

CORRESPONDENCE

Open Access



Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs

Thomas Burnett^{1*} , Pavel Mozgunov¹, Philip Pallmann², Sofia S. Villar³, Graham M. Wheeler⁴ and Thomas Jaki^{1,3}

Abstract

Adaptive designs for clinical trials permit alterations to a study in response to accumulating data in order to make trials more flexible, ethical, and efficient. These benefits are achieved while preserving the integrity and validity of the trial, through the pre-specification and proper adjustment for the possible alterations during the course of the trial. Despite much research in the statistical literature highlighting the potential advantages of adaptive designs over traditional fixed designs, the uptake of such methods in clinical research has been slow. One major reason for this is that different adaptations to trial designs, as well as their advantages and limitations, remain unfamiliar to large parts of the clinical community. The aim of this paper is to clarify where adaptive designs can be used to address specific questions of scientific interest; we introduce the main features of adaptive designs and commonly used terminology, highlighting their utility and pitfalls, and illustrate their use through case studies of adaptive trials ranging from early-phase dose escalation to confirmatory phase III studies.

Keywords: Novel designs, Innovative trials, Efficient methods, Enrichment designs, Multi-arm multi-stage platform trials

What are adaptive designs?

In a traditional clinical trial, the design is fixed in advance, the study is carried out, and the data analysed after completion [1]. In contrast, adaptive designs pre-plan possible modifications on the basis of the data accumulating over the course of the trial as part of the trial protocol [2]. We consider designs that allow for modifications of the trial such as the sample size, the number of treatments, or the allocation ratio to different arms. We do not consider options such as stopping early due to failure to meet operational criteria or excessive safety events, although adaptive designs for some of these do also exist [3]. Adaptive design methodology has been around for more

than 25 years [4], with some methods such as group sequential designs being even older [5].

It is crucial that the adaptive nature of a design does not undermine the trial's integrity and validity [6]. By integrity of a trial, we mean that the data have not been used in such a way as to substantially alter the result, while the validity of the results requires that the study answers the original research questions appropriately. Adaptive designs require procedures to ensure that data is collected, analysed, and stored in an appropriate manner at every stage of the trial, with specialised statistical methodology for inference. The involved logistical and statistical nature of adaptive designs should also be reflected in their reporting [7].

Flexibility of a design is not a virtue in itself but rather a gateway to more efficient and ethical trials where futile treatments may be dropped sooner, more patients

*Correspondence: t.burnett1@lancaster.ac.uk

¹Department of Mathematics and Statistics, Lancaster University, Fylde College, Lancaster, LA1 4YF, UK

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

may receive a superior treatment, fewer patients may be required overall, treatment effects may be estimated with greater precision, a definitive conclusion may be reached earlier, etc. Adaptive designs can aid in these aspects across all phases of clinical development [2].

Despite the many clear benefits, many modern adaptive designs are still far from established as typical practice [8–10]. Many reasons for this have been identified [9–12], the main of which include the following: lack of expertise and experience (in the application of adaptive designs among clinicians, trialists, and statisticians), lack of design and analysis software, time required for planning and analysis, inadequate funding structure to account for design uncertainty, and the fact that chief investigators may prefer more familiar methods. We believe that the main reason investigators are not inclined to adopt adaptive designs is a lack of clarity about what these are and what they can (and cannot) accomplish and how they may be implemented. Ambiguous terminology and vague definitions add to this confusion [13], and hence, we provide a glossary of common types of adaptive design in Table 1. Other work providing reviews of recent uses of adaptive designs may provide insight into designs not covered in detail here [4, 14].

To demonstrate how and when adaptive designs can be useful, we focus on four key questions of scientific interest when developing and testing novel treatments: ‘What is a safe dose?’ ‘Which is the best treatment among multiple options?’ ‘Which patients will benefit?’ ‘Does the treatment work?’ For each of these questions, we briefly review several important adaptive designs, outlining their advantages and disadvantages. We illustrate their application through real-world examples.

When to use an adaptive design

What is a safe dose?

Phase I trials of new drugs are conducted to assess the safety of a treatment, the aim being to establish the safety profile across a range of available doses, in order to select a dose for further testing. In many therapeutic areas, the goal is to identify the maximum tolerated dose (MTD), that is the highest dose that controls the risk of unacceptable side effects [15] and hence is deemed safe. In practice, one seeks to identify the dose at which the probability of a dose-limiting toxicity (DLT) is equal to some pre-specified target level, usually around 20–33%. This is done by treating consenting patients sequentially at increasing doses until too high a proportion of unacceptable side effects are observed.

3+3 design

The most commonly used method for conducting dose-escalation studies in oncology is the 3+3 design [16, 17]. It is a simple, rule-based approach under which patients are

dosed in cohorts of three. Based on the number of DLTs observed in the current cohort of patients, recommendations are made to dose the next three patients at either the next escalating dose or the current dose. Upon observing a pre-specified number of toxic outcomes at a dose level (say DLTs in more than 2 in 6 patients), the trial is terminated and the dose level below is considered to be the MTD. The 3+3 design is a special case of the more general A+B design [18]; when a new dose is introduced, a cohort of A patients are dosed, and if further observations are required on the same dose, a cohort of B further patients are then dosed.

Example Park et al. [19] performed a phase I dose-escalating study of docetaxel in combination with 5-day continuous infusion of 5-fluorouracil (5-FU) in patients with advanced gastric cancer. The study used a 3+3 design to find the MTD from four dose levels of 5-FU. The treatment consisted of docetaxel 75 mg/m² on day 1 in a 1-hour infusion followed by 5-FU in continuous infusion from day 1 to day 5, according to the escalating dose levels. The starting dose of 5-FU was 250 mg/m²/day for 5 days. In the absence of any DLTs (defined as febrile neutropenia and/or grade 3/4 toxicity of any other kind apart from alopecia), dose escalation in additional cohorts continued, increasing the dose by 250 mg/m²/day for each increment.

The first DLT was observed at dose level 2 (5-FU 500mg/m²/day for 5 days). Three additional patients were enrolled at this dose level, none of whom experienced any DLT. Thus, dose escalation proceeded to dose level 3 (5-FU 750 mg/m²/day for 5 days) where a further 2 patients experienced DLTs and so dose escalation was stopped. Dose level 2 was therefore the recommended regimen with docetaxel 75 mg/m² on day 1 and 5-FU 500 mg/m²/day in a 5-day continuous infusion.

Advantages The key advantage of the 3+3 design is that it does not require any time to design. In addition, this method is well-known to clinicians, often leading to its use being well motivated within the trial team. Web applications [20] are available to understand the performance of such designs.

Disadvantages The major disadvantages of the 3+3 design will become clear as we draw comparisons to the methods that follow. In particular, we note that the MTD is not explicitly defined; this means that the most likely dose to be chosen as the MTD can have a probability of DLT far from the assumed target and can be highly variable [21–23].

Rule-based dose-escalation methods such as 3+3 designs are seriously flawed, which runs afoul of the part of our definition of an adaptive design that demands

Table 1 Glossary of adaptive designs and descriptions of their typical applications

Method	a.k.a	Phase of development	Definition	Targeted benefits
What is a safe dose?				
Continual re-assessment method	CRM	I	Dose-escalation design for finding the maximum tolerated dose (MTD)	More accurate and precise estimation of the MTD than with 3+3 designs, more patients treated at or close to the MTD
Escalation with overdose control	EWOC	I	Dose-escalation design to find MTD using an allocation criterion to avoid overdosing	More accurate and precise estimation of the MTD than with 3+3 designs, avoiding undesirable overdosing of patients
Which is the best treatment of multiple options?				
Adaptive treatment switching		II/III	Allow trial participants to switch from allocated treatment to an alternative	More trial participants receive preferred treatment
Drop the loser	DTL	II/III	Drop inferior treatment arms (control group typically retained)	Fewer trial participants assigned to less effective treatments
Multi-arm multi-stage trial	MAMS	II/III	Compare multiple treatments to a common control, allow for early stopping for efficacy or futility	Common control requires fewer patients than conducting separate trials, early stopping for efficacy or futility
MCP-Mod		II	Combination of multiple comparisons and modelling approaches to establish dose-response model	Efficient use of available data vs pairwise comparisons
Response-adaptive randomisation	Adaptive allocation, RAR	II/III	Shift allocation ratio towards more promising treatment(s)	More trial participants receive effective treatment
Which patients will benefit?				
Basket trials			Examine a single experimental treatment in multiple sub-types of a biomarker	Identify and target biomarker sub-types that benefit from the treatment
Biomarker adaptive	Adaptive signature	II/III	Identify and utilise biomarker information to modify trial in progress to target population	Target the correct patient population
Covariate-adjusted response adaptive	CARA	II/III	Shift allocation ratio towards promising treatment(s) using covariate information	More trial participants receive effective treatment
Population enrichment	Adaptive enrichment	II/III	Allow for selection of target population during the trial based on pre-defined patient populations	Target the correct patient population
Umbrella trials			Multiple biomarkers each paired with specific treatments	Target the appropriate to treatment within each patient group
Does the treatment work?				
Group sequential design		II/III	Early stopping for futility or efficacy	Reduction in the expected sample size, typically allowing for faster trials requiring fewer patients (for a small increase in the possible maximum sample size)

Table 1 Glossary of adaptive designs and descriptions of their typical applications (*continued*)

Method	a.k.a	Phase of development	Definition	Targeted benefits
Sample size re-assessment	Sample size re-estimation/re-calculation	II/III	Mid-course adjustment of the sample size, in either a blinded or unblinded fashion	Raise the probability of a successful trial
Broader topics in adaptive designs				
Bayesian adaptive		I/II/III	Bayesian methodology may be incorporated into many other designs in the analysis and/or the interim decision-making	Lower sample size due to utilisation of prior information
Seamless design	Portfolio decision-making	I/II/III	Merge trials from different phases of development, e.g. phase I/II or phase II/III, can be inferentially and/or operationally seamless	(Inferential) More efficient use of data from each phase of clinical development/(operational) faster clinical development process and moving between stages

integrity and validity. Thus, this method being well-known to clinicians, possibly allowing them to avoid collaboration with a statistician, can also present a serious problem.

Continual re-assessment method

The continual re-assessment method (CRM) [24, 25] models the relationship between dose and the risk of a patient experiencing a DLT, using an iterative process to make use of all available trial data when choosing the dose for the next patient cohort. Based on all available data from the trial, the relationship between dose and toxicity is modelled to inform the choice of dose for the next cohort. The dose for the next patient or cohort is chosen as either that with an estimated probability of DLT closest to the target toxicity level, or the highest available dose below the target level. This process is iterated for each new cohort of patients, ensuring that at all times all available data are used. The application of the CRM process is highly flexible, allowing the investigators to adjust the design to suit the particular trial and trialist (making use of all trial data wherever it is introduced, as is seen in the example to follow). Both the cohort size and the sample size of a CRM trial are determined before the trial begins; sample sizes are often planned with practical constraints in mind rather than statistical properties while simulation may be used to understand statistical operating characteristics [26].

Example Paoletti et al. [27] provide a tutorial of the practical considerations for designing CRM trials; they describe the design, conduct, and analysis of a multicentre phase I trial to find the MTD (defined as the dose with probability of a DLT closest to 20%) of rViscumin in patients with solid tumours. A DLT was defined as any haematological grade 4 or non-haematological grade 3 or

grade 4 adverse event as defined by the National Cancer Institutes Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 2, with the exclusion of nausea, vomiting, or rapidly controllable fever. The starting dose of the trial was 10 ng/kg, with fixed dose levels for further exploration of 20, 40, 100, 200, 400, and 800 ng/kg; if no adverse events of grade 2 or higher were observed after escalation to 800 ng/kg, additional doses would be added in increments of 800 ng/kg (i.e. 1600, 2400 ng/kg).

The trial used a two-stage CRM design [24] allowing the low doses to be rapidly moved through while utilising the model-based approach in the selection of the MTD. During the first stage, one patient was assigned to the starting dose of 10 ng/kg, and if adverse events were absent or grade 1, a new patient was given the next highest dose; if a non-DLT adverse event of grade 2 or higher was observed, a further two patients would be given the same dose. Escalation continued in this manner until the first DLT was observed, at which point the model-based design took over. A one-parameter model [25] was fitted to the data, and the dose with an estimated probability of DLT closest to 20% was recommended for the next patient, subject to the constraint that no untested dose level is skipped. The trial was stopped when the probability of the next five patients being given the same dose was at least 90% (i.e. the trial would be unlikely to gain further information that would affect dose allocation).

The first DLT was observed in the 11th patient who was given 4000 ng/kg, at which point the CRM part of the design took over. In total, 37 patients were recruited to the trial before it was terminated under the aforementioned rule, and the MTD declared as 5600 ng/kg, with an estimated DLT probability of 16%; the estimated probability of a DLT at the next highest dose (6400 ng/kg) was 31%. It is worth noting that during the ongoing trial,

the first DLT was recoded to a non-DLT; this change is easily incorporated in a CRM design by simply re-estimating the DLT risks at each dose using the updated data [26]. This recoding of the first DLT had an impact on the overall trial outcome as without this a lower MTD would have been selected [27]. This illustrates one of the benefits of a model-based approach; deviations from the planned course of the trial are handled without compromising the validity of the design.

