risks of Adaptive Trial Designs

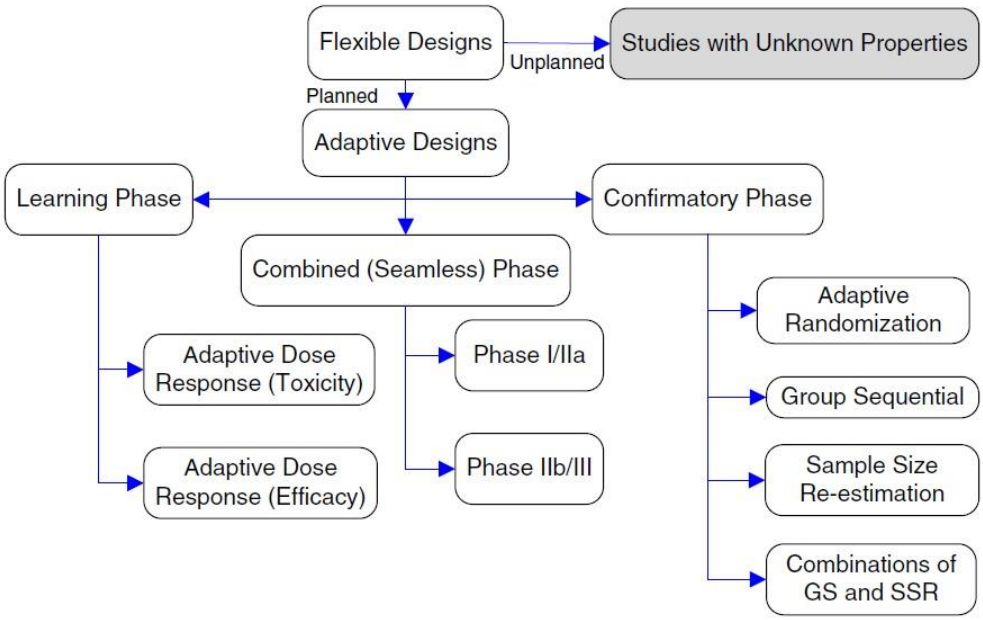
It has been suggested that the flexibility of ATDs has the potential to improve trial efficiency, providing clinical evidence in a shorter time or using fewer study resources (e.g., fewer participants).[4] Alternately, an early assessment of accumulating trial data may indicate that the study is potentially underpowered to demonstrate the true differences between the groups.[4] An adaptation may allow for the study to be extended by permitting the enrolment of additional participants, thereby increasing the sample size and statistical power.[4]

While there are some potential benefits of ATDs, there are methodological considerations that should be noted. Adaptations of a trial have the potential to compromise the trial's methodological validity and integrity and increase the risk of type I error[4] (leading to the conclusion that a difference exists, when it actually does not). Some ATDs also carry the risk that the modifications will sufficiently change the trial so that it can no longer answer the question it was originally intended to answer.[3,8] Major adaptations can create variations in data collection, shift the target population, and create inconsistencies in the statistical analysis plan.[9]

## Types of Adaptive Trial Designs

Kairalla et al. (2012) categorized ATDs on the basis of study phase (learning, confirmatory, or combined) as the adaptations used may be specific to the stage of research and development (Figure 1).[10]

## Figure 1: Categorization of Adaptive Trial Designs




GS = group sequential; SSR = sample size re-estimation.
Source: Reproduced from Kairalla JA, Coffey CS, Thomann MA, Muller KE. Adaptive trial designs: a review of barriers and opportunities. Trials. 2012;13:145.

Generally, ATDs are more common and accepted in the early learning phases of drug development, as they assist with identifying promising therapies and appropriate dosage ranges.[4,10] Adaptive dose-response studies, examining either efficacy or toxicity, fall into the learning phase.[10] Since phase I and II clinical studies were historically not typically considered pivotal trials for regulatory approval, FDA has stated that exploratory studies may be a more suitable setting for gaining experience with the less well-understood adaptive designs.[4] However, FDA also emphasizes that it is important to be aware that the inflation of the type I error rate or biased estimates may occur with some adaptive designs.[4] As such, it is important to follow good principles of study design in exploratory trials given the risk of adversely affecting the drug development program if decisions are made on the basis of poorly designed studies.[4] Unrecognized biased estimates or inflated type I error rates in early phase trials can lead to poor decision-making in the design of later stage trials.[4]

A combined or seamless phase is a study design in which phase I and II or phase II and III trials are conducted sequentially under a single protocol using adaptive techniques.[10] These combined studies can potentially save time and reduce the number of participants being exposed to therapies of unknown benefit and toxicity.[10] In the case of a phase II/III combined trial, both the learning and confirmatory data are collected and used in the regulatory approval, and appropriate statistical techniques ensure that the type I error is not compromised or the estimated treatment effect biased.[10]

Adaptations in phase III studies, or confirmatory trials, are less commonly encountered than in phase I or early phase II studies but have greater relevance to economic evaluation. As such, phase III studies are the focus of this report. As phase III trials are used in regulatory decisions, it is critical that phase III trials appropriately reflect the efficacy and safety of a

therapy without increasing the risk of type I error. The types of adaptations that are used in confirmatory trials include adaptive randomization, adaptive enrichment, group sequential analysis, sample size re-estimation, and combinations of group sequential analysis and sample size re-estimation.[10]

Historically, regulatory authorities have been reluctant to incorporate trials conducted using adaptive methods into their decision frameworks.[6] However, over the past decade there has been progress made toward incorporating adaptive trials into the regulatory process, and so specific guidelines on appropriate procedures for conducting ATDs have been developed both within the pharmaceutical industry and within the regulatory bodies.[4-7] Some types of adaptations in clinical trials are considered to be well understood and accepted by regulatory agencies, such as FDA; others are considered to be "less well understood" (typically designs for which there is relatively little regulatory experience and for which the magnitude of the risk of bias, the size of the potential bias, and the means of eliminating bias are not yet well established).[4] When designed in accordance with FDA's draft guidance, the trial designs of enrichment, classic group sequential analysis, and sample size re-estimation can fall into FDA's "well understood" category; adaptive randomization would be considered "less well understood," which creates difficulty in interpreting results when a treatment effect is demonstrated.[4] The combined seamless phase II/III trial is also considered to be a "less well understood" design.[4]

## Combined Phase II/III Trials

Combined or seamless adaptive trials incorporate study objectives into a single study that would otherwise typically be addressed in separate clinical trials.[9] An adaptive phase II/III trial consists of an exploratory phase or learning phase (which is the phase II component) followed by a confirmatory phase of the trial (which is the phase III component). In the final analysis, data from both phases may be combined. This process is referred to as a "seamless design."[9] Investigators may choose to analyze the data from each phase separately; in that case, the design is referred to as "operationally seamless" and the advantage gained is simply one of time savings. "Inferentially seamless" designs use data from participants enrolled before and participants enrolled after the adaptation.[11] In its most common form, the combined phase II/III trial randomizes participants to a control group and several dosage groups in the exploratory phase II stage, one or more of which (along with the control group) would continue into the confirmatory phase III stage of the trial (the unsuccessful dosage groups are dropped).[6] Data are combined from the exploratory phase II stage and the confirmatory phase III stage using statistical methods that control the type I error rate, adjust confidence intervals, and provide reasonable point estimates.[11,12] These adjustments are necessary given that multiple statistical testing occurs for the primary end point: first with the phase II data alone and then with the combined phase II and phase III data. Further, early stopping of some groups (the unsuccessful ones) occurs during the first stage, which has the potential to create a bias (i.e., only those treatment groups that were successful are analyzed in the next phase).[12] The seamless design may improve trial efficiency in that the same participants are included in both phases of the trial.[13] Gains in trial efficiency can potentially lead to cost savings and may have the advantage of providing longer-term safety data as some participants are followed from the phase II stage.[13] The ability to fully utilize data from both stages is purported to be an advantage of the combined phase II/III approach in terms of duration of follow-up and number of participants.[14] As well, the ability to incorporate additional adaptations such as adaptive randomization or "drop the loser" at the end of the first phase after interim analysis are also seen as potential advantages to this approach.[14]

## Limitations

The combined phase II/III design is considered by FDA to be "less well understood" and has a risk of inflated type I error rate unless statistical methods (e.g., *P* value combination tests, Pocock and O'Brien-Fleming type boundaries, and step down–like procedures) are used to make adjustments to control the type I error rate.[4,11] The statistical analysis of data from inferentially seamless trials must be adjusted to take into consideration aspects of the design (for example, related to treatment selection in approaches such as "drop the loser" where unsuccessful treatment arms are dropped) that can lead to statistical bias (e.g., inaccurate coverage for the associated confidence intervals) if unaccounted for and adjustments are not made.[11]

