

Current Issues in Clinical Trials Utilizing Master Protocol Design

Grace Zhou*, Shein-Chung Chow

Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Durham, North Carolina

***Corresponding Author:** Grace Zhou, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Durham, North Carolina.

Received: 18 February 2023; **Accepted:** 24 February 2023; **Published:** 28 February 2023.

Citation: Grace Zhou, Shein-Chung Chow (2023), Current Issues in Clinical Trials Utilizing Master Protocol Design, *Clinical Trials and Bioavailability Research*. 2(1). DOI: 10.58489/2836-5836/008.

Copyright: © 2023 Grace Zhou, this is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In recent years, the use of master protocol design in clinical trials has received much attention. Master protocol design, which includes basket, umbrella, and platform design, has the flexibility of testing multiple therapies in one indication, one therapy for multiple indications or both. In this manner, master protocol design can expedite clinical development of drug products under investigation in a more efficient way. In practice, however, the integrity, quality, and validity of clinical studies utilizing master protocol trial design is a concern due to various issues such as heterogeneity across target patient populations, multiplicity in hypotheses testing, and population drift when using non-concurrent control subjects. In this article, current statistical issues in clinical trials utilizing master protocol design are described and potential solutions are provided. These issues include sample size and power determination, controlling type I error rate, multiplicity, pooled analysis, treatment imbalance/heterogeneity, and population drift. Various approaches have been proposed to control and/or eliminate these issues where many follow Bayesian-based approaches. These practical issues are illustrated through some case studies that demonstrate the implementation of master protocol design within the field of oncology where it is most used and in other therapeutic areas. Some recommendations for resolutions of the practical issues encountered in the case studies are provided.

Keywords: *Basket Trial; Umbrella Trial, Platform Trial; Sample Size; Power; Type I Error Rate; Pooled Analysis; Heterogeneity; Multiplicity; Population Drift*

Introduction

Woodcock and LaVange (2017) originally proposed the concept of master protocol for testing multiple therapies in one indication, one therapy for multiple indications or both to expedite drug/clinical development of drug products under investigation. A master protocol design is defined as a protocol designed with multiple substudies, which may have different objectives and involves coordinated efforts to evaluate one or more investigational drugs in one or more disease subtypes within the overall trial structure. Woodcock and LaVange (2017) indicated that a master protocol may be used to conduct the trial(s) for exploratory purposes or to support a marketing application and can be structured to evaluate different drugs compared to their respective controls or to a single common control. In practice, it is suggested that the sponsor can design the master protocol with a fixed or adaptive design with the intent

to modify the protocol to incorporate or terminate individual substudies within the master protocol. Individual drug substudies under the master protocol can incorporate an initial dose-finding phase. As an example, adult data, when sufficient, can inform a starting dose for the investigational drug in pediatric patients as long as the drug provides the prospect of direct clinical benefit to pediatric patients.

Under the broad definition for master protocol, a master protocol consists of three distinct entities: basket, umbrella, and platform trials, which are defined in Section 2. A master protocol may involve direct comparisons of competing therapies or be structured to evaluate, in parallel, different therapies relative to their respective controls. Some take advantage of existing infrastructure to capitalize on similarities among trials, whereas others involve setting up a new trial network specific to the master protocol. All require extensive pretrial discussion

Clinical Trials and Bioavailability Research

among sponsors contributing therapies for evaluation and parties involved in the conduct and governance of the trials to ensure that issues surrounding data use, publication rights, and the timing of regulatory submissions are addressed and resolved before the start of the trial.

In practice, each of the three distinct trial designs that utilize master protocol have their own statistical considerations that arise from the trial's objective, and these considerations also vary within each distinct entity as the trial designs can vary widely even within each entity. Basket trials commonly face issues with heterogeneity between different subgroups, which can lead to an inflated false positive rate when exchanging information across subgroups. Umbrella trials, on the other hand, commonly face multiplicity issues as multiple treatment effects are tested in the same study, leading to a greater likelihood of a false positive. Lastly, platform trials face multiplicity concerns similar to umbrella trials, but they also face statistical considerations that arise from population drift. Due to the complexity of master protocol trial designs, these statistical considerations need to be considered for an accurate and reliable assessment of the treatment effect under study.

The purpose of this article is not only to provide a review of master protocol trial design but also to outline current issues and potential challenges from multiple perspectives. In the next section, types of master protocols are briefly introduced. Regulatory perspectives to inform master protocol trial design are discussed in Section 3. Section 4 provides current issues and challenges regarding statistical validity and efficiency from statistical perspectives. Case studies of master protocol trial design in oncology drug development and other therapeutic areas are discussed in Section 5. Concluding remarks are given in the last section of this article.

Master Protocol Trial Design

Classical clinical trials investigate one therapy in the context of one population group while master protocol clinical trials investigate either multiple therapies or multiple populations at the same time in one study. In what follows, commonly considered master protocol trial designs are briefly described.