The authors describe how the statistical work of the trial helped to inform study clinicians and the Trial Steering Committee, with whom any final decisions rest. For example, a decision was made to dose another patient at 3200 ng/kg rather than escalate to 4000 ng/kg as per the design in order to gather more PK data at this level. Furthermore, 10 extremely tolerable (but presumably ineffective) dose levels were cleared quickly and with far fewer patients than the 3+3 design would require.

Advantages Conceptually, the CRM is a far wiser approach (and more efficient/economic) than the 3+3 design because it uses all available trial data to make decisions, rather than solely the data from the last cohort [28, 29]; the CRM also targets a pre-specified toxicity level. Numerous comparative simulation studies have shown the CRM to supersede the 3+3 design by dosing more patients in trial at or near the correct MTD and also by selecting the correct dose as the MTD more often [30–34], which in turn can result in higher probability of success in subsequent phase II and phase III clinical trials [35].

Furthermore, the CRM can easily be adapted to include more informative endpoints as follows: multiple graded toxicities to incorporate severity of side effects [36–38], combinations of safety and efficacy [39–44], time-to-event outcomes to distinguish between toxicity events occurring sooner or later [45, 46], or even developed to escalate multiple treatments at once [47–54]. Regulatory authorities are also recognising that novel adaptive designs using statistical models are of great importance, and actively encourage sensible usage of them in phase I trials [55, 56].

Disadvantages The main disadvantage of the CRM design is the time and effort required at the design stage to assess how the trial is expected to perform. This requires close collaboration between the clinical team and a suitably trained statistician who is able to guide this optimisation process, although this opportunity to consider the design more carefully can only be a good thing. The clinical team may still see the CRM as a ‘black box’; to resolve this concern, Dose Transition Pathways provide a tool for visualisation of the CRM escalation/de-escalation decisions [57]. Several computer programs are available

(see MD Anderson Cancer Center software library [58], Vanderbilt University, and packages within R [59, 60]) for conducting simulation studies, some of which can offer comparisons to other popular dose-escalation designs [60] and tutorial papers are available offering further guidance [26]. Web-based solutions are on offer for both the CRM and conceptual equivalents [61–63].

Escalation with overdose control

The escalation with overdose control (EWOC) approach [64] is similar to the CRM in that all available data are used to make dose-escalation decisions with a target toxicity level used to choose which dose level the next patient or cohort should receive. However, the EWOC approach assigns the next patient using a skewed allocation criterion to account for the fact that the overdosing of patients is much more undesirable compared to the underdosing. This results in a more conservative patient allocation approach, with fewer patients being exposed to possible overdosing compared to the CRM, while still benefiting from the model attempting to allocate the patients near the MTD [65, 66]. Additionally, the same statistical model is re-expressed in a way that allows focus on the clinically relevant parameters, the MTD, and the probability of a DLT at the lowest dose. This means that prior information about the treatment being investigated can easily be incorporated and one can visualise how the distribution of the MTD changes over the course of the trial.

Example Nishio et al. [67] conducted a dose-escalation study of ceritinib in patients with advanced anaplastic lymphoma kinase-rearranged, non-small-cell lung cancer, or other tumours. A Bayesian EWOC approach was used to allocate the dose for the next patient. This allocates the next patient to the largest dose with an estimated probability of less than 25% that the risk of a DLT exceeds 33%. In total, 19 patients were recruited to the trial: three patients received doses of 300 mg, six patients received doses of 450 mg, four patients received doses of 600 mg, and six patients received doses of 750 mg. Two patients experienced DLTs, one at 600 mg and the other at 750 mg. The MTD was chosen as 750 mg, the largest dose at which the estimated probability of the risk of a DLT exceeding 33% was less than the target 25% (the probability was 7.3% for the chosen dose). Although the aim of the trial was not to evaluate efficacy of ceritinib in this population, 10 patients achieved partial responses to their cancers.

Advantages The EWOC approach offers a more cautious dose-escalation design that reduces the chance of patients being treated at excessively toxic doses [64]. Similar to the CRM, the EWOC approach has also been adapted to be used in trials with more complex outcomes, such as time-to-event data [68], and for combinations of treatments

[69, 70]. Furthermore, the escalation control threshold can be altered depending on the trial context and may change during the conduct of the trial [65, 66]; this offers a conservative dose-escalation schema at the start of the trial when there is little data available, but as more data are accrued, dose-escalation gradually becomes less conservative, and the MTD can be targeted more quickly than with the standard EWOC approach [65, 66, 71].

Disadvantages A slower dose-escalation approach may increase the number of patients treated at sub-therapeutic doses. Similar to implementation of the CRM, care is required when designing trials using the EWOC approach. For example, choice of the MTD estimator needs to be considered; several trials use the same criterion as that by Nishio et al. [67], i.e. the MTD is the dose that would be given to a new patient had they entered a trial. The implications of each choice need to be considered well in advance [72]. Furthermore, if the investigators plan to relax the escalation control mechanism, as has been done in practice before [66, 73, 74], the implications of this decision need to be considered. The EWOC approach may recommend to escalate the dose even when the most recently evaluated patient experienced a DLT [75].

Summary

Despite the 3+3 designs frequent use in phase I clinical trials over the last 30 years [8, 76–78], there is overwhelming consensus among statisticians and methodologists that it is sub-optimal, and more efficient designs for identifying the MTD should be used [35, 79, 80]. Many alternative designs propose the use of statistical models, such as the two alternatives we have presented here; both of which have superior operating characteristics over the 3+3 design.

The model-based approaches above serve as the main framework for other proposed approaches designed for trials with novel drug combinations, endpoints that use time-to-event data and/or efficacy outcomes, or information about the severity of observed toxicities. These designs have found their way into clinical practice in recent years, primarily in oncology for cytotoxic treatments. However, they can be used for novel molecularly targeted anti-cancer therapies [81], and in other disease areas altogether: O’Quigley et al. [82] proposed CRM-type designs for anti-retroviral drugs to treat human immunodeficiency virus (HIV), Lu et al. [83] conducted a dose-escalation study of quercetin in patients with hepatitis C; Whitehead et al. [84] proposed a model-based design for trials in healthy volunteers, and Lyden et al. [85] used a CRM design in the RHAPSODY trial in stroke patients.

Combining the advantages of both model-based and rule-based designs can also be desirable, with proposals such as the Bayesian optimal interval (BOIN) design [86, 87].

There is no ‘one size fits all’ approach for conducting adaptive dose-escalation studies, but there is overwhelming evidence that model-based designs are far better than rule-based designs, such as the 3+3. Model-based designs are on the whole more efficient in their use of data, less likely to dose patients at sub-therapeutic doses, more likely to recommend the correct MTD at the end of the trial, and provide an MTD estimate that directly relates to a specified target level of toxicity. We have discussed two approaches here for brevity, though many other alternatives have also been proposed, including designs based on optimal design theory [88–91] and model-free designs [86, 92–94], without the shortcomings of common rule-based designs like 3+3. The increasing usage of model-based designs in practice, as well as their acknowledgement in regulatory guidance and the provision of guideline documents [95], formal courses and computer software is indicative of the changing tide of clinical practice for phase I trials. Although the fact remains that such designs are more complex making implementation more challenging, for a trained statistician unfamiliar with the design planning, such a trial would require a significant investment of time. Such issues in implementing novel statistical methods are well recognised [96].

Which is the best treatment among multiple options?

After establishing the safety of a treatment, we next examine its efficacy. In this section, we consider randomised clinical trials that aim to select the best treatment among multiple experimental treatment arms (these can be different treatments, doses of the same treatment, or combinations of the two). The methods we explore are typically considered for use in phase II of the development process, where we wish to select a treatment or dose for further study. We explore methods that seek to remove less beneficial treatments from the trial quickly, giving patients a better chance of receiving an efficacious treatment. In trials where the different arms correspond to a series of exposure levels (such as doses of a drug, duration or intensity of radiotherapy, or number of therapy sessions), model-based approaches examine the dose-response relationship to provide a deeper understanding of this in an efficient way.

Multi-arm multi-stage

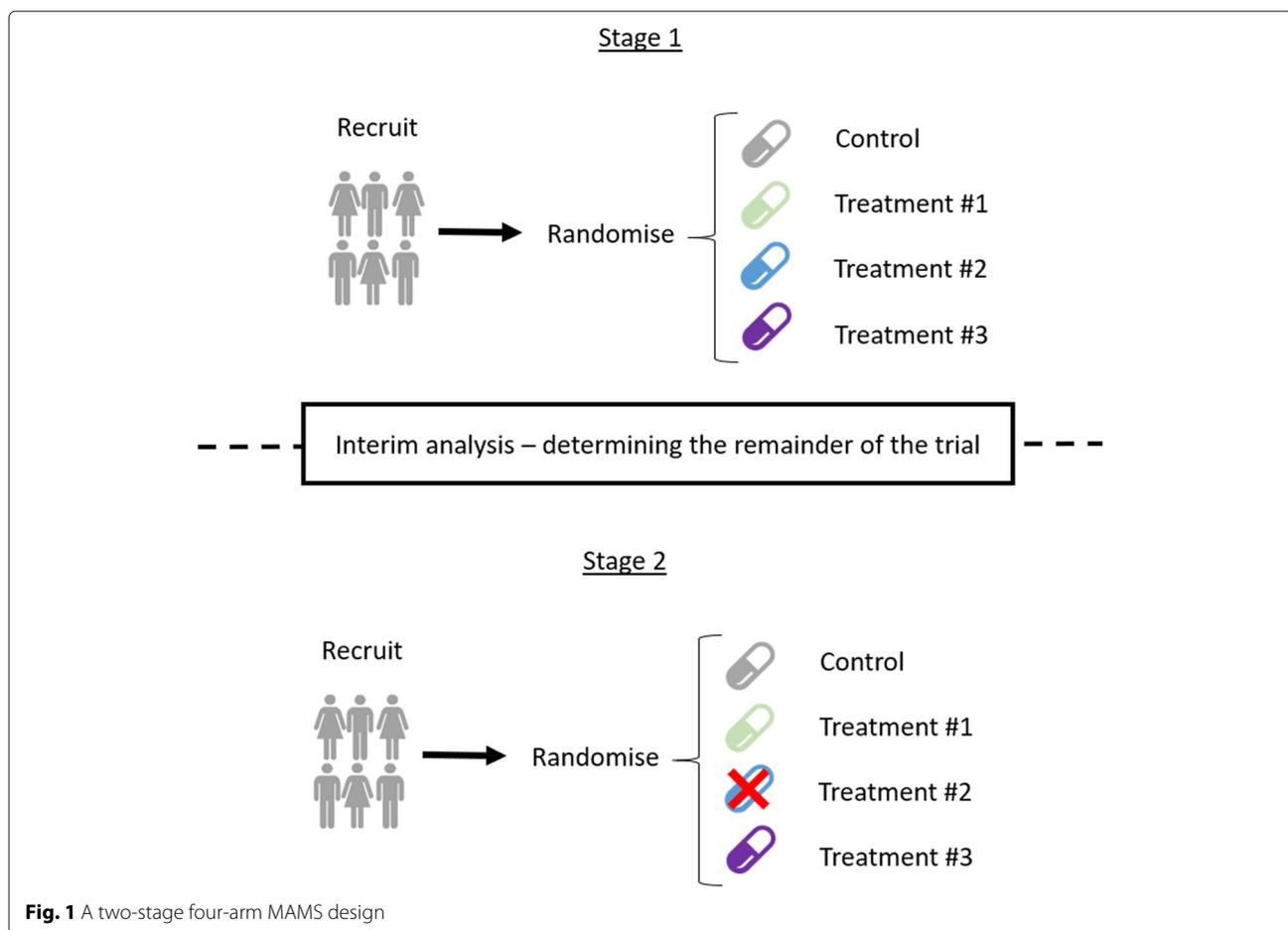
Multi-arm multi-stage (MAMS) [97] trials allow the simultaneous comparison of multiple experimental treatment arms with a single common control. They are conducted over multiple stages: allowing for the early stopping of recruitment for either efficacy or futility. For example, if an experimental treatment is found to be

performing poorly, it may be dropped for futility at a pre-planned interim analysis (if all experimental arms are dropped, the trial is stopped for futility); alternatively, the trial may end early when a treatment is shown to be sufficiently efficacious. We cover group sequential designs in more detail in the ‘Group sequential designs’ section, but MAMS designs apply similar methodology while testing multiple experimental treatments simultaneously. A simple schematic of how a two-stage four-arm trial using MAMS design can progress is given in Fig. 1.

MAMS trials are designed using a pre-planned set of adaptation rules to find the best treatment to carry forward for further study [98] or carry forward all promising treatments [99]. Alternatively, more flexible testing methods [100–102] allow methodological freedom as to how adaptation decisions are made. Open source software is available to assist in the design and analysis of MAMS trials in the form of the ‘MAMS’ package for R [103]. Alternatively, in STATA, there are several modules available such as ‘nStage’ [104], ‘nStagebin’ [105], and ‘DESMA’ [106].

Example The TAILoR trial [107] was a phase II, multi-centre, randomised, open-labelled, dose-ranging trial of telmisartan using a two-stage MAMS design. The trial planned recruitment of up to 336 HIV-positive individuals over a 48-week period, with a single interim analysis planned after 168 patients had completed 24 weeks on either an intervention or a control treatment. Patients were randomised with equal probability to one of four groups: no treatment (control), 20 mg telmisartan daily, 40 mg telmisartan daily, or 80 mg telmisartan daily.