For the combined phase II/III design, given that the review of the exploratory phase data dictates the design of the remainder of the trial, the results of the interim analysis are generally available, as it is known which groups continue on. This process has the potential to create operational biases (subjective decision-making during the course of the study regarding the conduct of the trial)[4] but is unavoidable. Although the direction of the treatment effect or the groups that were selected to continue into the confirmatory phase on the basis of interim analyses are known, the actual results of the interim analyses should not be made available beyond the members of the data monitoring committee (DMC).[11]

## Guidance

The guidance from FDA[4] and PhRMA[6] working groups on the combined phase II/III design are summarized in Table 1. The two organizations take different views on these designs. FDA has few recommendations specific to the design because the organization indicates that combined phase II/III design does not differ from other adaptive designs.[4] FDA does not use the term "seamless" or "phase II/III design" in its guidance on adaptive designs and provides an explanation about why it does not support this terminology.[4] FDA states that these terms do not add to the understanding beyond simply using the term "adaptive" to describe a well-designed exploratory study (e.g., one with multiple dosages or end points) with an interim analysis that permits an adaptation to a design that is similar to an adequate and well-controlled confirmatory study (e.g., an adaptation to a single comparison with a single-dose group or a single primary end point).[4] FDA feels that the term "phase II/III" leads to confusion about the original design of the study (i.e., whether the study was initially an exploratory study or a confirmatory study to demonstrate efficacy). FDA emphasizes the potential drawbacks of shortening the time between exploratory and confirmatory trials; PhRMA focuses on the potential benefits related to a shortened time between phases (Table 1). PhRMA recommends considering a phase II/III design when the design of the confirmatory (phase III) trial can be clearly envisioned and the follow-up to the end point is short relative to the duration of the trial.[6] They emphasize specific design aspects to help preserve the integrity of the trial design by reducing the risk of operational biases (Table 1).[6]

## Table 1: Recommendations for Phase II/III Designs from FDA[4] and PhRMA[6]

| FDA[4] | PhRMA[6] |
|---|---|
| • Although efficiency is gained by reducing the time lag between separate exploratory and confirmatory trials using the adaptive approach, there is also risk to this approach.<br>• A longer time between the exploratory and confirmatory trials may allow for the better design of subsequent confirmatory trials. For example, potential safety issues may be identified that should be addressed in the confirmatory trial.<br>• Adaptive designs that combine exploratory and confirmatory studies are felt to be lower risk when there is already considerable relevant prior experience that reduces the risk of unforeseen issues with a drug. | • The benefits of the phase II/III design relative to the conventional separate phase approach are emphasized and include:<br>◦ potentially substantial time savings<br>◦ the need for fewer participants to achieve the same quality of evidence<br>◦ the ability to obtain longer-term follow-up data by the end of the confirmatory phase for participants who started in the exploratory phase.<br>• It is recommended that the phase II/III design be strongly considered when:<br>◦ the design of the confirmatory phase of the trial can be clearly envisioned<br>◦ the follow-up time to the end point is short compared with the duration of the trial.<br>• Because these trials provide confirmatory data, it is critical that the processes for data review, decision-making, and implementation be carefully specified and adhered to.<br>• Interim trial results must be confidentially reviewed by a designated board without other trial responsibilities.<br>• At times, the nature of the trial and selection decision may justify sponsor participation in review, but without threatening the integrity of the trial. |

PhRMA = Pharmaceutical Research and Manufacturers of America.

## Adaptive Randomization

Randomization within controlled trial settings has been the cornerstone of the conduct of rigorous clinical trials since it was first introduced by Ronald Fisher in 1935.[15] The process of randomization involves the unbiased allocation of participants into two or more groups in a manner that balances both known and unknown covariates. Traditional approaches to randomization include simple randomization, stratification, and block randomization. Descriptions of these methods can be found in Table 2.[16]

## Table 2: Traditional Randomization Strategies[16]

| Method | Description | Comments |
|---|---|---|
| **Simple randomization** | • Participants are allocated according to a single sequence of random assignments. | • This form of randomization is the simplest and easiest, and is thought to be the least predictable.<br>• This method could result in a long run of the same treatment.<br>• There is a risk of covariate imbalance, particularly in smaller trials. |
| **Stratification or stratified randomization** | • This form of randomization accounts for the possible influence of covariates (baseline characteristics). Covariates that are deemed likely to influence the outcomes (e.g., gender) are identified, and separate blocks of allocations are created for each combination of covariates.<br>• Participants are then assigned to their appropriate block, and once all participants have been enrolled and assigned, a simple randomization procedure is used to allocate them to treatment or control groups. | • This method results in greater balance between treatment and control groups in covariates that were used as stratification variables.<br>• This method can be complicated if several confounding variables are identified.<br>• Covariates must be measurable in a timely fashion so as not to delay the randomization process. |
| **Block randomization** | • This technique is used to ensure that participants are randomized into groups of equal sample sizes, in which small blocks are created with predetermined group assignments. The size of blocks will be set as a multiple number of treatment groups, assuming that equal randomization to groups is desired. | • There remains the risk of covariate imbalance.<br>• This method can be used to help balance allocation over time or across geographic locations. |

Source: Suresh, 2011.[16]

Despite the merits of traditional randomization methods, there are some limitations to these approaches. First, it is possible that traditional randomization procedures can result in imbalances between treatment groups simply by random chance, particularly in trials with smaller sample sizes (i.e., fewer than 100).[17] Such imbalances may be in the treatment assignment or in the covariates across groups and can lead to an inflation of the error variance and, possibly, a failure to recognize a significant treatment difference when one exists.[18]

Adaptive methods of randomization have been designed such that parameters are set in which the randomization scheme can be altered after the study has started. Depending on the specific design, adaptive randomization may consider covariates (i.e., participant characteristics) and the allocation of the previous participant or outcome data in allocating the next participant (i.e., data that has already been collected influences treatment allocation). The adaptive randomization designs discussed here can be grouped into three categories, depending on the types of data that are used to decide how the randomization is altered (Table 3).

## Table 3: Adaptive Randomization Designs[10]

| Method | Description |
|---|---|
| Covariate-adaptive randomization | Allows the allocation probabilities to change as a function of the current distribution of covariates. |
| Treatment-adaptive randomization | Aims to ensure that there is balance with respect to treatment allocation by using a varied allocation probability. |
| Response-adaptive randomization | Uses observed treatment outcomes from preceding participants to change allocation probabilities. |

### Covariate-Adaptive Randomization

Covariate-adaptive randomization tries to balance the treatment groups with respect to key covariates. This balance is particularly important when certain covariates could influence the final results of the trial and, therefore, have prognostic significance. Imbalanced covariates can compromise the credibility of a trial, and adjusted statistical analyses (to deal with the imbalance) may create difficulties in interpretation. Further, if statistical adjustments are not planned a priori, there could be a risk of bias created when adjustments are made post hoc.

In covariate-adaptive randomization, also known as adaptive stratification, balance can be achieved using different methods, examples of which include minimization models, Wei's marginal urn design, and Frane's method (Table 4).[14] Minimization is the most common method of covariate-adaptive randomization.[19] Minimization may be useful for small to medium trials because it can balance a large number of covariates, which could not be achieved with other randomization strategies.[19]

There are several variations of minimization approaches for covariate-adaptive randomization, all of which are designed to minimize the imbalance in covariates. In its simplest form, minimization does not use a random process and is completely deterministic: each recruited participant is allocated to the treatment group that minimizes imbalance in groups.[20] Key algorithms for minimization are summarized in Table 4.