At the interim analysis, there were three possible outcomes based on assessment of change in HOMA-IR index from baseline to 24 weeks: if one telmisartan dose was substantially more effective than control, the study would stop and that dose would be recommended for further study; if all telmisartan doses were less effective than control, the study would stop with no dose recommended for further study; if one or more doses were better than control but none met the first criterion, the study would continue and patients would have been randomised between these remaining dose(s) and control. If a second stage was conducted, then a final analysis would be conducted with



two possible outcomes: either the best dose is significantly more effective than the control in which case it is recommended for further study, or no dose is significantly better than control in which case no dose is recommended.

A total of 377 patients were recruited [108] (note this difference in sample size was due to higher than expected dropout). In stage 1, 48, 49, 47, and 45 patients were randomised to control and 20, 40, and 80 mg telmisartan, respectively. At the interim analysis, the 20- and 40-mg telmisartan groups performed worse than control on average and so only 80 mg telmisartan was taken forward into stage 2. At the end of stage 2, 105 patients had been recruited to control and 106 to the 80-mg arm (in total); there was no difference in HOMA-IR (estimated effect, 0.007; SE, 0.106) at 24 weeks between the telmisartan (80 mg) and control arm. If a traditional fixed sample design had been used in place of a MAMS design, all experimental arms would have been studied throughout the trial, requiring a further 100 or so patients for arms that ultimately did not demonstrate an effect of the experimental treatment.

Advantages MAMS designs are useful when there are multiple promising treatments with no strong belief that one treatment will be more beneficial. The use of a shared control group considerably reduces the number of patients that need to be recruited compared to separate RCTs testing each treatment. Other advantages are as follows: treatments that provide no benefit to patients are dropped from the trial; patients have a higher chance of receiving an experimental treatment compared to a 2-arm trial, which may improve recruitment to the study [109, 110]; administratively and logistically, effort is only required for one trial and thus can substantially speed up the development process [111].

Disadvantages MAMS trials require an outcome measure that allows a timely decision about the worth of each treatment. Consequently, the primary endpoint needs to be relatively quickly observed (in comparison to patient accrual) or an intermediate measure that is strongly associated with the primary endpoint is required for interim decision-making. MAMS designs require a larger potential maximum sample size than a corresponding multi-arm fixed design (although smaller than several separate trials). The MAMS approach has a variable sample size depending on which decisions are made during the trial, making planning more cumbersome, although the possible pathways are pre-defined; this is more variable than is typical of even other adaptive designs because decisions relate to each treatment individually.

Drop the loser

Drop the loser (DTL) designs [112, 113] are closely related to MAMS designs [114] in that they compare several experimental treatments to a common control over multiple stages. The key difference is that in a DTL design, it is pre-determined how many arms will be dropped after each stage of the trial. As the name suggests, the worst performing experimental treatments are dropped at interim analysis, leaving only one treatment to compare to control at the final analysis.

Example The ELEFANT trial [115] is a randomised controlled, multicentre, three-armed trial testing whether early elimination of triglycerides and toxic free fatty acids from the blood is beneficial in HyperTriGlyceridemia-induced Acute Pancreatitis (HTG-AP). The study uses a two-stage DTL design; in the first stage, patients with HTG-AP are randomised with equal probability into three groups: patients who undergo plasmapheresis and receive aggressive fluid resuscitation, patients who receive insulin and heparin treatment with aggressive fluid resuscitation, and patients with aggressive fluid resuscitation only (the control). At the interim analysis, the two experimental treatments will be compared and the one demonstrated to be the best will be retained for the remainder of the trial. Thus, in the second stage, patients are randomised into two groups, the control and the selected experimental treatment. At the conclusion of the trial, formal statistical comparisons may be drawn between the control and the experimental treatment selected at the interim analysis. The target sample size is 495 in order to detect a 66% relative risk reduction, using a 10% dropout rate with 80% power at 5% significance level. The study began in February 2019 and is expected to finish December 2024.

Advantages As with MAMS designs, the key advantage is the use of the shared control group which reduces the number of patients required. Of further practical benefit is the guarantee that a pre-specified number of arms will be dropped during the course of the trial, meaning that the required sample size is known before recruitment begins [114]. At the conclusion of the trial, only one experimental treatment remains to be compared against the control making for a clear interpretation of results.

Disadvantages The choice to only continue the most efficacious treatments may cause concern; consider for example an interim analysis where a dropped treatment has demonstrated almost equivalent performance to a treatment that continues in the trial, it is possible that a suitable treatment has been dropped by chance. In addition, there is a similar operational complexity to MAMS designs as it is unknown which treatments will be carried through the trial.

Response-adaptive randomisation

At the beginning of the trial, the comparable performance of the experimental treatment arms may be unknown; hence, equal randomisation is sensible under a clinical equipoise principle. However, as data accumulates, it becomes challenging from an ethical standpoint to randomise patients to a treatment arm that data suggest may be inferior. To resolve this, response-adaptive randomisation (RAR) aligns the randomisation probabilities with the observed efficacy of the different arms (Fig. 2).

RAR dates back to 1933 [116], and since then, several methods to align randomisation probabilities and observed evidence of efficacy have been proposed [117, 118]. Thompson [116] proposed to randomise patients to arms with a probability that is proportional to

the probability of an arm being the best arm. Regardless of how these probabilities are defined and applied, they can be also used to define further adaptations to the trial depending on the values they assume. For example, if the allocation probability goes below or rises above a certain value, arms can be dropped for futility or selected similar to a MAMS study [119, 120]. Free software is available from the MD Anderson website [58]. The R package ‘bandit’ offers an alternative to the implementation of such designs [121].

Example A prospective, randomised study reported by Giles et al. [122] was conducted in patients aged 50 years or older with untreated, adverse karyotype, acute myeloid leukaemia to assess three troxacitabine-based regimens:

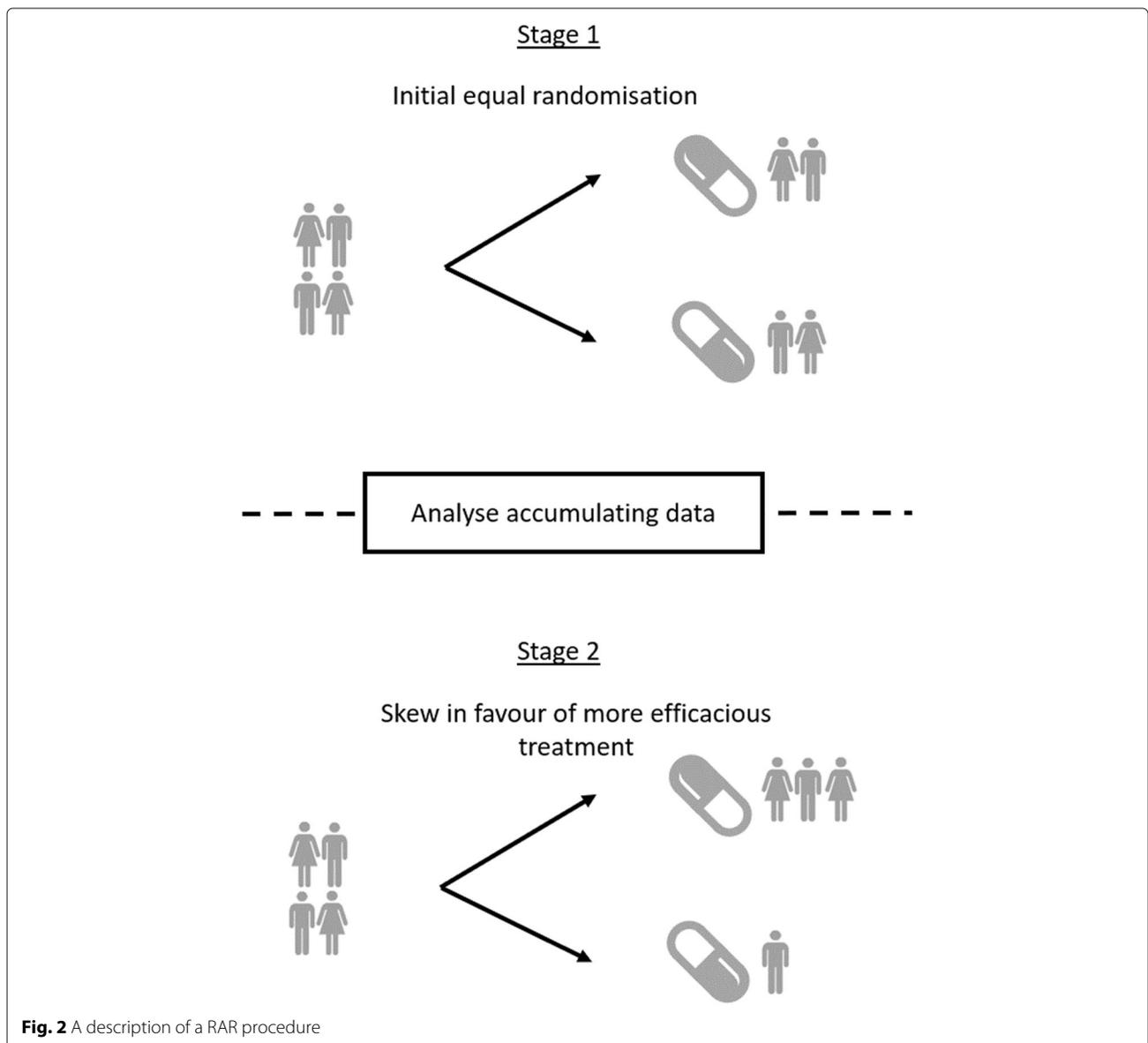


Fig. 2 A description of a RAR procedure

idarubicin and cytarabine (the control arm), troxacitabine and cytarabine, and troxacitabine and idarubicin. The trial used a Bayesian RAR design along the lines proposed by Thompson [116]. Thirty-four patients were recruited and randomised to one of the three arms. Initially, there was an equal chance of randomisation to each arm. The randomisation probabilities were updated after every patient such that treatment arms with a higher success rate, defined as the proportion of patients having complete remission within 49 days of starting treatment, would receive a greater proportion of patients. The design would drop arms if their assignment probabilities became too low or promote them to phase III if their assignment probability was high enough. The probability of a patient being randomised to the control arm was fixed until the first experimental arm was dropped. This occurred when the randomisation probability for the dropped experimental arm was 0.072 [123].

Of the thirty-four patients recruited, 18 were randomised to idarubicin and cytarabine, randomisation to troxacitabine and idarubicin stopped after five patients, and randomisation to troxacitabine and cytarabine stopped after 11 patients. Success rates were 55% (10 of 18 patients) with idarubicin and cytarabine, 27% (three of 11 patients) with troxacitabine and cytarabine, and 0% (zero of five patients) with troxacitabine and idarubicin.

Advantages RAR can increase the overall proportion of patients enrolled in the trial who benefit from the treatment they receive while controlling the statistical operating characteristics [124–126]. This mitigates potential ethical conflicts [127] that can arise during a trial when equipoise is broken by accumulating evidence and makes the trial more appealing to patients [110] possibly improving trial recruitment [128]. The main motivation for RAR designs is to ensure that more trial participants receive the best treatments; it is possible to use such methods to optimise other characteristics of the trial [129]. In a multi-armed context, RAR can shorten the development time and more efficiently identify responding patient populations [130].

Disadvantages RAR designs have been criticised for a number of reasons [131] although many of the raised concerns can be addressed. Logistics of trial conduct is a noticeable obstacle in RAR due to the constantly changing randomisation [132]; requiring more complex randomisation systems may in turn impact things such as drug supply and manufacture. When the main advantage pursued is patient benefit, this may compromise other characteristics; for example, a two-arm RAR trial will require larger sample sizes than a traditional fixed design with equal sample sizes in both arms; methods to account for such compromise have been proposed [124, 133].

Choosing an approach from the variants of RAR can be challenging; in most cases, balancing the statistical operating characteristics and randomly assigning patients in an ethical manner is required. Most RAR methods require the availability of a reliable short-term outcome (although the exact form of the data may vary [124, 134]); however, this can result in bias, requiring the use of extra correction methods for estimation purposes [135]. Another statistical concern is control of the type I and type II error rates. As discussed above, this is possible but requires intensive simulations or the use of specific theoretical results [118, 129, 130, 136]; this creates an additional burden at the design stage, requiring additional time and support.

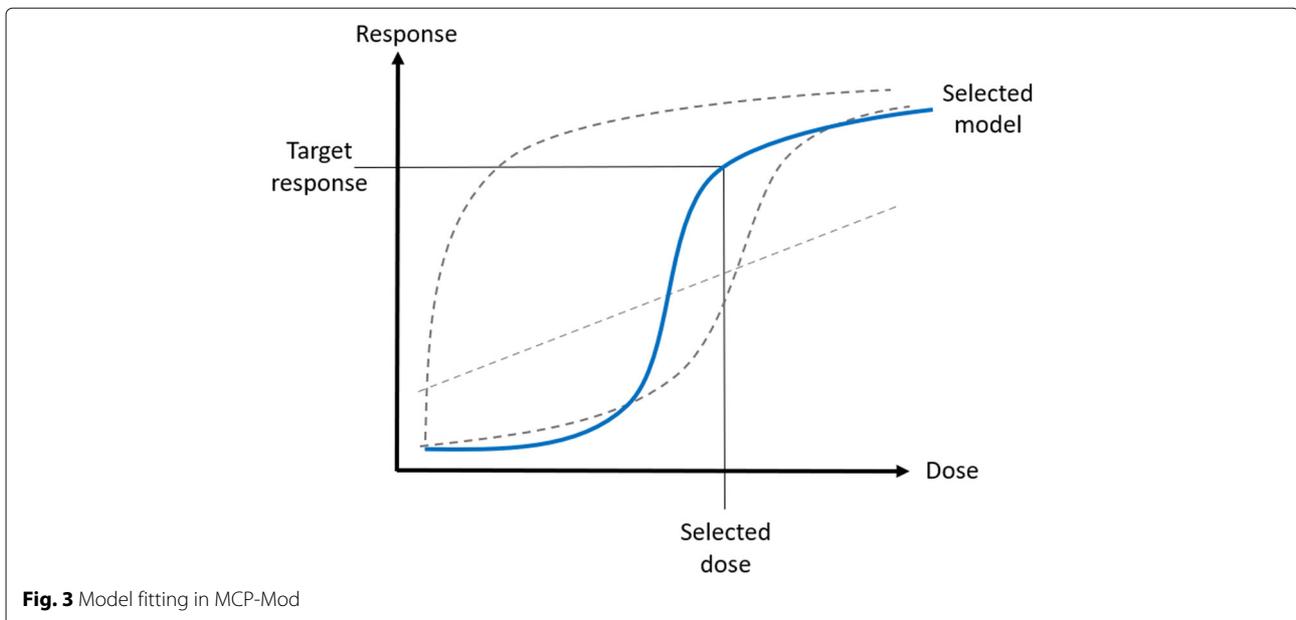
Multiple comparison procedures and modeling approaches (MCP-Mod)

In phase II dose-ranging studies, patients are typically randomised to either one of a number of doses and possibly a placebo. The target dose is often the minimum effective dose, the smallest dose giving a particular clinically relevant effect. A traditional approach to find the target dose is based on pairwise comparisons. However, this only uses the information from the corresponding doses and typically results in larger sample sizes required in the trial [137]. As an alternative, MCP-Mod [138, 139] employs a dose-response model allowing for interpolation between the doses.

MCP-Mod is a two-stage method that combines Multiple Comparison Procedures (of dose levels) and MODELing approaches. At the planning stage, the set of possible models for the relationship between dose and response are defined, such as those shown in Fig. 3. The inclusion of several models addresses the issue of some of the models being mis-specified. At the trial stage, the MCP step checks whether there is any dose-response signal. This is done through hypothesis tests for each model, adjusting for the fact that there are multiple candidate models. This controls the probability of making an incorrect claim of a dose-response signal (a type I error). If no models are found to be significant, it is concluded that the dose-response signal cannot be detected given the observed data.

With a dose-response signal established, a single model is selected, or if multiple models are selected, an average is made. The selection of models can be based either on tests performed at the MCP step or on some other measures such as information criterion [140]. The chosen model is used to select the best dose. We refer the reader to the works focusing on the step-by-step application of MCP-Mod in practice [137, 141].

Example Verrier et al. [142] described the application of MCP-Mod in a placebo-controlled parallel group study undertaken in hypercholesterolaemic patients, which evaluated the change in low-density lipoprotein



cholesterol (LDLC, mg/dL) following 12 weeks treatment as the primary endpoint. Three active doses were studied: 50, 100, and 150 mg, and nearly 30 patients per treatment group were recruited. The objective of the trial was twofold: (i) to demonstrate the dose effect of the compound and (ii) to select the dose providing at least a 50% decrease in LDLC.

The set of candidates was composed of linear, logistic (four possible pairs of parameters obtained from the guess that 50% of the maximum effect occurred at 50, 75, 100, or 125 mg, respectively, defined four models), and quadratic (corresponding to the maximum effect at 125 mg) models. The hypothesis tests selected a logistic model, and the estimated target dose was 76.7 mg. Alternatively, an information criterion approach selected the quadratic model and the estimated target dose was 78.2 mg. To check the robustness of the results, model averaging was used and resulted in nearly the same estimated target dose. This analysis informed the selection of the dose for phase III trials for which the dose of 75 mg was chosen. This choice of dose was not one of those three active doses directly studied but could be selected due to using a model-based approach.

Advantages MCP-Mod allows for a more efficient use of data. There are many practical recommendations available [137], and it has been successfully applied in a number of trials [143]. The European Medicines Agency issued a qualification opinion of MCP-Mod [144] concluding that MCP-Mod uses available data better than the traditional pairwise comparisons, and the FDA also designated the method as fit for purpose [145]. There is software that implements the methodology, e.g. an

R-package, *DoseFinding* [146] and PROC MCPMOD in SAS.

Disadvantages The method can be sensitive to the model assumptions [147], which can result in significantly lower power if the dose-response relationship is not well approximated by one of the pre-specified candidate models. The number of doses to be included should inform the choice of the candidate models. When the treatment regimens consist of various drugs and schedules (and doses within each), such disadvantages are amplified and MCP-Mod should be approached with much care.

Summary

The methods presented in this section are suitable for selecting a treatment or dose for further study and can allow for formal testing in a confirmatory setting. With RAR, we see the goal of focusing on the more effective treatments was thought of as an important topic almost 90 years ago; however, it is only in the last 30 years or so that this topic has gained traction as a more active area of methodological research. The knowledge on these more modern methods will need to be broadly shared before we start to see their use more widely in clinical practice [148].

Each of the methods discussed uses adaptation to make efficient comparisons of several treatments or doses. There is a common advantage of reducing the (expected) number of patients required to achieve the same strength of evidence when compared with fixed sampling alternatives. The key challenge is making design decisions given the uncertainty about how the trial will develop.

For RAR, MAMS, and DTL designs, the trial focuses on those treatments demonstrating effectiveness. These

methods are appealing as they increase the chance of receiving a treatment that is more likely to be effective. Further to this, the model-based approach of MCP-Mod increases understanding of the relationship between dose and response in order to better allocate patients based on current trial data, in turn allowing greater confidence about the choice of dose.

Which patients will benefit?

Late in the development cycle, such as phase III of drug development, we wish to confirm the treatment is effective. An important aspect of this is to ensure the right patients receive the treatment (i.e. those who will gain a meaningful benefit). Here, we focus on trials that use clinically relevant biomarkers to identify patients who may be sensitive to a treatment and therefore likely to respond.

Covariate-adjusted response adaptive

A form of RAR (see the “[Response-adaptive randomisation](#)” section) that accounts for patient differences is covariate-adjusted response adaptive (CARA). Randomisation probabilities are aligned to the patient’s observed biomarker information skewing allocation probabilities towards the best performing arms according to a patient’s set of characteristics. Such changes based upon available biomarker information are one of the most common adaptations in biomarker adaptive designs [149].

CARA procedures are sometimes (incorrectly) referred to as minimisation procedures or dynamic allocation; the methods referred to with these names are very different in their goals and nature. For example, some CARA procedures have been proposed altering the randomisation (similar to RAR) [150] while other methods do not do so in a randomised fashion determining allocation probabilities based solely on covariates [151–153]. Some CARA procedures are designed to minimise imbalances on important covariates only [150, 151] while other methods have an efficiency goal, being designed to minimise the variance of the treatment effect in the presence of covariates [154]. Finally, some CARA rules will aim to assign the largest number of patients to the best treatment while accounting for patients’ differences in biomarkers [155].

Example The BATTLE [156] trial is a prospective, biopsy-mandated, biomarker-based, adaptively randomised [157] study conducted in 255 pre-treated lung cancer patients. Initially, 97 patients were randomised equally to four arms: erlotinib, vandetanib, erlotinib plus bevacizumab, or sorafenib, based on relevant biomarkers. Then, for the remaining 158 patients, the allocation probabilities were adapted using a CARA procedure. This procedure used a Bayesian adaptive algorithm: the data from the first 97 patients were assessed giving a prior distribution (describing the likely values of the effect of

the treatment) for disease control rate (DCR, the primary endpoint) in each biomarker group; this prior distribution was continuously updated using the accumulating data as more patients were observed giving a posterior distribution (describing the likely values of the effect of the treatment having combined information from the prior distribution and the available trial data) for DCR in each biomarker group; upon recruiting, any new patient to the trial their randomisation was governed by the currently available data using posterior distribution. Results include a 46% 8-week disease control rate (primary endpoint) and evidence of an impressive benefit from sorafenib among mutant-KRAS patients.

Advantages/disadvantages Because of the similarity to RAR, advantages and disadvantages of CARA designs are very similar. The main advantage being that they allow flexibility, introducing balance, efficiency, or ethical goals according to what might be more relevant. In addition, CARA designs make assumptions about the model for biomarker interactions for patients in the trial and thus are further sensitive to these assumptions.

Population enrichment

Population enrichment designs are useful when biomarker-defined sub-groups are known before the trial commences. With uncertainty about which populations should be recruited, they select which sub-populations to recruit from for the remainder of the trial based on data available at the interim analysis. Figure 4 illustrates how an adaptive enrichment design can progress. In a non-adaptive trial, the study team must make this sub-population selection before the trial begins. Population enrichment designs are typically planned with one of two goals in mind: target the sub-population where patients receive the greatest benefit, or stop recruiting from sub-populations where the treatment may not provide a benefit.

Flexible hypothesis testing methods [101, 102, 158] preserve the statistical integrity of the trial while allowing freedom in the selection of sub-populations. The decision-making methodology is key in population enrichment with several approaches, both Bayesian [159–161] and classical techniques being applied [162, 163].

Example TAPPAS [164, 165] is a trial of TRC105 (an antibody) and pazopanib versus pazopanib alone in patients with advanced angiosarcoma. The study identified two sub-groups, those with cutaneous advanced angiosarcoma and those with non-cutaneous advanced angiosarcoma. There was an indication of greater tumour sensitivity to TRC105 in the cutaneous sub-group. The primary endpoint for this study was progression-free survival, with

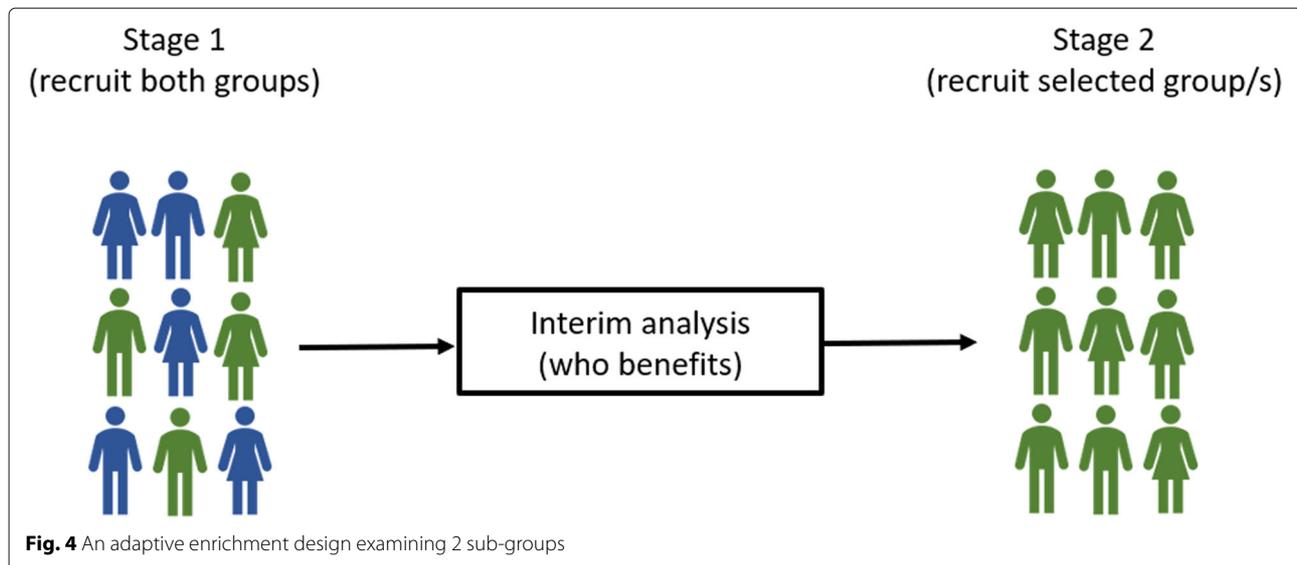


Fig. 4 An adaptive enrichment design examining 2 sub-groups

an initial sample size of 124 patients to be followed until 95 events (progression or death) have been observed.

A population enrichment design was used where the data monitoring committee was able to recommend one of three pre-planned actions at the interim analysis: continue as planned with the full population (recruiting 124 patients followed until 95 events in total), continue with the full population and an increase in sample size and progression-free survival events (recruiting 200 patients followed until 170 events in total), and continue with only the cutaneous sub-group, thereby enriching the study population (recruiting 170 patients followed until 110 events in total). This decision-making procedure followed the promising zone approach [166], where the choice about which option would be followed is based upon the probability of rejecting the null hypotheses given the data available at the interim analysis.

The study recruited from the full population throughout, as dictated by the data observed at the interim analysis. In total, 128 patients were recruited (close to the targeted recruitment of 124), with 64 patients randomised to each the experimental treatment and the control. It was concluded that TRC105 did not demonstrate activity when combined with pazopanib.

Advantages Population enrichment designs recruit fewer patients from sub-groups that do not benefit from the treatment focusing on those patients who receive a benefit. When there is uncertainty about which sub-groups benefit from the new treatment, the adaptive method can offer an improvement over non-adaptive alternatives. It is able to offer the benefit of increasing the sample size in one sub-group without sacrificing the opportunity to test the new treatment in all patients before the observation of any patients. In addition, popu-

lation enrichment offers a compromise between the possible fixed designs (for example recruiting only from one of the two sub-groups, or the full population). This is appealing where there is disagreement between which (sub)group(s) should be recruited.