## Table 4: Methods for Covariate-Adaptive Randomization[14,19,21]

| Method Authors (Year of Development) | Description of Method |
|---|---|
| **Taves (1974)** | A minimization algorithm that compares marginal totals of corresponding covariates for each treatment group. The group with the lowest marginal total is where the next participant is allocated. |
| **Pocock and Simon (1975)** | A minimization algorithm that builds on the model proposed by Taves (1974). This method:<br>• temporarily assigns a participant to both groups and then calculates the absolute difference in marginal totals for comparison<br>• places the next participant in the group with the lowest sum of absolute differences among the covariates. |
| **Zelen (1974)** | A minimization model that follows a simple randomization sequence; however, once a predetermined threshold for balance is met, the next participant is forced to go to the group with fewer participants. |
| **Chen (2011)** | A two-way minimization procedure that assesses the "imbalance in the total numbers of participants" and the "imbalance in the distributions of prognostic factors" and is therefore considered both covariate adaptive and treatment adaptive. This method:<br>• chooses to minimize one of these two imbalances by probability when allocating a participant to a treatment group<br>• has appropriate type I error rates when performed correctly<br>• is most useful for small trials. |
| **Wei's marginal urn design** | Method that attempts to resolve, using a modified urn design, the potential for treatment imbalances that can occur in stratified procedures when there is a large number of strata with small strata sizes. |
| **Frane et al. (1998)** | Covariate-adaptive randomization method for both continuous and categorical variables that uses *P* values to assess imbalance between treatment groups. The smaller the *P* value, the greater the imbalance between treatment groups. This method:<br>• starts by assigning a new participant to either the treatment or the control group, and then calculates *P* values for each covariate using a t-test, with an analysis of variance for continuous variables and a chi-square test for categorical variables<br>• repeats this process by temporarily placing this participant in each of the other treatment groups; the minimum *P* value is then identified from all the test statistics, and the participant is ultimately allocated to this group. |

### *Limitations*

Despite the effectiveness of covariate-adaptive randomization for balancing prognostic factors, there are limitations to this approach. Given that the allocation of a participant is dependent on the balance of covariates of the preceding participants, the treatment allocation can be predictable.[22] Being able to predict the treatment to which a participant will be allocated can influence the decision of whether or not to enroll that participant, thereby introducing selection bias.[22]

Another limitation of covariate-adaptive randomization is that it depends on the correct determination of the participant's prognostic factors. Poor data quality (i.e., incorrect information on the covariates) can result in randomizing the participant to the incorrect treatment group. With minimization, imbalance may still arise from participants withdrawing from a study or being lost to follow-up after they have been randomized to treatment.[22] An important limitation of covariate-adaptive randomization is that the assumptions of the statistical inference of standard statistical tests are difficult to achieve because of the complicated probability structure arising from the strategies used for adaptive randomization.[3] As such, standard statistical tests may lack validity, although robust statistical test methods for covariate-adaptive randomization have become available.[19]

## Treatment-Adaptive Randomization

Treatment-adaptive randomization is used to help balance the treatment allocation, irrespective of covariates. Under treatment-adaptive randomization, the allocation probabilities are updated on the basis of the allocation history within the trial. Several techniques have been developed to increase treatment balance; urn randomization, Efron's biased coin, and adaptive coin randomization are summarized in Table 5. Although the Efron's biased coin method is easy to implement and maintains some aspect of randomization, this approach does not distinguish between large and small imbalances (i.e., regardless of the imbalance, the same probability constant is employed to help restore balance).[23] The adaptive biased coin model was developed to address this limitation.[24]

## Table 5: Methods for Treatment-Adaptive Randomization[14,24]

| Design | Method |
|---|---|
| **Urn randomization (Friedman 1949)** | • Equal numbers of two different coloured balls are placed in an urn, each colour representing a treatment group.<br>• Randomization is achieved by selecting a ball. The selected ball is replaced in the urn, and two balls of the opposite colour are added to the urn.<br>• The probability that the opposite treatment will be selected the next time increases, as does the likelihood of treatment balance.<br>• This approach is helpful in achieving balance in smaller trials and at the beginning of larger trials. Once trials reach a certain size, urn randomization behaves like simple randomization and therefore reduces the potential for selection bias as the sample size increases. |
| **Efron's biased coin** | • The existing treatment balance influences future treatment allocation.<br>• If balance exists, then the probability of the next participant being assigned to either treatment group is 0.50. If one group has more participants, then the probability that the next participant will be assigned to the opposite treatment will be a constant between 0 and 1 (0 < eta < 1).<br>• Any constant could be used, though ⅔ is thought to be optimal, and unlike the urn model, the probability does not change as a function of the degree of imbalance. |
| **Adaptive biased coin design** | • The design is based on Friedman's urn model; however, the existing balance is scaled by the total sample size to date.<br>• The ratio of the imbalance to the total sample size is then used to determine the probability that will be used.<br>• The greater the imbalance, the greater the probability that the next participant will be assigned to the group with fewer participants.[23] |

### *Limitations*

Treatment-adaptive randomization can have some limitations. Covariates are not considered in the balancing algorithm, which can create imbalances in prognostic variables in smaller studies. There is also a greater chance of predictability because the probabilities for treatment assignment are known in advance, possibly leading to selection bias, as previously outlined.[23] Similar to covariate-adaptive randomization, inferential statistical testing under treatment-adaptive randomization should consider that treatment allocation is not fully random.[25]

## Response-Adaptive Randomization

Response-adaptive randomization takes into consideration the treatment outcomes of preceding study participants to adjust allocation probabilities.[10] Traditional randomization strategies for two-arm studies may, for example, involve allocating participants into treatment groups in a 1:1 ratio. Response-adaptive randomization shifts the ratio in favour of the therapy with more favourable outcomes as the trial progresses. Typically, a probability threshold for accepting or rejecting the null hypothesis would be identified a priori, and once that probability is met, the trial is closed.[26] "Play the winner" and "randomized play the winner" are examples of approaches to conducting a response-adaptive randomization trial

that have been used in practice (Table 6). Response-adaptive randomized designs are most efficient when the disease is rare and the anticipated differences between the treatments are large.[27]

## Table 6: Response-Adaptive Randomized Designs[9]

| Design | Method |
|---|---|
| **Play the winner** | • The trial starts by using a simple randomization procedure, and each participant's response is labelled as a "success" or a "failure."<br>• If a participant's response is a success, then the subsequent participant enrolled will be given the same treatment; if the response is a failure, then the subsequent participant will be given the opposite therapy.<br>• The trial stops when a predefined number of failures is documented or when the trial reaches its target sample size. |
| **Randomized play the winner** | • An urn is filled with balls labelled either "T" (treatment) or "C" (control).<br>• At the beginning of the trial, there is an equal number of Ts and Cs.<br>• As participants are recruited, if they are deemed to have success with treatment or to have failed the control, then a pre-specified number of "T" balls are added to the urn.<br>• This process increases the probability of receiving the more promising therapy while maintaining the possibility of receiving the control. |

*Limitations*

Response-adaptive can be administratively difficult to implement and manage. For large trials or trials with a long duration of treatment, feasibility becomes an issue given that the randomization of a participant is dependent on the response of the previous participant, and this process also creates a risk of selection bias.[3] Like covariate-adaptive randomization, predictability is also a concern, as recent outcomes influence the probability of being allocated to certain groups. With response-adaptive randomization, it is also difficult to calculate appropriate sample sizes in advance of the study because the allocation and size of treatment groups are dependent on the participant outcome. As well, the ability to maintain control over the type I error rate in response-adaptive designs is complex.[26] A further limitation of response-adaptive randomization is that significant imbalances of covariates may exist between groups at the end of the trial. The smaller the sample in the inferior treatment group, the more magnified the imbalance in covariates can become.

Bias can be introduced in a study that uses response-adaptive randomization if healthier participants decide (or their physicians decide on their behalf) to delay enrolment in a trial to increase the likelihood of receiving the superior treatment. Sicker or refractory participants tend to be enrolled earlier in clinical trials. This situation can create an imbalance in participants and a biased effect estimate since the probability of being randomized to the superior treatment increases later in the trial with adaptive randomization[10] and decreases the generalizability of the findings. Enrolling healthier participants later in the trial also has the potential to increase the study duration, as the healthier participants who are recruited later will tend to have a longer time to event, which can delay the release of results.