Disadvantages Computing optimal decision rules and evaluating overall trial performance are non-trivial typically requiring some form of simulation [160], increasing the workload in setting up the trial. While possible changes to eligibility criteria during the course of a trial add challenges to the operational planning of patient recruitment.

Summary

In this section, we have discussed designs suitable for phase II and phase III of clinical development with a focus on a biomarker-guided approach to treatment. Despite active research in this area, these methods appear to have the lowest uptake at the time of writing, although with the drive towards personalised healthcare they are likely to become increasingly relevant.

The designs discussed target pre-defined sub-populations allowing formal testing for efficacy of an experimental treatment in biomarker-defined sub-populations. Of note, these methods assume these biomarker groups are well defined, with little work on how to properly account for this should this assumption be violated [167]. Beyond being structured to allow formal analysis of sub-groups, there is also a large ethical benefit to these designs, exposing as few patients as possible to treatments from which they may not receive a benefit.

Should no such pre-defined biomarkers be available, one might consider adaptive signature designs [168, 169], also referred to as biomarker adaptive designs [149, 170, 171].

They aim to identify and use predictive biomarkers [170] during the trial. They help to improve the chances of identifying patients who will benefit from the treatment, while still providing accurate treatment effect estimates; this maximises the use of the trial data. However, the identification of predictive biomarkers adds a large amount of complexity.

Umbrella [172] and basket trials [173] make use of more detailed biomarker information. Broadly speaking, a basket trial examines a single experimental treatment in multiple sub-types of a single biomarker, while an umbrella trial may consider multiple biomarkers each with a specific treatment.

Does the treatment work?

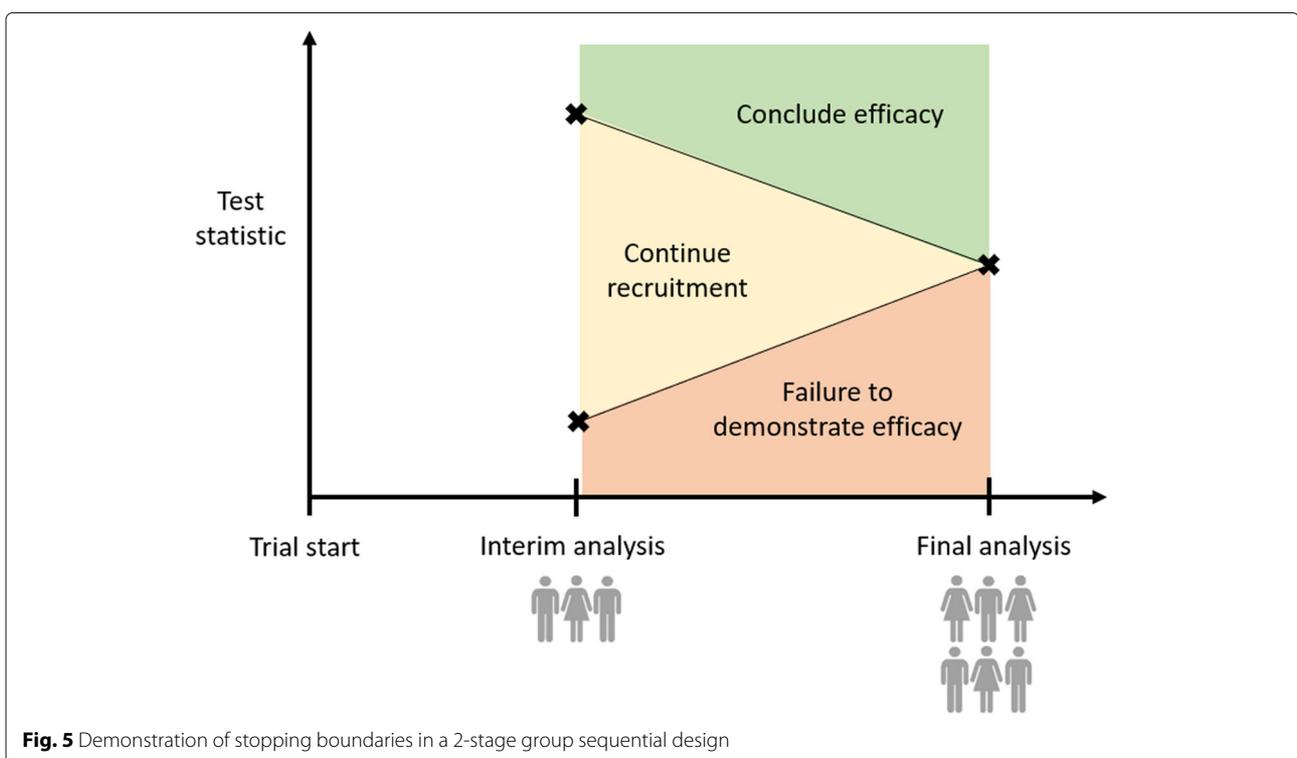
In this section, we consider treatments late in the development cycle, typically phase III (although viable in phase II also), where we wish to show conclusively that a novel treatment is an improvement over the current standard of care. A conventional aim at this stage of development is to be efficient, both in terms of the time that is required to conduct the trial and the number of patients; the methods that follow attempt to make the trial more efficient while also ensuring that the overall probability of a successful outcome is maintained or increased. Methods aiming to be efficient in recruiting the correct number of patients have both an operational and ethical benefit, ensuring patients are not subject to an experimental treatment for a

trial that is underpowered (i.e. has a low chance of detecting a meaningful effect) or when their contribution to the overall result is not required.

Group sequential designs

Group sequential designs are the most widespread [174] of the adaptive designs we consider. They differ from a more traditional phase III clinical trial in that one or more pre-planned interim analyses may be used to assess efficacy; if there is strong evidence the experimental treatment is superior to control, or indeed that both have the same effect, then the trial may be terminated early as depicted in Fig. 5.

To enable early stopping of the trial while maintaining control of the type I error rate, group sequential trials use pre-defined stopping rules. At each interim analysis, the current data for the experimental and control arms are compared to construct a test statistic. If the test statistic is sufficiently high/low, the trial is stopped early for efficacy/lack of a demonstrated benefit of the treatment; if neither criterion is met, the trial continues to the next interim or final analysis. Several approaches to the definition of stopping rules for group sequential designs have been proposed [5, 175–177]. Jennison and Turnbull [178] provide a comprehensive guide to group sequential designs and their application. Whitehead [179] describes the implementation of group sequential methods using SAS; in addition, the gsDesign [180] and optGS [181]



packages in R are open source alternatives allowing the construction and analysis of group sequential designs.

Example The INTERCEPT trial [182, 183] was a randomised, double-blind trial in patients with acute myocardial infarction. The trial used a group sequential design to achieve 80% power to detect a 33% between-group difference in cumulative first event rate of cardiac death, non-fatal reinfarction, or refractory ischaemia. Interim analyses were conducted after enrolment of 300 patients, and at intervals of roughly 300 patients thereafter. The stopping boundaries were constructed using a double triangular test [177].

Recruitment was stopped after the third interim analysis as the pre-specified 33% between-group difference for the primary endpoint would likely not be observed. In total, the trial recruited 874 patients. A total of 430 patients were randomised to 300 mg oral diltiazem once daily, and 444 patients were randomised to placebo initiated within 36–96 h of infarct onset, and given for up to 6 months.

Advantages Group sequential designs reduce the expected sample size of the trial compared to a traditional trial design with a fixed sample size. This is due to the possibility the trial will stop early due to strong evidence that the treatment either is or is not effective. For O'Brien-Fleming [175] boundaries, this reduction is typically around 15% compared to a fixed design [55]. Group sequential designs are optimal in terms of minimising the expected sample size [178]. Thus, any other design aiming to reduce the expected sample size can only perform as well as the group sequential option. In addition, with group sequential methods being widespread [174], their implementation is likely to be more familiar to regulators and ethics committees.

Disadvantages The opportunity for early stopping requires careful definition of stopping boundaries, and there is a cost to this adjustment. While the expectation is for the design to reduce the sample size by stopping early, it is possible that the group sequential design will recruit more patients than a non-adaptive method would have (when no early stopping criterion is met), usually by approximately 10% [55]. Due to uncertainty about the length of the trial before it commences, logistics and planning are more complex than for a traditional fixed sample design.

Sample size re-assessment

Sample size re-assessment (also known as sample size re-estimation or sample size re-calculation) seeks to ensure an appropriate sample size for the trial despite uncertainty about key design parameters (such as the variance in the observations) by re-assessing the required sample

size during the trial. Typically, the re-assessment focuses on the estimation of design parameters at an interim analysis to re-calculate the sample size in order to achieve, for example, a desired conditional power [184–186] (the probability of rejecting the null hypothesis given the currently available data). Usually, this allows an increase but not a decrease to the sample size, often with a limit on what maximum sample size is possible, as large increases to the sample size in this way can be inefficient [187]; in practice, if the estimated sample size exceeds the pre-set maximum, then recruitment is usually stopped.

Sample size re-assessment may be either blinded or unblinded [188], while maintaining the statistical integrity of the trial. In unblinded sample size re-estimation, interim analyses are conducted based on unblinded trial data; that is, the statistician performing the interim analysis will know which participants are in which trial arm. Updated estimates of parameters related to the sample size are used to re-assess the sample size for the remainder of the trial [189–191]. Conversely, blinded sample size re-estimation only makes use of the combined trial data from all treatment arms [184, 185, 192, 193]. Alterations to the sample size in this way have a minimal impact on the type I error rate, even without formal adjustment [193]. The suitability of blinded or unblinded re-assessment will depend on the parameters that require re-estimation.

The sample size re-assessment must be conducted before the originally planned conclusion of the trial. That is, it must not be used where a trial has completed and failed to reject the null hypothesis. Indeed, it should not be used at all with the intention of recovering a significant result. Equally severely reducing the number of patients to be recruited using sample size re-estimation can be problematic, where it may inflate the probability of falsely rejecting the null hypothesis [194]. Note also that it can be counterproductive to make changes to a trial in progress based on estimates of the treatment effect from the trial itself, as small or no treatment effect results in increased sample sizes so that less useful treatments require more resource. In such a case, a group sequential design where the interim analyses are planned to correspond to different treatment effects of interest (e.g. using an optimistic effect at first interim, moderate effect at second, and minimally relevant effect at final interim) is likely a better choice.

Example Hade et al. [195] discuss a sample size re-assessment in a randomised trial in breast cancer where the primary outcome is disease-free survival. Women were randomised either to immediate surgery in the next 1–6 days, which was expected to be in the follicular phase of the menstrual cycle, or to scheduled surgery during the next mid-luteal phase of the menstrual cycle. Based on primary analysis by log-rank test with a target of 80%

power, with 5% two-sided type I error rate, to detect a hazard ratio (HR) of 0.58 in favour of scheduled surgery, this required 113 events. With accrual time of 2 years and 4 additional years of follow-up and a 2–3% loss to follow-up, the initial study design planned to randomise 340 women.

The HR of 0.58 was felt to be optimistic based on available information and sources external to the trial. A blinded sample size re-assessment using the available data increased the sample size by 170 patients, to a total of 510 randomised patients (with a required 175 events) in order to target a revised HR of 0.65.

Advantages The main aim of sample size re-assessment is to ensure that the trial recruits an appropriate number of patients. Sample size re-assessment designs are not as complex as many other adaptive methods, allowing the trial to be planned more quickly. The fact that both unblinded and blinded methods are available means that sample size re-assessment can be applied in many different settings. There is an upward trend in the use of sample size re-assessment in clinical practice, and as these designs become more widespread, it will become increasingly easy to put them forward.

Disadvantages This is a method with few drawbacks. There is a small additional burden at the interim analysis to properly estimate the required sample size for the remainder of the trial, which requires appropriate expertise in the case of both blinded and unblinded interim analyses. Most of the practical issues that a sample size re-assessment design may bring (time constraints or securing sufficient funding in advance) are similar to those faced when using other adaptive designs, while further concerns may be raised from extending the trial beyond the originally planned end. An example is the comparability of patients recruited early and those recruited late to a modified trial [196].

Summary

These methods are the most widespread of the adaptive designs we have considered [174], to the extent group sequential trials may even be considered a standard approach. The group sequential framework forms a foundation for many other adaptive methods due to its preservation of the integrity of the trial results.

The fact that such methods are so well established in practice shows promise for the use of the adaptive methods discussed in this paper. Despite the complexity, these methods are sufficiently well-known in the trials community with software support and common practices established allowing their implementation. The other methods discussed throughout this work share many similarities in terms of methodological complexity while each has their own advantages/disadvantages, but these are not so far

removed from those that have been overcome for group sequential designs so as to make overcoming these hurdles an impossibility.

Conclusions

Research into adaptive designs has become more prevalent across all stages of clinical development, although this increase is not necessarily reflected by their uptake in practice. The suitability of adaptive methods depends largely on the clinical question being addressed. We have presented four key clinical questions for which adaptive designs may be of use across a wide range of disease areas, study settings, and endpoints. For each possible design, there are advantages and disadvantages with some key themes: increase in efficiency of the design in terms of the expected number of patients or a clear benefit in understanding the question of scientific interest, clear ethical advantages to ensuring the right patients are given the best available treatment whenever possible, and the key disadvantage is the additional burden, both in planning the trial and the interim analyses. Importantly, while the adaptive methods can be highly effective when used in the correct scenario, an adaptive method is not always the best choice [197], so careful consideration must be taken before their use.

At the design stage of any trial (adaptive or not), some design assumptions must be made. These will influence the performance of the trial and inadequate assumptions can lead to a sub-optimal design. With the additional complexity of many adaptive designs, there are more assumptions to be made and it is critical that these are well understood by the trial team to consider the impact of these choices, although for a corresponding fixed sample design many of these assumptions must be made anyway. As noted in the “Summary” of the “Does the treatment work?” section, many such problems have been well worked out for group sequential designs and are not insurmountable for the other designs discussed. Communication and establishing a common practice between the methodology and trial community will be key in seeing the wider spread application of such methods.

Regulatory bodies are increasingly recognising the desire for the use of adaptive designs and accepting their use although it is recommended regulators be engaged early in the process whenever using any novel methodology [55, 198]. Funding bodies are also increasingly comfortable with the use of adaptive designs; the TAI-LoR trial discussed in the ‘Multi-arm multi-stage’ section appears as a case study on the National Institute for Health Research website [199]. Additionally, new reporting guidance for adaptive designs [200] has recently been published to facilitate uptake further.

We have not been exhaustive in our discussion of adaptive designs, focusing on the key designs to answer the

most common questions of clinical interest. Seamless designs [101, 102, 201], which we have not discussed in detail, use similar adaptive design methodology to combine phases of clinical development: designs may be inferentially seamless, where data from the earlier stage are incorporated into the overall trial results; operationally seamless, avoiding any break in recruitment between the stages of the development process but excluding data from earlier stages from the final analysis of the latter; or both. There are many motivations for conducting seamless designs [202], making it an active area of research [203].

For many years, one major obstacle to the use of adaptive designs in practice has been the lack of suitable software to aid both the design and conduct of trials. This issue is increasingly being tackled by those researching the methods, with many open source packages available for the design and analysis of adaptive methods some of which have been cited in this work. For example *rpact* [204] is an R package that assists in the design and analysis of confirmatory clinical trials. In addition, there is a steep learning curve to the implementation of such designs; training courses are becoming increasingly available to address this.

From a methodology standpoint, there are some further issues that go beyond the level of detail we have discussed that should be considered when proposing an adaptive design, for example potential information leakage or the introduction of bias [205, 206]. The Practical Adaptive and Novel Designs and Analysis (PANDA) toolkit [207] is under development at the time of writing and will be an online resource that addresses and explains broader issues in the use of adaptive designs.

Despite the challenges in the design and analysis of an adaptive trial, we believe that under the right circumstance, the benefits introduced by the increased flexibility clearly outweigh these issues.

Acknowledgements

Not applicable.

Authors' contributions

All authors read and approved the final manuscript.

Funding

TB was supported by the MRC Network of hubs for Trials Methodology HTMR Award MR/L004933/1. SSV is supported by UK Medical Research Council (grant number: MC_UU_00002/3). GMW is supported by Cancer Research UK. This report is an independent research arising in part from Prof Jaki's Senior Research Fellowship (NIHR-SRF-2015-08-001) and Dr Pavel Mozgunov's Fellowship (National Institute for Health Research Advanced Fellowship, Dr Pavel Mozgunov, NIHR300576) supported by the National Institute for Health Research. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health and Social Care (DHCS). TJ is also supported by UK Medical Research Council (grant number: MC_UU_0002/14).

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Mathematics and Statistics, Lancaster University, Fylde College, Lancaster, LA1 4YF, UK. ²Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University, Cardiff, UK. ³MRC Biostatistics Unit, University of Cambridge School of Clinical Medicine, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK. ⁴Cancer Research UK & UCL Cancer Trials Centre, University College London, 90 Tottenham Court Road, London, W1T 4TJ, UK.

Received: 29 June 2020 Accepted: 7 October 2020

Published online: 19 November 2020

References

- Friedman L, Furberg C, DeMets D, Reboussin D, Granger C, et al. *Fundamentals of clinical trials*, vol. 4. New York: Springer; 2010, pp. 85–115.
- Pallmann P, Bedding A, Choodari-Oskooei B, Dimairo M, Flight L, Hampson L, Holmes J, Mander A, Sydes M, Villar S, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med*. 2018;16(1):29.
- Hampson L, Williamson P, Wilby M, Jaki T. A framework for prospectively defining progression rules for internal pilot studies monitoring recruitment. *Stat Methods Med Res*. 2018;27(12):3612–27.
- Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Stat Med*. 2016;35(3):325–47.
- Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977;64(2):191–9.
- Chow S-C, Chang M, Pong A. Statistical consideration of adaptive methods in clinical development. *J Biopharm Stat*. 2005;15(4):575–91.
- Dimairo M, Coates E, Pallmann P, Todd S, Julious S, Jaki T, Wason J, Mander A, Weir C, Koenig F, et al. Development process of a consensus-driven consort extension for randomised trials using an adaptive design. *BMC medicine*. 2018;16(1):210.
- Le Tourneau C, Lee J, Siu L. Dose escalation methods in phase I cancer clinical trials. *JNCI: J Natl Cancer Inst*. 2009;101(10):708–20.
- Jaki T. Uptake of novel statistical methods for early-phase clinical studies in the UK public sector. *Clin Trials*. 2013;10(2):344–6.
- Chevret S. Bayesian adaptive clinical trials: a dream for statisticians only? *Stat Med*. 2012;31(11–12):1002–13.
- Dimairo M, Julious S, Todd S, Nicholl J, Boote J. Cross-sector surveys assessing perceptions of key stakeholders towards barriers, concerns and facilitators to the appropriate use of adaptive designs in confirmatory trials. *Trials*. 2015;16(1):585.
- Dimairo M, Boote J, Julious S, Nicholl J, Todd S. Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials. *Trials*. 2015;16(1):430.
- Dragalin V. Adaptive designs: terminology and classification. *Drug Inf J*. 2006;40(4):425–35.
- Stallard N, Hampson L, Benda N, Brannath W, Burnett T, Friede T, Kimani P, Koenig F, Krisam J, Mozgunov P, Posch M, Wason J, Wassmer G, Whitehead J, Williamson S, Zohar S, Jaki T. Efficient adaptive designs for clinical trials of interventions for covid-19. *Stat Biopharm Res*. 2020;0(0):1–15.
- of Health N. National Cancer Institute. NCI Dictionary of Cancer Terms. 2014. <https://www.cancer.gov/publications/dictionaries/cancer-terms>. Accessed 26 Apr 2020.
- Carter S. Study design principles for the clinical evaluation of new drugs as developed by the chemotherapy program of the national cancer institute. *The design of clinical trials in cancer therapy*. 1973;242–89.
- Storer B. Design and analysis of phase I clinical trials. *Biometrics*. 1989;45(3):925–37.

18. Lin Y, Shih W. Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics*. 2001;2(2):203–15.
19. Park S, Bang S-M, Cho E, Shin D, Lee J, Lee W, Chung M. Phase I dose-escalating study of docetaxel in combination with 5-day continuous infusion of 5-fluorouracil in patients with advanced gastric cancer. *BMC cancer*. 2005;5(1):87.
20. Wheeler G, Sweeting M, Mander A. Aplusb: a web application for investigating a+b designs for phase I cancer clinical trials. *PLoS one*. 2016;11(7):0159026.
21. Kang S-H, Ahn C. The expected toxicity rate at the maximum tolerated dose in the standard phase I cancer clinical trial design. *Drug Inf J*. 2001;35(4):1189–99.
22. Kang S-H, Ahn C. An investigation of the traditional algorithm-based designs for phase I cancer clinical trials. *Drug Inf J*. 2002;36(4):865–73.
23. He W, Liu J, Binkowitz B, Quan H. A model-based approach in the estimation of the maximum tolerated dose in phase I cancer clinical trials. *Stat Med*. 2006;25(12):2027–42.
24. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*. 1990;46:33–48.
25. O'Quigley J, Shen L. Continual reassessment method: a likelihood approach. *Biometrics*. 1996;52(2):673–84.
26. Wheeler G, Mander A, Bedding A, Brock K, Cornelius V, Grieve A, Jaki T, Love S, Weir C, Yap C, et al. How to design a dose-finding study using the continual reassessment method. *BMC Med Res Methodol*. 2019;19(1):1–15.
27. Paoletti X, Baron B, Schöffski P, Fumoleau P, Lacombe D, Marreaud S, Sylvester R. Using the continual reassessment method: lessons learned from an early phase I dose finding study. *Eur J Cancer*. 2006;42(10):1362–8.
28. Ratain M, Mick R, Schilsky R, Siegler M. Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents. *J Natl Cancer Inst*. 1993;85(20):1637–43.
29. O'Quigley J, Zohar S. Experimental designs for phase I and phase II dose-finding studies. *Br J Cancer*. 2006;94(5):609.
30. O'Quigley J, Shen L, Gamst A. Two-sample continual reassessment method. *J Biopharm Stat*. 1999;9(1):17–44.
31. Thall P, Millikan R, Mueller P, Lee S-J. Dose-finding with two agents in phase I oncology trials. *Biometrics*. 2003;59(3):487–96.
32. Iasonos A, Wilton A, Riedel E, Seshan V, Spriggs D. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in phase I dose-finding studies. *Clin Trials*. 2008;5(5):465–77.
33. Onar A, Kocak M, Boyett J. Continual reassessment method vs. traditional empirically based design: modifications motivated by phase I trials in pediatric oncology by the pediatric brain tumor consortium. *J Biopharm Stat*. 2009;19(3):437–55.
34. Onar-Thomas A, Xiong Z. A simulation-based comparison of the traditional method, rolling-6 design and a frequentist version of the continual reassessment method with special attention to trial duration in pediatric phase I oncology trials. *Contemp Clin Trials*. 2010;31(3):259–70.
35. Conaway M, Petroni G. The impact of early-phase trial design in the drug development process. *Clin Cancer Res*. 2019;25(2):819–27.
36. Lee S, Cheng B, Cheung Y. Continual reassessment method with multiple toxicity constraints. *Biostatistics*. 2010;12(2):386–98.
37. Iasonos A, Zohar S, O'Quigley J. Incorporating lower grade toxicity information into dose finding designs. *Clin Trials*. 2011;8(4):370–9.
38. Van Meter E, Garrett-Mayer E, Bandyopadhyay D. Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method. *Clin trials*. 2012;9(3):303–13.
39. Braun T. The bivariate continual reassessment method: extending the crm to phase I trials of two competing outcomes. *Control Clin Trials*. 2002;23(3):240–56.
40. Zohar S, O'Quigley J. Optimal designs for estimating the most successful dose. *Stat Med*. 2006;25(24):4311–20.
41. Zohar S, O'Quigley J. Identifying the most successful dose (msd) in dose-finding studies in cancer. *Pharm Stat*. 2006;5(3):187–99.
42. Zhong W, Koopmeiners J, Carlin B. A trivariate continual reassessment method for phase I/II trials of toxicity, efficacy, and surrogate efficacy. *Stat Med*. 2012;31(29):3885–95.
43. Yeung W, Whitehead J, Reigner B, Beyer U, Diack C, Jaki T. Bayesian adaptive dose-escalation procedures for binary and continuous responses utilizing a gain function. *Pharm Stat*. 2015;14(6):479–87.
44. Yeung W, Reigner B, Beyer U, Diack C, Sabanes Bové D, Palermo G, Jaki T. Bayesian adaptive dose-escalation designs for simultaneously estimating the optimal and maximum safe dose based on safety and efficacy. *Pharm Stat*. 2017;16(6):396–413.
45. Cheung Y, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*. 2000;56(4):1177–82.
46. Braun T. Generalizing the tite-crm to adapt for early-and late-onset toxicities. *Stat Med*. 2006;25(12):2071–83.
47. Kramar A, Lebecqz A, Candali E. Continual reassessment methods in phase I trials of the combination of two drugs in oncology. *Stat Med*. 1999;18(14):1849–64.
48. Wang K, Ivanova A. Two-dimensional dose finding in discrete dose space. *Biometrics*. 2005;61(1):217–22.
49. Yuan Y, Yin G. Sequential continual reassessment method for two-dimensional dose finding. *Stat Med*. 2008;27(27):5664–78.
50. Wages N, Conaway M, O'Quigley J. Dose-finding design for multi-drug combinations. *Clin Trials*. 2011;8(4):380–9.
51. Harrington J, Wheeler G, Sweeting M, Mander A, Jodrell D. Adaptive designs for dual-agent phase I dose-escalation studies. *Nat Rev Clin Oncol*. 2013;10(5):277.
52. Riviere M-K, Yuan Y, Dubois F, Zohar S. A bayesian dose-finding design for drug combination clinical trials based on the logistic model. *Pharm Stat*. 2014;13(4):247–57.
53. Riviere M-K, Dubois F, Zohar S. Competing designs for drug combination in phase I dose-finding clinical trials. *Stat Med*. 2015;34(1):1–12.
54. Wages N, Slingluff Jr C, Petroni G. Statistical controversies in clinical research: early-phase adaptive design for combination immunotherapies. *Ann Oncol*. 2016;28(4):696–701.
55. Food and Drug Administration. Adaptive design clinical trials for drugs and biologics guidance for industry. 2019;22:2623–9.
56. Nie L, Rubin E, Mehrotra N, Pinheiro J, Fernandes L, Roy A, Bailey S, de Alwis D. Rendering the 3+3 design to rest: more efficient approaches to oncology dose-finding trials in the era of targeted therapy. *AACR*. 2016;22:2623–9.
57. Yap C, Billingham L, Cheung Y, Craddock C, O'Quigley J. Dose transition pathways: the missing link between complex dose-finding designs and simple decision-making. *Clin Cancer Res*. 2017;23(24):7440–7.
58. MD Anderson Center: Biostatistics Software. <https://biostatistics.mdanderson.org/SoftwareDownload/>. Accessed 26 Apr 2020.
59. Bové D, Yeung W, Palermo G, Jaki T. Model-based dose escalation designs in R with crmpack. *J Stat Softw*. 2019;89(10):1–22.
60. Sweeting M, Mander A, Sabin T, et al. Bcrim: Bayesian continual reassessment method designs for phase I dose-finding trials. *J Stat Softw*. 2013;54(13):1–26.
61. Wages N, Petroni G. A web tool for designing and conducting phase I trials using the continual reassessment method. *BMC cancer*. 2018;18(1):133.
62. Chen S-C, Shyr Y. A web application for optimal selection of adaptive designs in phase I oncology clinical trials. In: *Trials*. London: Biomed Central LTD; 2017.
63. Pallmann P, Wan F, Mander A, Wheeler G, Yap C, Clive S, Hampson L, Jaki T. Designing and evaluating dose-escalation studies made easy: the MoDESt web app. *Clin Trials*. 2020;17(2):147–56.
64. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Stat Med*. 1998;17(10):1103–20.
65. Wheeler G, Sweeting M, Mander A. Toxicity-dependent feasibility bounds for the escalation with overdose control approach in phase I cancer trials. *Stat Med*. 2017;36(16):2499–513.
66. Tighiouart M, Rogatko A. Dose finding with escalation with overdose control (ewoc) in cancer clinical trials. *Stat Sci*. 2010;25(2):217–26.
67. Nishio M, Murakami H, Horiike A, Takahashi T, Hirai F, Suenaga N, Tajima T, Tokushige K, Ishii M, Boral A, et al. Phase I study of ceritinib (ldk378) in Japanese patients with advanced, anaplastic lymphoma kinase-rearranged non-small-cell lung cancer or other tumors. *J Thorac Oncol*. 2015;10(7):1058–66.
68. Tighiouart M, Liu Y, Rogatko A. Escalation with overdose control using time to toxicity for cancer phase I clinical trials. *PLoS one*. 2014;9(3):93070.
69. Shi Y, Yin G. Escalation with overdose control for phase I drug-combination trials. *Stat Med*. 2013;32(25):4400–12.