### Guidance

FDA's *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics* includes commentary on response-adaptive randomization designs, but not on covariate-adaptive or treatment-adaptive trial designs.[4] The European Medicines Agency and PhRMA have not put forth any recommendations or statements on adaptive randomization strategies.[6,7]

According to FDA, response-adaptive randomization designs, such as the "play the winner" model, are "less well understood." In the guidance document, FDA indicates that such methods are valuable in exploratory studies, as they may permit more dosage options to be studied, but states that such designs should be used with caution in confirmatory studies. FDA highlights several key limitations, one being that the analysis is not as easily interpreted compared with cases where fixed randomization probabilities are used. FDA further emphasizes the need to control the type I error rate and avoid bias. Concern is also raised that changing allocation probabilities will result in imbalances in participant characteristics (both known and unknown) and that if those characteristics are associated with the outcome, the estimated difference in treatment effect between groups could be biased.[4]

## Adaptive Group Sequential Design

In a conventional group sequential design, an interim analysis is conducted at one or more planned points in the trial to assess efficacy and safety, and to identify a lack of treatment benefit (i.e., futility), which would make continuing the trial unethical and lead to early termination. In clinical trials that assess therapies of unknown clinical benefit, it is ethical to monitor data to ensure that participants are not exposed to unnecessary risks. It may be necessary to stop a trial early if there is evidence that the therapy is ineffective or unsafe. Group sequential designs are considered by FDA to be "well understood" if they use accruing study data in a planned and confidential manner (i.e., by a DMC), with appropriate control of the type I error rate, and maintain study integrity.[4]

A group sequential design can incorporate additional, planned adaptive features at the time of interim analysis, such as sample size re-estimation and changes to the study arms (modifications, deletions, or additions), study end points, dosage or the duration of treatment, or the randomization schedule.[4] These study designs are referred to as adaptive group sequential designs. For example, one or more unblinded interim analyses of the apparent treatment effect could be carried out, and groups that meet the prospectively defined futility criteria would be terminated early from the study. When these additional adaptive features are incorporated, the design may be considered by FDA to be "less well understood," depending on the adaptation that is made.[4]

Blinded analyses are "those in which the treatment group assignments of study subjects are not known and are therefore not used in any manner in the analysis" and unblinded analyses are "those in which the treatment group assignments of subjects are known and used in some manner in the analysis, usually (but not always) as a formal comparison between treatment groups."[4]

## Limitations

In a group sequential design, multiple "looks" at the data will increase the risk of type I error (unless adequately controlled) and can potentially introduce operational bias if there is knowledge of the interim analysis.[4] The implementation of multiple points of interim analysis can be particularly troublesome. Different approaches can be taken to control the type I error rate in a group sequential design. Haybittle and Peto maintain a very stringent alpha level throughout enabling the use of a *P* value at or close to 0.05 at the final analysis.[14,28] With this approach, it can be difficult to stop a trial early. O'Brien and Fleming use an approach that makes the level for significance less and less stringent in each analysis while still reserving the majority of the alpha for the last analysis.[14] The advantage of this approach is that the trial is more likely to stop early than with the Haybittle and Peto approach but allocates only a small amount of the total nominal significance to the early stages of interim analysis.[14] The Lan and DeMets alpha spending function is an approach to maintaining the type I error rate in a group sequential design where the number of interim analyses does not need to be pre-specified.[29]

Another major limitation of group sequential designs is that stopping a trial early because of unanticipated large treatment effects may overestimate the true effect, particularly early in a trial, when sample sizes are smaller and estimated treatment effects can be quite variable.[4,14] This limitation is, in part, why some alpha spending functions for interim analyses make it more difficult to stop a trial in the earliest phases. FDA only supports the early termination of trials when there is a combination of compelling ethical concern (improved survival or the avoidance of irreversible disability) and robust statistical evidence.[4] However, FDA further states that it is difficult to determine if the observed treatment effect is reflective of "random highs" related to the limited amount of data early in the trial.[4] The overestimation of treatment effect that can occur early in a trial also presents a challenge for adaptive group sequential designs when adaptations to the design (e.g., dropping treatment arms) are made on the basis of those treatment estimates.[4,14]

The early termination of a trial for efficacy reasons reduces the sample size and trial duration (and follow-up), which will also results in the loss of safety data.[4,14] Further, the reduced sample size may mean that there are fewer participants available for important pre-specified subgroup analyses.[4,14]

## Guidance

Despite the potential for limitations, FDA considers group sequential designs to be "well understood" if they are based on "unblinded interim analyses of accruing study data that are used in a planned and confidential manner (i.e., by a DMC [data monitoring committee]) that controls type I error and maintains study integrity."[4]

For the group sequential methods to be valid, FDA states that they must adhere to the prospective analytic plan, terminating groups if a futility criterion is met and not terminating the study for efficacy unless the prospectively defined efficacy criteria are satisfied. It should be noted, however, if the trial is not conducted according to the FDA guidance, a group sequential design would not be considered "well understood."

The PhRMA working group has not put forth any recommendations or statements on group sequential designs.

## Sample Size Re-Estimation

Sample size re-estimation is an adaptive design that permits the sample size to be adjusted or re-estimated on the basis of an interim analysis of observed data from a study.[9] For a superiority study, the estimation of a study's sample size is based on values obtained from previous studies, from literature values or from consulting with content experts. As a result, the parameters are only an approximation of the anticipated value that will be observed in the study. Inaccurate estimates for these variables can result in an underpowered or overpowered study, leading to the inefficient use of resources. For example, a study may be originally designed to detect a larger treatment effect than is actually observed in a clinical trial. Regardless of whether the smaller treatment effect is still clinically relevant, the trial may be underpowered to detect the difference statistically. A clinical trial may be similarly underpowered if the variance used in the sample size calculation is lower than the variance observed in the actual trial.[30]

A sample size re-estimation design enables researchers to update their parameter estimates on the basis of accumulated data from the ongoing trial to estimate a more accurate sample size. Group sequential methods can be used for sample size re-estimation at planned points within the trial, the timing of which can be defined by the number of participants recruited along with their follow-up period, or on the basis of a fixed number of events.

Statistical techniques must be applied to control the type I error rate when methods of sample size re-estimation are used in a clinical trial.[4] As well, it is important to consider that in the interim analysis, short-term data will be predominant and the treatment effect may vary over time. Thus, the treatment effect from the interim analysis could potentially differ from the treatment effect expected at the study conclusion. As such, FDA recommends acting conservatively when deciding to make changes to sample size based upon early estimates.[4]

### Limitations

With sample size re-estimation, there is a risk of type I error rate inflation, which must be adjusted for in the analysis.[9] When using a group sequential approach to sample size re-estimation, the limitations highlighted in the previous section on group sequential designs also apply. An additional concern with sample size re-estimation is the ability of blinded investigators to predict the treatment effect. That is, investigators may become effectively unblinded with all the associated risks for selection bias and potentially inflated estimates of effect based upon changes to the sample size. Further, there is the risk that the trial's interim result, which is based on a limited number of participants, will be considered to represent the "true" treatment effect at the study completion when, in fact, it could differ considerably. The variability in event rates or treatment effect at different points in a study's timeline can be substantial; therefore, such early data should be interpreted cautiously,[4] as misinterpretation can result in misleading or biased sample size re-estimation. Thus, it remains possible that despite having re-estimated the sample size, a clinical trial could potentially remain underpowered at its completion.

### Guidance

According to FDA, sample size re-estimation based solely on the interim analysis of nuisance parameters (any parameter, other than the treatment effect, that can affect the overall power of the study) is considered to be a "well understood" adaptation, and sample size re-estimation that uses unblinded data of treatment effect is considered to be "less well understood."[4] Although the PhRMA working group does distinguish between methods, it recognizes the importance of sample size re-estimation for trial efficiency and has put forth recommendations for maximizing benefits while maintaining integrity.[6] Table 7 summarizes the recommendations from each document.

## Table 7: Recommendations for Sample Size Re-Estimation

| FDA[4] | PhRMA[6] |
|---|---|
| • Analysis should be conducted in a blinded manner when possible to prevent the introduction of bias.<br>• Since sample size re-estimation using a blinded assessment of nuisance parameters can be done while controlling for type I error, it should be considered in most clinical trials.<br>• The timing of a sample size re-estimation should be considered carefully, as the variability of the estimate can be substantial at different points in the trial.<br>• Sample size re-estimation strategies should be reserved for increasing study size and not decreasing study size, as there is the risk of making the wrong choice because of the high variability in effect size throughout the study. | • "The need for sample size re-estimation should be carefully evaluated during trial planning, and the extent to which it is planned to re-evaluate sample size should be described in the protocol. Sample size re-estimation should never be a substitute for adequate up-front planning; rather, it is an acknowledgment of potential limitations of the information available at the time of trial design."<br>• "These methods should generally be implemented minimally within the trial as needed to achieve a more satisfactory sample size. A single re-evaluation may suffice, particularly for nuisance parameters. If a minimum sample size has been prespecified, then it will often be sensible to perform the re-estimation shortly before reaching that minimum enrollment."<br>• "Logistic concerns must be adequately planned for in advance, such as the potential need for additional drug supply if the sample size is increased."<br>• "Where relevant, consider whether to withhold from the protocol and document elsewhere details of the re-estimation procedure, to decrease the amount of information that observers can infer from any changes made to the sample size."<br>• "We recommend that sample size re-estimation based on nuisance parameters should be routinely considered, particularly when there is a good deal of uncertainty about those parameters or the sample size is very sensitive to initial assumptions. Frequently, this can be addressed sufficiently well in a blinded manner and this should often be the recommended approach, because re-estimation can then be implemented in a manner that does not compromise the trial and which minimizes operational difficulties."<br>• "Methods exist for sample size re-estimation based on updated information on treatment effects obtained at interim analyses and these can be considered."<br>• "However, such methods must be applied cautiously. They can introduce operational biases because 1) they may provide to observers an unacceptable amount of information about the interim effects; 2) different values of the treatment effect may not be of the same level of clinical or commercial relevance; or 3) interim estimates are often too imprecise to be used efficiently in this regard." |

| FDA[4] | PhRMA[6] |
|---|---|
| | • "As an alternative, it should be strongly considered whether an appropriate group sequential scheme as a method of sample size determination could better meet the study objectives, as operational concerns are minimized and statistical behavior is often superior." |

PhRMA = Pharmaceutical Research and Manufacturers of America.

# Stepped Wedge Cluster Randomization (Novel Design)

Stepped wedge randomization aims to allow the comparison of policy interventions or service delivery interventions.[31] The stepped wedge cluster randomized trial is a newer study design that is a modification of the more traditional cluster randomized trial in which clusters of individuals within a larger unit (e.g., a geographic area, a hospital, or a school) are all randomized to receive the same treatment (i.e., the unit is randomized so that all the individuals in that unit receive the same intervention). The stepped wedge design is similar in the sense that the units of randomization are larger than the individual, but rather than the units being randomized to receive one intervention or another, the units are randomized to receive the novel intervention beginning at different times and act as control units up until the designated switching time. All units have a run-in period in which they receive standard therapy so that the intervention-control comparison can be made both within units and between units. At regular intervals (referred to as the "steps") a cluster (in some cases a group of clusters) is then randomized to cross from the control to receive intervention that is being studied.[31] The rationale for cluster randomization, in general, is that certain types of interventions cannot be randomized at the individual level without a high risk of contamination between individuals within a unit.

From a practical perspective, the stepped wedge design overcomes a shortcoming of the traditional parallel cluster trial in the sense that all the participating units in the study will eventually get the study treatment.[31] Two systematic reviews indicate that the use of this design is increasing in popularity and being used in diverse circumstances.[32,33]

## Limitations

There are some technical challenges to designing a stepped wedge trial, depending on the particular characteristics of the design. In the design phase, the number of clusters, number and length of steps, and number of clusters randomized at each step must be determined,[31] which can be challenging. Further, methods for power and sample size calculations are only available for cross-sectional designs, in which it is assumed that the individuals being measured within a unit are different from one period to the next.[31] As well, it is necessary to adjust the calculation for the degree of correlation between individuals within a cluster (the intracluster correlation coefficient); this coefficient is often unknown.[34] From a statistical power perspective, the more usual parallel cluster trial is preferred to the stepped wedge when the intracluster correlation is small.[35]

Studies that use stepped wedge randomization and have a requirement for participants to give consent may have a potential for bias as a result.[36,37] There is still debate about the necessity to recruit and obtain participant consent at the individual level when investigating a new mode of delivery. If individual consent is required, it would be best to recruit

participants before they are aware of treatment allocation, which is often not feasible. Clusters must also be ready to cross to the intervention according to the randomization, and this move is not always possible.

Carry-over effects typically are not a concern in trials that use a stepped wedge design, as the control condition precedes the intervention (i.e., crossover is always from control to intervention). However, there may be confounding between treatment and time. At the beginning of the study, few clusters will be receiving the intervention and many will be receiving the standard care or treatment. Toward the end of the study, the reverse is true. Thus, any time trends that have an impact on the outcomes being measured must be adjusted for.[38] Further, if the study population has a medical condition that is unstable and fluctuates during the control condition, this fluctuation could affect the response to the intervention.[39]

## Enrichment Designs (Conventional and Adaptive Approaches)

Enrichment designs are used to focus recruitment on those participants who are more likely to respond to the therapy under investigation. Enrichment is defined as the "prospective use of any participant characteristic to select a study population in which detection of a drug effect is more likely than it would be in an unselected participant."[5] Enrichment characteristics should almost always be specified before study initiation,[5] and would be defined as part of the inclusion and exclusion criteria. This is the conventional enrichment design. When enrichment characteristics are identified after study initiation, the study design is an "adaptive enrichment" study.[5]

Enrichment designs can be applied to personalized medicine, where biomarkers or specific participant characteristics (e.g., genetic, demographic, or physiologic) are used to identify those participants who are more likely to respond to treatment (or experience an adverse event). Specific enrichment strategies used in randomized controlled trials can be divided into three main categories, as outlined in Table 8.

## Table 8: Main Categories of Enrichment Strategies[5]

| Strategy | Description | Comments |
|---|---|---|
| **Strategies to decrease heterogeneity** | • Participants are selected with baseline measurements in a narrow range. This selection measure decreases interparticipant variability.<br>• Participants are excluded if their disease or symptoms improve spontaneously or if their measurements are highly variable. This exclusion measure decreases intraparticipant variability.<br>• Study power is increased by the decreased variability resulting from these strategies. | • The key purpose is to increase study power by decreasing non-drug-related variability.<br>• These approaches are widely used in practice and are not adaptive. |
| **Prognostic enrichment strategies** | • Participants are selected who have a higher likelihood of a disease-related end point event (for event-driven studies) or a substantial worsening in condition (for continuous measurement end points).<br>• These strategies increase the magnitude of the absolute effect difference between groups but will not alter the relative effect. | • The prognostic factors may include clinical and laboratory measures, medical history, genomic measures, and proteomic measures.<br>• These strategies allow the treatment effect to be more easily discerned. With higher event rates, the power is higher as well. |
| **Predictive enrichment strategies** | • Participants are selected who are more likely to respond to the drug treatment than other participants with the condition being treated.<br>• This selection can lead to larger absolute and relative effect sizes and permit the use of a smaller study population. | • The selection of participants may be based on a specific aspect of a participant's physiology or a disease characteristic that is related to the study drug's mechanism, or selection could be empiric (e.g., the participant has had a prior response to a drug in the same class). |

### Strategies to Decrease Heterogeneity

Enrichment to reduce sample heterogeneity is not an adaptive approach and is widely used in clinical trials in the form of the a priori inclusion and exclusion criteria to reduce non-drug-related variability.[5] Narrowing the characteristics of the sample population increases the power of the study and, therefore, the ability to detect a difference if one truly exists. This approach also increases the efficiency of the clinical trial by decreasing the sample size required to demonstrate an effect, given the reduction in the variance of the primary outcome being measured.

Some strategies that might be considered include:[5]

- clearly defining the disease characteristics that participants must exhibit to be included, and ensuring that participants meet those criteria
- identifying participants who are more likely to adhere to therapy and clinic visits or who are less likely to drop out
- excluding participants who improve spontaneously by using a placebo run-in period before randomization
- minimizing baseline variability by only recruiting participants with consistent values for key baseline characteristics measured over a pre-specified run-in period
- excluding participants who are using other pharmacotherapies or treatment modalities that are similar to, or likely to interact with, the study drug

- excluding participants who have a high risk of intolerable side effects
- excluding participants who are likely to drop out for non-medical reasons.

### Limitations

Although making efforts to decrease the heterogeneity of the study sample may improve trial efficiency and help to identify those participants who are more likely to benefit from a given therapy,[5] there are some limitations to this approach. Although very narrow selection criteria can increase the likelihood of demonstrating efficacy within a tightly defined sample,[5] such restrictions can potentially decrease the external validity and generalizability of the study results to the general clinical population. Moreover, studies of wider populations may not be carried out.[5] As such, it remains unclear if similar treatment effects could be anticipated in a more broadly defined participant population.