70. Tighiouart M, Piantadosi S, Rogatko A. Dose finding with drug combinations in cancer phase i clinical trials using conditional escalation with overdose control. *Stat Med*. 2014;33(22):3815–29.
71. Mozhgunov P, Jaki T. Improving safety of the continual reassessment method via a modified allocation rule. *Stat Med*. 2020;39(7):906–22.
72. Berry S, Carlin B, Lee J, Muller P. Bayesian adaptive methods for clinical trials. Raton, FL: CRC press; 2010.
73. Babb J, Rogatko A. Patient specific dosing in a cancer phase i clinical trial. *Stat Med*. 2001;20(14):2079–90.
74. Cheng J, Babb J, Langer C, Aamdal S, Robert F, Engelhardt L, Fernberg O, Schiller J, Forsberg G, Alpaugh R, et al. Individualized patient dosing in phase i clinical trials: the role of escalation with overdose control in pnu-214936. *J Clin Oncol*. 2004;22(4):602–9.
75. Wheeler G. Incoherent dose-escalation in phase i trials using the escalation with overdose control approach. *Stat Papers*. 2018;59(2):801–11.
76. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri F, Porter A. Translation of innovative designs into phase i trials. *J Clin Oncol*. 2007;25(31):4982–6.
77. Le Tourneau C, Gan H, Razak A, Paoletti X. Efficiency of new dose escalation designs in dose-finding phase i trials of molecularly targeted agents. *PLoS one*. 2012;7(12):51039.
78. Rivoirard R, Vallard A, Langrand-Escure J, Mrad M, Wang G, Guy J-B, Diao P, Dubanchet A, Deutsch E, Rancoule C, et al. Thirty years of phase i radiochemotherapy trials: latest development. *Eur J Cancer*. 2016;58:1–7.
79. Paoletti X, Ezzalfani M, Le Tourneau C. Statistical controversies in clinical research: requiem for the 3+3 design for phase i trials. *Ann Oncol*. 2015;26(9):1808–12.
80. Jaki T, Clive S, Weir C. Principles of dose finding studies in cancer: a comparison of trial designs. *Cancer Chemother Pharmacol*. 2013;71(5):1107–14.
81. Mandrekar S, Cui Y, Sargent D. An adaptive phase i design for identifying a biologically optimal dose for dual agent drug combinations. *Stat Med*. 2007;26(11):2317–30.
82. O'Quigley J, Hughes M, Fenton T. Dose-finding designs for hiv studies. *Biometrics*. 2001;57(4):1018–29.
83. Lu N, Crespi C, Liu N, Vu J, Ahmadiyeh Y, Wu S, Lin S, McClune A, Durazo F, Saab S, et al. A phase i dose escalation study demonstrates quercetin safety and explores potential for bioflavonoid antivirals in patients with chronic hepatitis c. *Phytother Res*. 2016;30(1):160–8.
84. Whitehead J, Zhou Y, Stevens J, Blakey G, Price J, Leadbetter J. Bayesian decision procedures for dose-escalation based on evidence of undesirable events and therapeutic benefit. *Stat Med*. 2006;25(1):37–53.
85. Lyden P, Pryor K, Coffey C, Cudkovic M, Conwit R, Jadhav A, Sawyer Jr R, Claassen J, Adeoye O, Song S, et al. Randomized, controlled, dose escalation trial of a protease-activated receptor-1 agonist in acute ischemic stroke: final results of the rhapsody trial: a multi-center, phase 2 trial using a continual reassessment method to determine the safety and tolerability of 3k3a-apc, a recombinant variant of human activated protein c, in combination with tissue plasminogen activator, mechanical thrombectomy or both in moderate to severe acute ischemic stroke. *Ann Neurol*. 2019;85(1):125.
86. Yuan Y, Hess K, Hilsenbeck S, Gilbert M. Bayesian optimal interval design: a simple and well-performing design for phase i oncology trials. *Clin Cancer Res*. 2016;22(17):4291–301.
87. Zhang L, Yuan Y. A practical bayesian design to identify the maximum tolerated dose contour for drug combination trials. *Stat Med*. 2016;35(27):4924–36.
88. Haines L, Perevozskaya I, Rosenberger W. Bayesian optimal designs for phase i clinical trials. *Biometrics*. 2003;59(3):591–600.
89. Azriel D. Optimal sequential designs in phase i studies. *Comput Stat Data Anal*. 2014;71:288–97.
90. Haines L, Clark A. The construction of optimal designs for dose-escalation studies. *Stat Comput*. 2014;24(1):101–9.
91. Liu S, Yuan Y. Bayesian optimal interval designs for phase i clinical trials. *J R Stat Soc: Ser C: Appl Stat*. 2015;64(3):507–23.
92. Gasparini M, Eisele J. A curve-free method for phase i clinical trials. *Biometrics*. 2000;56(2):609–15.
93. Mander A, Sweeting M. A product of independent beta probabilities dose escalation design for dual-agent phase i trials. *Stat Med*. 2015;34(8):1261–76.
94. Mozhgunov P, Jaki T. An information theoretic phase i–ii design for molecularly targeted agents that does not require an assumption of monotonicity. *J R Stat Soc: Ser C: Appl Stat*. 2019;68(2):347–67.
95. LoRusso P, Boerner S, Seymour L. An overview of the optimal planning, design, and conduct of phase i studies of new therapeutics. *Clin Cancer Res*. 2010;16(6):1710–8.
96. Chevret S. Bayesian adaptive clinical trials: a dream for statisticians only? *Stat Med*. 2012;31(11–12):1002–13.
97. Jaki T. Multi-arm clinical trials with treatment selection: what can be gained and at what price? *Clin Investig*. 2015;5(4):393–9.
98. Stallard N, Todd S. Sequential designs for phase iii clinical trials incorporating treatment selection. *Stat Med*. 2003;22(5):689–703.
99. Magirr D, Jaki T, Whitehead J. A generalized dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika*. 2012;99(2):494–501.
100. Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. *Stat Med*. 1999;18(14):1833–48.
101. Bretz F, Schmidli H, König F, Racine A, Maurer W. Confirmatory seamless phase ii/iii clinical trials with hypotheses selection at interim: general concepts. *Biom J*. 2006;48(4):623–34.
102. Schmidli H, Bretz F, Racine A, Maurer W. Confirmatory seamless phase ii/iii clinical trials with hypotheses selection at interim: applications and practical considerations. *Biom J*. 2006;48(4):635–43.
103. Jaki T, Pallmann P, Magirr D. The r package mams for designing multi-arm multi-stage clinical trials. *J Stat Softw*. 2019;88(4):.
104. Royston P, Bratton D, Choodari-Oskooei B, Barthel F-S. Nstage: Stata module for multi-arm, multi-stage (mams) trial designs for time-to-event outcomes. 2019. Boston College Department of Economics.
105. Bratton D. Nstagebin: Stata module to perform sample size calculation for multi-arm multi-stage randomised controlled trials with binary outcomes. 2014. Boston College Department of Economics.
106. Grayling M. Desma: Stata module to design and simulate (adaptive) multi-arm clinical trials. 2019. Boston College Department of Economics.
107. Pushpakom S, Taylor C, Kolamunnage-Dona R, Spowart C, Vora J, García-Fiñana M, Kemp G, Whitehead J, Jaki T, Khoo S, et al. Telmisartan and insulin resistance in hiv (tailor): protocol for a dose-ranging phase ii randomised open-labelled trial of telmisartan as a strategy for the reduction of insulin resistance in hiv-positive individuals on combination antiretroviral therapy. *BMJ Open*. 2015;5(10):.
108. Pushpakom S, Kolamunnage-Dona R, Taylor C, Foster T, Spowart C, García-Fiñana M, Kemp G, Jaki T, Khoo S, Williamson P, Pirmohamed M, for the TAILOR Study Group. TAILOR (Telmisartan and Insulin Resistance in Human Immunodeficiency Virus [HIV]): an adaptive-design, dose-ranging phase IIb randomized trial of telmisartan for the reduction of insulin resistance in HIV-positive individuals on combination antiretroviral therapy. *Clin Infect Dis*. 2019. <https://doi.org/10.1093/cid/ciz589>.
109. Dumville J, Hahn S, Miles J, Torgerson D. The use of unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials*. 2006;27(1):1–12.
110. Meurer W, Lewis R, Berry D. Adaptive clinical trials: a partial remedy for the therapeutic misconception? *JAMA*. 2012;307(22):2377–8.
111. Parmar M, Barthel F-S, Sydes M, Langley R, Kaplan R, Eisenhauer E, Brady M, James N, Bookman M, Swart A-M, et al. Speeding up the evaluation of new agents in cancer. *J Natl Cancer Inst*. 2008;100(17):1204–14.
112. Thall P, Simon R, Ellenberg S. A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics*. 1989;45(2):537–47.
113. Sampson A, Sill M. Drop-the-losers design: normal case. *Biom J*. 2005;47(3):257–68.
114. Wason J, Stallard N, Bowden J, Jennison C. A multi-stage drop-the-losers design for multi-arm clinical trials. *Stat Methods Med Res*. 2017;26(1):508–24.
115. Zádori N, Gede N, Antal J, Szentesi A, Alizadeh H, Vincze A, Izbéki F, Papp M, Czákó L, Varga M, et al. Early elimination of fatty acids in hypertriglyceridemia-induced acute pancreatitis (ELEFANT trial): Protocol of an open-label, multicenter, adaptive randomized clinical trial. *Pancreatology*. 2019;20(3):369–76.
116. Thompson W. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*. 1933;25(3/4):285–94.