### Guidance

In its guidance document on enrichment strategies, FDA suggests that the approaches to decreasing heterogeneity as outlined in the previous section are useful and generally accepted in enrichment designs.[5] However, for other approaches (e.g., excluding participants with comorbidities that could have an impact on survival, or broadly excluding participants on concomitant therapies), concerns have been raised that studies with these exclusions may be too restrictive (which would limit the study's generalizability to the broader patient population).[5] FDA states that, under such restrictions, there is a risk that too little information may be provided about the broader participant population that would receive the treatment in practice.[5]

## Prognostic Enrichment Strategies

Prognostic enrichment strategies use prognostic indicators to identify participants who are most likely to experience an outcome of interest in a study.[5] Investigators generally seek to enroll participants at the greatest risk for the primary outcome such that, if a treatment or intervention is efficacious, there is a greater likelihood of detecting a benefit with a given sample size. Generally, the larger the absolute effect size, the smaller the sample size required to detect a statistically significant difference. By focusing on a high-risk population, fewer participants are required for the detection of a significant difference between treatment alternatives. While focusing on a high-risk population potentially translates into a greater absolute effect in the subpopulation enrolled, typically there is no impact on the relative effect. This type of trial design is particularly useful for examining preventive therapies with indications such as cardiovascular disease (i.e., stroke or myocardial infarction prevention) or oncology.[5]

### Limitations

To capture participant populations that are more likely to experience an outcome, a balance is required between identifying those who are at higher risk and are therefore good candidates for participating in the trial versus those who are at high risk of rapid deterioration and are therefore beyond the point at which therapy can potentially improve outcomes.[5] Thus, when interpreting the findings of an enriched clinical trial with particularly restrictive inclusion and exclusion criteria, it is critical to consider how these criteria limit the generalizability of the results to populations more broad than those included in the trial. Given the overlap between the limitations of prognostic and predictive enrichment strategies, further limitations are discussed later in this report.

### Guidance

FDA provides no specific guidance on prognostic enrichment strategies but provides examples of their use.[5]

## Predictive Enrichment

Predictive enrichment studies aim to focus on those participants who are more likely to respond to therapy; the studies can involve approaches that are both adaptive and not adaptive in design. Predictive characteristics might include factors such as disease presentation, gender, age, genetic or nongenetic biomarkers, or a history with or response to previous therapies.[5] Like prognostic strategies, narrowing the study population to those more likely to respond translates to a smaller required sample size and, therefore, a more efficient clinical trial. Additionally, employing predictive strategies has the potential to increase the benefit-to-risk ratio for participants by avoiding exposure and potential toxicity in participants unlikely to respond.[5]

Predictive enrichment strategies are particularly useful in early clinical trials, as these strategies can help to establish proof of concept and target dosage selection.[5] These strategies are also commonly used in pharmacogenetics and in circumstances in which response rates are traditionally low (e.g., oncology). By focusing on the potential responders, there is an increased chance of detecting an effect that may not be possible to detect in the broader disease population. Unlike prognostic enrichment designs, in predictive designs there is the potential to detect both a larger absolute effect and a larger relative effect.

Specific predictive enrichment strategies can be categorized into five groups:[5]

- genomic strategies
- empiric strategies
- pathophysiologic strategies
- randomized withdrawal studies
- studies in nonresponders or patients intolerant to other therapies.

### Genomic Strategies

Genomic enrichment strategies typically involve tumour genomics and can be applied pre- or post-randomization. When implemented post-randomization (according to an a priori plan) genomic enrichment strategies are considered adaptive (i.e., biomarker-adaptive designs).[5] The strategy of integrating a biomarker evaluation for the enrichment of the study sample, particularly for prediction, can be a reasonable strategy when the performance characteristics of the biomarker are well established.[5] From a design perspective, multiple strategies can be applied. The simplest form (a nonadaptive approach) is to restrict enrolment to only participants who are biomarker positive, which suggests a prediction of response.[40] A simple extension of this approach is the biomarker stratified design, in which both biomarker-positive and biomarker-negative participants are randomized to the treatment or control. This design allows for the evaluation of the effect in both biomarker-positive and biomarker-negative participants; however, it is most often designed for a definitive assessment of the effect in biomarker-positive participants and so is powered as such, incorporating one of several different analytic strategies with the appropriate allocation of alpha error across multiple tests.[5]

There are other approaches for the evaluation of a genetic biomarker for enrichment. One involves testing the treatment effect in biomarker-positive participants at significance level alpha and, if significant, testing the treatment effect in biomarker-negative participants[41] or the overall population.[42] The most complex approach is the marker sequential test design — which incorporates analyses of biomarker-positive participants, biomarker-negative participants, and the overall population into the final analysis, with the effect in each group tested sequentially at a given alpha level — in which each subsequent test is dependent on the previous test.[43,44] To make this process more efficient, Freidlin and Simon designed a novel study approach that combines the genetic marker discovery and confirmatory trials into one.[5,44] This process is described in Table 9.

## Table 9: Freidlin and Simon Approach[5,44]

| Step | Procedure |
|---|---|
| 1 | Design the study as usual, but divide it into first and second halves. Prospectively allocate the overall study alpha as 0.04 for the whole population and 0.01 for a participant subset to be identified in the first half of the study. |
| 2 | Run the first half of the study and conduct unblinded data analyses, searching for a genetic predictor of response. There would be no limit to the number of such analyses conducted. A single genetic subset appearing to predict response may be identified. |
| 3 | Complete the remainder of the study, entering participants according to the original eligibility criteria (both the predicted responders and predicted nonresponders) as before. |
| 4 | At the conclusion of the study, test the effect in the entire study population at an alpha of 0.04; the genetically identified subset is tested only in the second half, at an alpha of 0.01. |
| 5 | Determine evidence of effectiveness (the study shows such evidence if either analysis is positive). When the responder population is a small fraction of the total population but exhibits a large response, this design can improve the chance of detecting a treatment effect. It also retains good power for the overall study if the drug is more broadly effective. |

In addition to the example in Table 8, FDA outlines the following potentially applicable adaptive enrichment designs in which the sample size may change after the study initiation, or in which other planned adaptations may occur on the basis of information accrued during the trial:[5]

- In a study with biomarker-positive and biomarker-negative participants, an interim look at the data could demonstrate that the biomarker-negative population has a much lower response rate, leading to a reduced or terminated enrolment of the biomarker-negative group.
- When the optimal marker cut point is not well known before the study, an adaptive design could be used to obtain more precise information on the performance classifying characteristic (e.g., by using several pre-specified thresholds and correcting multiplicity issues).
- Interim analyses could suggest changing entry criteria to emphasize a better-responding subgroup.

### Nonadaptive Predictive Enrichment Strategies

The remaining predictive enrichment strategies that do not involve adaptations are summarized in Table 10. Each approach is described in the table and some general comments are provided. General limitations and key considerations for economic analysis are found in the sections that follow.

## Table 10: Predictive Enrichment Strategies[5]

| Strategy | Description | Comments |
|---|---|---|
| **Empiric strategies** | • The selection of participants is not based on the baseline characteristics but rather on data collected on therapy response during a screening period or prior experience with related therapies.<br>• Participants who are recruited for a study are given open-label therapy, and those who achieve a predetermined outcome or biomarker are then randomized to either active therapy or placebo.<br>• These strategies may give all participants the open-label drug, identify responders, withdraw the drug, and then randomize. | • These strategies are **useful** when it is known, **a priori,** that there will likely be a very low response rate; however, these strategies raise some ethical concerns.<br>• It is impossible to assess the risk of adverse events that occur in a screening period because of the lack of a control group.<br>• There is also a risk that adverse event rates will be underestimated at the end of the trial if those who are intolerant to the therapy drop out or are excluded before the randomization.<br>• This study design gives very specific evidence of potential treatment response rates, particularly with treatments for which a low initial response rate might be expected. |
| **Pathophysiologic strategies** | • These strategies require prior knowledge of a participant's physiological characteristics or disease pathophysiology when it is suspected that a drug will only be effective in participants with certain characteristics.<br>• These characteristics might relate to a participant's ability to metabolize a drug, the effect on tumour metabolic response, or the presence of proteomic markers (i.e., test and treat).<br>• It is important for investigators to carefully define these characteristics and the tests used for their detection. | • Tests used to detect certain markers should have acceptable sensitivity and specificity.<br>• An inaccurate marker can undermine the enrichment exercise and be detrimental to determining the efficacy of a particular therapeutic agent for a specific indication.<br>• The application of an inaccurate marker can potentially result in the erroneous rejection of an effective intervention for a specific indication (i.e., committing a type II error), thereby resulting in the erroneous rejection of a potentially beneficial treatment. |
| **Randomized withdrawal studies** | • Only participants deemed to be responders are included in the trial.<br>• Responders can be identified using an open-label period or can be selected from a treatment group of a randomized trial.<br>• Participants are typically given active treatment for an extended duration and then randomized, in a blinded fashion, to either remain on therapy or be withdrawn from therapy and receive a placebo.<br>• The study is designed to remove participants from treatment if a certain end point is achieved (i.e., the return of disease or symptoms) and, therefore, reduce an extended exposure to ineffective treatment. | • **This design could be used to generate long-term dose-response data.**<br>• **These studies are useful in determining the duration of the benefit of long-term treatments, particularly for drugs with long-term toxicity.** |

| Strategy | Description | Comments |
|---|---|---|
| **Studies in nonresponders or participants intolerant to other therapies** | • Enrichment is achieved by recruiting participants who failed or were intolerant of an existing drug.<br>• Participants are randomized to a new drug or the existing or failed drug.<br>• The selected participants are less likely to respond to an existing therapy than participants from an unselected population, which gives the study an enrichment advantage in that if an unselected population were used, many would respond to the existing therapy, requiring a larger sample size to show a treatment effect (difference between the new and the existing or failed drug).<br>• The findings from these studies are useful for comparisons between drugs within the same pharmacological class (an existing drug that participants failed on versus a newer agent in that class) or to show that a drug from a different pharmacological class is useful in participants who fail on the previous treatment. | • This design can be used to demonstrate that different drugs used to treat the same condition can have different responder populations.<br>• The design can also be used to demonstrate that a new drug may be superior to existing therapies, but to statistically demonstrate this superiority in an unselected population would be difficult because of the large sample size required.<br>• The use of a selected population can reduce the sample size to enable a demonstration of the comparative effectiveness.<br>• Thus, the results of studies in nonresponders may prove useful for the evaluation of new treatment options that may be considered as second- or third-line alternatives for use in participants who do not respond to the first-line standard of care. |