117. Berry D, Eick S. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Stat Med*. 1995;14(3):231–46.
118. Hu F, Rosenberger W. The theory of response-adaptive randomization in clinical trials. Hoboken, NJ: John Wiley & Sons; 2006.
119. Wason J, Trippa L. A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Stat Med*. 2014;33(13):2206–21.
120. Lin J, Bunn V. Comparison of multi-arm multi-stage design and adaptive randomization in platform clinical trials. *Contemp Clin Trials*. 2017;54:48–59.
121. Lotze T, Loecher M. Bandit: functions for simple A/B split test and multi-armed bandit analysis. 2015. R package version 0.5.0.
122. Giles F, Kantarjian H, Cortes J, Garcia-Manero G, Verstovsek S, Faderl S, Thomas D, Ferrajoli A, O'Brien S, Wathen J, et al. Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia. *J Clin Oncol*. 2003;21(9):1722–7.
123. Grieve A. Response-adaptive clinical trials: case studies in the medical literature. *Pharm Stat*. 2017;16(1):64–86.
124. Villar S, Bowden J, Wason J. Response-adaptive designs for binary responses: how to offer patient benefit while being robust to time trends? *Pharm Stat*. 2018;17(2):182–97.
125. Gutjahr G, Posch M, Brannath W. Familywise error control in multi-armed response-adaptive two-stage designs. *J Biopharm Stat*. 2011;21(4):818–30.
126. Robertson D, Wason J. Familywise error control in multi-armed response-adaptive trials. *Biometrics*. 2019;75(3):885–94.
127. London A. Learning health systems, clinical equipoise and the ethics of response adaptive randomisation. *J Med Ethics*. 2018;44(6):409–15.
128. Tehranisa J, Meurer W. Can response-adaptive randomization increase participation in acute stroke trials? *Stroke*. 2014;45(7):2131–3.
129. Hu F, Rosenberger W. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *J Am Stat Assoc*. 2003;98(463):671–8.
130. Berry D. Adaptive clinical trials: the promise and the caution. *J Clin Oncol*. 2010;29(6):606–9.
131. Proschan M, Evans S. The temptation of response-adaptive randomization. *Clin Infect Dis*. 2020.
132. Korn E, Freidlin B. Outcome-adaptive randomization: is it useful? *J Clin Oncol*. 2011;29(6):771.
133. Viele K, Broglio K, McGlothlin A, Saville B. Comparison of methods for control allocation in multiple arm studies using response adaptive randomization. *Clin Trials*. 2020;17(1):52–60.
134. Smith A, Villar S. Bayesian adaptive bandit-based designs using the gittins index for multi-armed trials with normally distributed endpoints. *J Appl Stat*. 2018;45(6):1052–76.
135. Bowden J, Trippa L. Unbiased estimation for response adaptive clinical trials. *Stat Methods Med Res*. 2017;26(5):2376–88.
136. Simon R, Simon N. Using randomization tests to preserve type I error with response adaptive and covariate adaptive randomization. *Stat Probab Lett*. 2011;81(7):767–72.
137. Xub X, Bretz F. Handbook of methods for designing, monitoring, and analyzing dose-finding trials. In: O'Quigley J, Iasonos A, Bornkamp B, editors. Boca Raton: CRC Press, Taylor and Francis Group; 2017. p. 205–27. Chap. 12.
138. Bretz F, Pinheiro J, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*. 2005;61(3):738–48.
139. Pinheiro J, Bornkamp B, Glimm E, Bretz F. Model-based dose finding under model uncertainty using general parametric models. *Stat Med*. 2014;33(10):1646–61.
140. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel. 1986;81. Taylor & Francis.
141. Bornkamp B, Pinheiro J, Bretz F, et al. Mcpmod: An R package for the design and analysis of dose-finding studies. *J Stat Softw*. 2009;29(7):1–23.
142. Verrier D, Sivapregassam S, Solente A-C. Dose-finding studies, mcp-mod, model selection, and model averaging: Two applications in the real world. *Clin Trials*. 2014;11(4):476–84.
143. Bornkamp B, Bretz F, Pinheiro J. Request for CHMP qualification opinion. 2013. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/02/WC500161026.pdf. Accessed 26 Apr 2020.
144. European Medicines Agency. Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. 2014.
145. Drug Development Tools: Fit-for-Purpose Initiative. <https://www.fda.gov/drugs/development-approval-process-drugs/drug-development-tools-fit-purpose-initiative>. Accessed 26 Apr 2020.
146. Bornkamp B. DoseFinding: planning and analyzing dose finding experiments. 2019. R package version 0.9-17. <https://CRAN.R-project.org/package=DoseFinding>.
147. Food and Drug Administration. FDA qualification of mcp-mod method. 2015.
148. Angus D, Alexander B, Berry S, Buxton M, Lewis R, Paoloni M, Webb S, Arnold S, Barker A, Berry D, et al. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov*. 2019;18(10):797.
149. Antoniou M, Jorgensen A, Kolamunnage-Dona R. Biomarker-guided adaptive trial designs in phase ii and phase iii: a methodological review. *PLoS one*. 2016;11(2):0149803.
150. Pocock S, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31(1):103–15.
151. Taves D. Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther*. 1974;15(5):443–53.
152. Altman D, Bland J. Treatment allocation by minimisation. *Bmj*. 2005;330(7495):843.
153. Scott N, McPherson G, Ramsay C, Campbell M. The method of minimization for allocation to clinical trials: a review. *Control Clin Trials*. 2002;23(6):662–74.
154. Atkinson A. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*. 1982;69(1):61–7.
155. Rosenberger W, Vidyashankar A, Agarwal D. Covariate-adjusted response-adaptive designs for binary response. *J Biopharm Stat*. 2001;11(4):227–36.
156. Kim E, Herbst R, Wistuba I, Lee J, Blumenschein G, Tsao A, Stewart D, Hicks M, Erasmus J, Gupta S, et al. The battle trial: personalizing therapy for lung cancer. *Cancer Discov*. 2011;1(1):44–53.
157. Zhou X, Liu S, Kim E, Herbst R, Lee J. Bayesian adaptive design for targeted therapy development in lung cancer: a step toward personalized medicine. *Clin Trials*. 2008;5(3):181–93.
158. Marcus R, Peritz E, Gabriel K. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655–60.
159. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A. Confirmatory adaptive designs with bayesian decision tools for a targeted therapy in oncology. *Stat Med*. 2009;28(10):1445–63.
160. Burnett T. Bayesian decision making in adaptive clinical trials. PhD thesis, University of Bath. 2017.
161. Ondra T, Jobjörnsson S, Beckman R, Burman C-F, König F, Stallard N, Posch M. Optimized adaptive enrichment designs. *Stat Methods Med Res*. 2019;28(7):2096–111.
162. Götte H, Donica M, Mordenti G. Improving probabilities of correct interim decision in population enrichment designs. *J Biopharm Stat*. 2015;25(5):1020–38.
163. Magnusson B, Turnbull B. Group sequential enrichment design incorporating subgroup selection. *Stat Med*. 2013;32(16):2695–714.
164. Jones R, Attia S, Mehta C, Liu L, Sankhala K, Robinson S, Ravi V, Penel N, Stacchiotti S, Tap W, et al. Tappas: an adaptive enrichment phase 3 trial of trc105 and pazopanib versus pazopanib alone in patients with advanced angiosarcoma. *J Clin Oncol*. 2017;35:.
165. Mehta C, Liu L, Theuer C. An adaptive population enrichment phase iii trial of trc105 and pazopanib versus pazopanib alone in patients with advanced angiosarcoma (tappas trial). *Ann Oncol*. 2019;30(1):103–8.
166. Mehta C, Pocock S. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med*. 2011;30(28):3267–84.
167. Wan F, Titman A, Jaki T. Subgroup analysis of treatment effects for misclassified biomarkers with time-to-event data. *J R Stat Soc: Ser C: Appl Stat*. 2019;68(5):1447–63.
168. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res*. 2005;11(21):7872–8.

169. Bhattacharyya A, Rai S. Adaptive signature design-review of the biomarker guided adaptive phase-iii controlled design. *Contemp Clin Trials Commun.* 2019;15:100378.
170. Chen J, Lu T-P, Chen D-T, Wang S-J. Biomarker adaptive designs in clinical trials. *Transl Cancer Res.* 2014;3(3):279–92.
171. Wason J, Marshall A, Dunn J, Stein R, Stallard N. Adaptive designs for clinical trials assessing biomarker-guided treatment strategies. *Br J Cancer.* 2014;110(8):1950–7.
172. Park J, Siden E, Zoratti M, Dron L, Harari O, Singer J, Lester R, Thorlund K, Mills E. Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. *Trials.* 2019;20(1):1–10.
173. Cunanan K, Iasonos A, Shen R, Begg C, Gönen M. An efficient basket trial design. *Stat Med.* 2017;36(10):1568–79.
174. Hatfield I, Allison A, Flight L, Julious S, Dimairo M. Adaptive designs undertaken in clinical research: a review of registered clinical trials. *Trials.* 2016;17(1):150.
175. O'Brien P, Fleming T. A multiple testing procedure for clinical trials. *Biometrics.* 1979;35(3):549–56.
176. Gordon Lan K, DeMets D. Discrete sequential boundaries for clinical trials. *Biometrika.* 1983;70(3):659–63.
177. Whitehead J. The design and analysis of sequential clinical trials. Hoboken, NJ: John Wiley & Sons; 1997.
178. Jennison C, Turnbull B. Group sequential methods with applications to clinical trials. Raton, FL: Chapman and Hall/CRC; 1999.
179. Whitehead J. Group sequential trials revisited: simple implementation using sas. *Stat Methods Med Res.* 2011;20(6):635–56.
180. Anderson K. gsDesign: an R Package for designing group sequential clinical trials. 2009. Version 2.0 Manual.
181. Wason J. Optgs: an r package for finding near-optimal group-sequential designs. *J Stat Softw.* 2013.
182. Boden W, van Gilst W, Scheldewaert R, Starkey I, Carlier M, Julian D, Whitehead A, Bertrand M, Col J, Pedersen O, et al. Diltiazem in acute myocardial infarction treated with thrombolytic agents: a randomised placebo-controlled trial. *Lancet.* 2000;355(9217):1751–6.
183. Boden W, Scheldewaert R, Walters E, Whitehead A, Coltart D, Santoni J-P, Belgrave G. Design of a placebo-controlled clinical trial of long-acting diltiazem and aspirin versus aspirin alone in patients receiving thrombolysis with a first acute myocardial infarction. *Am J Cardiol.* 1995;75(16):1120–3.
184. Prochan M. Sample size re-estimation in clinical trials. *Biom J.* 2009;51(2):348–57.
185. Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review. *Biom J.* 2006;48(4):537–55.
186. Chuang-Stein C, Anderson K, Gallo P, Collins S. Sample size reestimation: a review and recommendations. *Drug Inf J.* 2006;40(4):475–84.
187. Wang S-J, James Hung H, O'Neill R. Paradigms for adaptive statistical information designs: practical experiences and strategies. *Stat Med.* 2012;31(25):3011–23.
188. Pritchett Y, Menon S, Marchenko O, Antonijevic Z, Miller E, Sanchez-Kam M, Morgan-Bouniol C, Nguyen H, Prucka W. Sample size re-estimation designs in confirmatory clinical trials-current state, statistical considerations, and practical guidance. *Stat Biopharm Res.* 2015;7(4):309–21.
189. Spiegelhalter D, Abrams K, Myles J. Bayesian approaches to clinical trials and health-care evaluation. Hoboken, NJ: John Wiley & Sons; 2004.
190. Lachin J. A review of methods for futility stopping based on conditional power. *Stat Med.* 2005;24(18):2747–64.
191. Broglio K, Connor J, Berry S. Not too big, not too small: a goldilocks approach to sample size selection. *J Biopharm Stat.* 2014;24(3):685–705.
192. Zucker D, Wittes J, Schabenberger O, Brittain E. Internal pilot studies ii: comparison of various procedures. *Stat Med.* 1999;18(24):3493–509.
193. Kieser M, Friede T. Simple procedures for blinded sample size adjustment that do not affect the type i error rate. *Stat Med.* 2003;22(23):3571–81.
194. Graf A, Bauer P, Glimm E, Koenig F. Maximum type 1 error rate inflation in multiarmed clinical trials with adaptive interim sample size modifications: Maximum type 1 error inflation. *Biom J.* 2014;56(4):614–30.
195. Hade E, Young G, Love R. Follow up after sample size re-estimation in a breast cancer randomized trial for disease-free survival. *Trials.* 2019;20(1):527.
196. Gould A. Sample size re-estimation: recent developments and practical considerations. *Stat Med.* 2001;20(17):2625–43.
197. Wason J, Brocklehurst P, Yap C. When to keep it simple—adaptive designs are not always useful. *BMC Med.* 2019;17(1):1–7.
198. Committee for Medicinal Products for Human Use (CHMP). Reflection paper on methodological issues in confirmatory clinical trials with an adaptive design. London: European Medicines Agency; 2007.
199. Adopting an adaptive approach to help HIV patients on antiretroviral therapy (cART). <https://www.nihr.ac.uk/documents/case-studies/adopting-an-adaptive-approach-to-help-hiv-patients-on-antiretroviral-therapy-cart/22259>. Accessed 28 Jan 2020.
200. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious S, Mander A, Weir C, Koenig F, Walton M, et al. The adaptive designs consort extension (ace) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ.* 2019;369.
201. Jennison C, Turnbull B. Confirmatory seamless phase ii/iii clinical trials with hypotheses selection at interim: opportunities and limitations. *Biom J.* 2006;48(4):650–5.
202. Cuffe R, Lawrence D, Stone A, Vandemeulebroecke M. When is a seamless study desirable? case studies from different pharmaceutical sponsors. *Pharm Stat.* 2014;13(4):229–37.
203. Graham E, Jaki T, Harbron C. A comparison of stochastic programming methods for portfolio level decision-making. *J Biopharm Stat.* 2019;13:1–25.
204. Wassmer G, Pahlke F. Rpac: confirmatory adaptive clinical trial design and analysis. 2019. R package version 2.0.6.
205. Sanchez-Kam M, Gallo P, Loewy J, Menon S, Antonijevic Z, Christensen J, Chuang-Stein C, Laage T. A practical guide to data monitoring committees in adaptive trials. *Ther Innov Regul Sci.* 2014;48(3):316–26.
206. Chow S-C, Corey R, Lin M. On the independence of data monitoring committee in adaptive design clinical trials. *Journal of biopharmaceutical statistics.* 2012;22(4):853–67.
207. A Practical Adaptive & Novel Designs and Analysis (PANDA) Toolkit. <https://www.sheffield.ac.uk/scharr/sections/dts/ctru/panda>. Accessed 26 Apr 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