## Limitations of Predictive Enrichment Strategies

The careful interpretation of study results is critically important when the study has enrolled an enriched population, as the potential generalizability to more broadly defined participant populations may be limited. As mentioned, both prognostic and predictive enrichment will likely overestimate the absolute effect of an intervention.[5] The magnitude of the absolute treatment effect in the broader participant population is not anticipated to be as large as that observed with a study population that has been enriched using prognostic or predictive strategies.[5] However, the relative treatment effect is generally unaffected when a study population has been selected using a prognostic enrichment strategy.[5] Predictive enrichment, however, is thought to overestimate the relative treatment effect as well in most cases.[5] Although these limitations of prognostic and enrichment design do not threaten the internal validity of the trials, they do result in concerns over the generalizability of the results.[5] When including studies with enrichment designs in systematic reviews with meta-analyses, consideration should be given to the clinical heterogeneity created by enrolling a selected population and whether the enriched study population is sufficiently homogenous with non-enriched study populations to permit pooling in meta-analysis. Further, in some conditions, enriched populations have been shown to have lower rates of adverse effects in meta-analysis.[45]

Adaptive enrichment designs have some key limitations. The designs may be too complex to operationalize in practice, can have limited generalizability, and provide no information on efficacy in the excluded groups.[10] The integration of biomarker evaluation within a specific study design results in concerns related to multiple testing and the increased risk of type I error.[5] Thus, it is imperative that consideration be given to multiple testing and that appropriate statistical techniques are employed to control or adjust the type I error.[5]

Moreover, sample size planning with adaptive enrichment can be difficult when there is uncertainty about the prevalence of a marker and its accuracy in prediction and the sample

size or entry criteria adjustments that may be made during the course of the trial.[5] Statistical simulations may be needed to estimate the anticipated power associated with different design choices.[5]

### Guidance for Predictive Enrichment Strategies

Enrichment strategies can be used in clinical trials from the learning to confirmatory phases.[5] As a regulatory body, FDA is supportive of the use of enrichment design strategies as a means of defining more targeted populations for therapies as well as identifying those populations that will not benefit or that are at risk of harm.[5] The important factor is ensuring the appropriate identification of the selected population and sufficient data collection on the unselected population.

When the likely responders can be identified in advance of treatment, such as those with some pathophysiologic or genomic characteristics, the procedure for the recruitment of the study population can be straightforward.[5] It is much more difficult to select a target population when the distinguishing characteristics are not as well understood or when screening tests are inaccurate. With less than ideal sensitivity and specificity, there is a risk that participants who could benefit would not be selected and those who will not benefit will be selected.[5] FDA suggests that this limitation should not discourage enrichment designs, as using an entirely unselected participant population may result in not finding a treatment effect for a promising therapy in certain subgroups.[5] FDA does, however, recommend that caution be taken in proceeding when the likely responders cannot be confidently selected, and indicates that statistical adjustments may be required to prevent an increase in the type I error rate, as previously discussed.[5]

From a drug approval perspective, enrichment designs provide fewer data on the effects of a treatment in an unselected population.[5] Although participants with certain characteristics may be identified as superior responders, there is still the possibility that some other participant groups may respond marginally, while other groups may be at greater risk of harm.[5] It is important to understand these risk-benefit ratios before implementing the therapy in clinical practice. FDA states in their guidance to industry on enrichment designs that the approval of a drug will not be delayed when it is tested only in an enriched population if the benefits to the selected group are substantial.[5] However, they also state that it is important to clarify with FDA beforehand how much information may be needed in an unselected population before drug approval, and that postmarket commitments (i.e., safety and efficacy studies) may still be required in a broader population.[5]

## Implications for Critical Appraisal

The use of adaptive designs can have implications for the interpretation of research findings from a randomized controlled trial. As a result, it is important to be able to recognize if any adaptive and novel features have been incorporated into the design of the randomized controlled trial or if the study population has been enriched using a conventional design. Identifying an adaptive or novel design feature or an enriched population may require a review of the study's research design, inclusion and exclusion criteria, and statistical methodology based on the publication or report; however, the trial design adaptations and features may not always be clear. Enrichment studies, for example, may be identified through run-in periods in which participants may be selected or excluded on the basis of their response to therapy. Extensive inclusion criteria and exclusion criteria may also be suggestive of an enriched study population, as may the exclusion of a relatively large

proportion of participants. Interim analyses and statistical procedures to adjust or control the type I error rate due to numerous statistical tests related to multiple "looks" at the data may also be suggestive. Changes to inclusion or exclusion criteria should be scrutinized to determine if they were planned adaptations (i.e., adaptations by design) or performed after a protocol amendment that was not pre-specified in the planning phase of the trial.

Once an adaptive or novel design has been identified, consideration should be given to the implication that the particular design feature has for the impact on the estimated treatment effect, generalizability, and risk of type I error, as well as the methods employed to mitigate against any risks. With prognostic and predictive enrichment designs as examples, the enrichment itself does not threaten the internal validity of the trials but does increase concerns over the generalizability of the observed treatment effects. Although concluding a positive benefit based on the results of such a trial would be correct, the extrapolation of the estimated effect size to a broader population must be carefully assessed. Although internal validity is critical when determining the efficacy of an intervention, the generalizability of results is a key consideration.

Enrichment strategies incorporated into adaptive clinical trials have been viewed as advantageous in being able to identify specific subgroups for whom a given intervention can be deemed effective and for facilitating the estimation of an undiluted effect size. However, these positive characteristics of enrichment must be weighed against concerns about the potential overestimation of the treatment effect, the lack of information on the excluded groups, and, thus, the lack of generalizability of the results beyond the enriched population.

Some clinical trials with adaptive features and designs may present data from interim analyses that do not reflect the actual primary and secondary end points of the study. Group sequential designs and sample size re-estimation are examples of such designs. These results are based on accumulating study data and have the potential to be biased for various reasons previously outlined in this paper. These estimates may not be generalizable to the overall study population (they can be overestimates of treatment effect) and may not reflect the treatment effect that is achieved at the study completion. As well, as described earlier, stopping a trial prematurely for a treatment difference will tend to overestimate the treatment effect had the study gone to completion.

It is also important to recognize adaptive randomization and the associated implications or potential for bias. As previously discussed, adaptive randomization strategies have the potential to increase the risk of introducing operational biases, particularly in the selection of participants. Some adaptive randomization approaches, although technically still concealed, can make the allocation sequence predictable, which could lead to selection bias.

Finally, enrichment, adaptive, and novel design features may result in the selection of participant populations that are less reflective of the overall target population for a given condition. Further, the distribution of clinical characteristics within the trial may differ from other studies because of adaptive randomization schemes. This issue has implications for meta-analysis and network meta-analysis in terms of clinical and statistical heterogeneity and may limit the ability to include studies with adaptive features into pooled analyses. Further, trials that stop early on the basis of group sequential procedures may overestimate the treatment effect,[4,14] which can be a potential source of heterogeneity. It is unclear if enriched, novel, and adaptive trials are consistently being included in meta-analyses without special consideration or if they are being excluded because of heterogeneity between study designs. Clarity and guidance will need to be brought to this issue so that

such trials can be appropriately combined in meta-analyses. The identification and assessment of findings from enriched, novel, and ATDs is critical to the interpretation of the results and the determination of whether the results can be generalized or included in evidence synthesis.

## Summary

Traditionally, ATDs have been applied more in the learning phase of drug development trials; however, in more recent years, there has been heightened interest in using adaptive techniques in confirmatory trials. The potential advantages of making adaptations in the confirmatory stage include the abilities to: 1) adjust the sample size of potential underpowered studies during interim analysis; 2) ensure balanced treatment groups in terms of important covariates; 3) shorten the development time of effective therapies; and 4) more efficiently identify ineffective treatments. Although there are potential advantages to adaptive designs, there are also potentially serious risks to these approaches — for example, the risk of inflating type I error or introducing sources of bias — that make the results of the study challenging to interpret. Further, with minimal or no time between exploratory and confirmatory phases in a combined phase II/III design, there is the risk that insufficient time will be taken to adequately reflect on the study's findings in the exploratory phase and design the confirmatory trial appropriately. As described in this report, specific procedures have been accepted to help mitigate against an increased type I error rate and bias; however, there is a general consensus that to maintain trial integrity, adaptations should be implemented by design and planned and specified a priori. Further, when possible, interim analyses should be blinded to reduce the risk of operational bias.

Some adaptive designs are considered to be "not well understood;" therefore, robust methods to control the type I error and reduce the risk of operational and other types of biases are still being developed. To date, only FDA has released some guidance on the use of ATDs in regulatory approval;[4,5] however, the recommendations are not detailed, and the documents highlight many areas in which there is insufficient knowledge of the methodologies. The European Medicines Agency has published a reflection paper on confirmatory trials planned with adaptive designs; however, this document focuses on application and concerns and offers little in the way of guidance.[7] No other regulatory agency has produced any official position statement or recommendations, which is a reflection on the novelty of this area of research.

# References


1. Guidelines for the economic evaluation of health technologies: Canada [Internet]. 4th ed. Ottawa: CADTH; 2017 Mar. [cited 2017 Nov 28]. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition
2. Marra C, Lynd L, Singer J. A review of novel trial designs [internal working paper]. Ottawa: CADTH; 2015.
3. Chow SC, Chang M. Adaptive design methods in clinical trials - a review. Orphanet J Rare Dis. 2008;3:11.
4. U.S. Food and Drug Administration. Guidance for industry: adaptive design clinical trials for drug and biologics [draft guidance] [Internet]. Rockville (MD): FDA; 2010. [cited 2018 Jan 3]. Available from: https://www.fda.gov/downloads/drugs/guidances/ucm201790.pdf
5. U.S. Food and Drug Administration. Guidance for industry: enrichment strategies for clinical trials to support approval of human drugs and biological products [draft guidance] [Internet]. Rockville (MD): FDA; 2012. [cited 2018 Jan 3]. Available from: https://www.fda.gov/downloads/drugs/guidances/ucm332181.pdf
6. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J, et al. Adaptive designs in clinical drug development--an executive summary of the PhRMA Working Group. J Biopharm Stat. 2006 May;16(3):275-83.
7. Committee for Medicinal Products for Human Use (CHMP). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design [Internet]. London: European Medicines Agency; 2007. [cited 2017 Nov 28]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf
8. Chow SC, Chang M, Pong A. Statistical consideration of adaptive methods in clinical development. J Biopharm Stat. 2005;15(4):575-91.
9. Chow SC. Adaptive clinical trial design. Annu Rev Med. 2014;65:405-15.
10. Kairalla JA, Coffey CS, Thomann MA, Muller KE. Adaptive trial designs: a review of barriers and opportunities. Trials. 2012;13:145.
11. Maca J, Bhattacharya S, Dragalin V, Gallo P, Krams M. Adaptive seamless phase II/III designs - background, operational aspects, and examples. Drug Inf J. 2006 Jan;40(4):463-73.
12. Coffey CS, Kairalla JA. Adaptive clinical trials: progress and challenges. Drugs R D. 2008;9(4):229-42.
13. Lai TL, Lavori PW, Tsang KW. Adaptive design of confirmatory trials: advances and challenges. Contemp Clin Trials. 2015 Nov;45(Pt A):93-102.
14. Chow SC, Chang M. Adaptive design methods in clinical trials. 2nd ed. Boca Raton (FL): Chapman and Hall/CRC; 2011. (Chapman & Hall/CRC biostatistics series).
15. Armitage P. Sequential medical trials. 2nd ed. Oxford: Blackwell; 1975.
16. Suresh K. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. J Hum Reprod Sci. 2011 Jan;4(1):8-11.
17. Cui L, Hung HM, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. J Biopharm Stat. 2002 Aug;12(3):347-58.
18. Colavincenzo J. Doctoring your clinical trial with adaptive randomization: SAS® macros to perform adaptive randomization. In: Proceedings of the SAS® Global Forum 2013 conference [Internet]. Cary (NC): SAS Institute Inc.; 2013 [cited 2017 Nov 28]. (Paper 181-2013). Available from: http://support.sas.com/resources/papers/proceedings13/181-2013.pdf
19. Hu F, Hu Y, Ma Z, Rosenberger WF. Adaptive randomization for balancing over covariates. WIREs Comput Stat. 2014 Jul 1;6(4):288-303.
20. Brown S, Thorpe H, Hawkins K, Brown J. Minimization--reducing predictability for multi-centre trials whilst retaining balance within centre. Stat Med. 2005 Dec 30;24(24):3715-27.
21. Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. J Athl Train. 2008 Apr;43(2):215-21.
22. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. a review. Control Clin Trials. 2002 Dec;23(6):662-74.
23. Colavincenzo J. Adaptive randomization designs [Internet]. San Luis Obispo (CA): California Polytechnic State University; 2012. [cited 2017 Nov 28]. Available from: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1024&context=statsp Senior project presented to: The Faculty of the Statistics and Mathematics Departments.
24. Wei LJ. The adaptive biased coin design for sequential experiments. Ann Stat. 1978;6(1):92-100.
25. Wei LJ, Lachin JM. Properties of the urn randomization in clinical trials. Control Clin Trials. 1988 Dec;9(4):345-64.
26. Hey SP, Kimmelman J. Are outcome-adaptive allocation trials ethical? Clin Trials. 2015 Apr;12(2):102-6.
27. Zang Y, Lee JJ. Adaptive clinical trial designs in oncology. Chin Clin Oncol. 2014 Dec;3(4):49.
28. Ellenberg SS, Fleming RR, DeMets DL. Data monitoring committees in clinical trials: a practical perspective. Chichester (GB): John Wiley & Sons, Ltd.; 2002.
29. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med. 1994 Jul 15;13(13-14):1341-52; discussion 1353-6.

30. Coffey CS, Levin B, Clark C, Timmerman C, Wittes J, Gilbert P, et al. Overview, hurdles, and future work in adaptive designs: perspectives from a National Institutes of Health-funded workshop. Clin Trials. 2012 Dec;9(6):671-80.

31. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391.

32. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol. 2006;6:54.

33. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. J Clin Epidemiol. 2011 Sep;64(9):936-48.

34. Campbell MJ, Walters SJ. How to design, analyse and report cluster randomised trials in medicine and health related research. Chichester (GB): John Wiley & Sons, Ltd.; 2014.

35. Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med. 2015 Jan 30;34(2):181-96.

36. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? BMJ. 2009;339:b4006.

37. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. BMJ. 2003 Oct 4;327(7418):785-9.

38. Bion J, Richardson A, Hibbert P, Beer J, Abrusci T, McCutcheon M, et al. 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. BMJ Qual Saf. 2013 Feb;22(2):110-23.

39. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials [Internet]. 2015 [cited 2017 Nov 28];16:352. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4538756

40. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med. 2011 Jun 30;364(26):2507-16.

41. Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. J Clin Oncol. 2010 Nov 1;28(31):4697-705.

42. Johnston S, Pippen J Jr, Pivot X, Lichinitser M, Sadeghi S, Dieras V, et al. Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor-positive metastatic breast cancer. J Clin Oncol. 2009 Nov 20;27(33):5538-46.

43. Freidlin B, Sun Z, Gray R, Korn EL. Phase III clinical trials that integrate treatment and biomarker evaluation. J Clin Oncol. 2013 Sep 1;31(25):3158-61.

44. Freidlin B, Korn EL. Biomarker enrichment strategies: matching trial design to biomarker credentials. Nat Rev Clin Oncol. 2014 Feb;11(2):81-90.

45. Furlan A, Chaparro LE, Irvin E, Mailis-Gagnon A. A comparison between enriched and nonenriched enrollment randomized withdrawal trials of opioids for chronic noncancer pain. Pain Res Manag [Internet]. 2011 Sep [cited 2017 Nov 28];16(5):337-51. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3206784