Basket Trial Design

As indicated by Woodcock and LaVange (2017), basket trials aim to study a single targeted therapy in the context of multiple diseases or disease subtypes. As a result, basket trials feature a single targeted agent in multiple patient populations across different indications.

Basket trials are commonly used in oncology studies. They have become more popular as the oncology drug development landscape increasingly aims to create therapies that are designed to improve outcomes for patients with cancers that harbor specific molecular aberrations rather than by specific tumour types. In this manner, basket trials have emerged as an approach to test the hypothesis that targeted therapies may be effective independent of tumour histology, as long as the molecular target is present. Basket trials offer the advantage of lowering the required sample size by increasing the power of the analyses through borrowing information across the different subgroups. However, a key scientific assumption for this kind of pooled analysis is that the subgroups are homogeneous and therefore exchangeable.

Umbrella Trial Design

In contrast to basket trials, Woodcock and Lavange (2017) defined umbrella trials as the study of multiple targeted therapies in the context of a single disease, dividing patients into multiple parallel treatment arms and often sharing one control. An umbrella trial with common control has the potential of reducing the total number of patients allocated to the control, making it more appealing to patients and drug developers. However, because umbrella trials involve multiple investigational drugs, they require collaboration among multiple stakeholders, which has proven to be a significant challenge.

Ouma and others (2022) found that most umbrella trials to date have been conducted in early phase settings and similar to basket trials, nearly exclusively in oncology. While new designs for the umbrella trial design have been studied very little, Simon (2017) demonstrated the potential for umbrella trials to be designed as integrated phase II/III trials through the implementation of enrichment designs.

Platform Trial Design

Platform trials, as characterized by Woodcock and LaVange (2017), also study multiple targeted therapies in the context of a single disease but are designed to continue in a perpetual manner, with therapies allowed to enter or leave the platform on the basis of a predetermined decision algorithm. Because platform trials are an ongoing process rather than following a predefined schedule, the study design can take advantage of adaptive strategies like response-adaptive randomization, using information generated early on in the study to determine and adjust how the study continues.

While some platform trials have been deemed to be

successful like BATTLE and I-SPY2 trials, featuring adaptive randomization and the use of Bayesian hierarchical models to evaluate treatment effects, statistical methodology research on the platform trial design remains sparse. While some new Bayesian modeling and adaptive randomization procedures have been proposed (Hobbs et al. 2016, Kaizer et al. 2018), there continues to be widespread debate and disagreement on not only problems specific to platform trials but also on response-adaptive randomization. In addition, platform trials face issues with controlling the Type I error rate as a result of multiplicity concerns.

While platform trials still primarily exist in the oncology setting, they have become used increasingly in the non-oncology setting as well. These include the REMAP-CAP trial, which studied community-acquired pneumonia, and DIAN-TU NexGen, which studied treatments for Alzheimer. Therefore, we can see that platform trials have potential to be used across numerous therapeutic areas including bacterial and viral diseases as well as mental and neurological diseases.

A Comparison

In practice, well-designed and conducted master protocols can accelerate drug development by maximizing the amount of information obtained from the research effort. Utilizing a master protocol eliminates the need to develop new protocols for every trial, and instead new agents can be introduced through amendments to an already-approved protocol, avoiding the need for repeated review of all study procedures. Furthermore, compared with conducting separate stand-alone trials, conducting an umbrella or platform trial can increase data quality and efficiency through shared infrastructure and can reduce overall sample size through sharing of a control arm. Basket trials are also capable of reducing overall sample size by taking advantage of the ability to borrow information across population groups. This reduction in minimum sample size is a crucial advantage of using master protocol, especially as the medical field continues to push toward personalized medicine and the standard randomized clinical trial design no longer suffices. Major differences between the three subsets of clinical trials that utilize master protocol and the standard clinical trial are summarized in the following table.

For simple interpretation, let θ_1, θ_k be the treatment effects, denoting the difference in the mean outcome between treatment and control, in the first to k th

subgroups. For the basket trial example, we consider the example of a pooled analysis where the study design incorporates pruning of indications, such that of the k initial subgroups, only p subgroups remain in the final pooled analysis. Statistics surrounding applications and sample sizes across each type of trial utilizing master protocol was determined by Park and others (2019). The analogous characteristics for the standard 1:1 trial is based off of FDA regulations. Comparing the median duration and sample size between trials that utilize master protocol and those recommended by the FDA for classical trials, the potential for such trials to substantially decrease the minimum study length and sample size needed becomes clear.

Regulatory Perspectives

While the FDA has endorsed the use of master protocols, this does not cover first-in-human or early-stage clinical trials using expansion cohorts to expedite drug development. FDA recommends that for drugs evaluated in a master protocol, the recommended phase II dose (RP2D) should have already been established. Furthermore, because of the increased complexity surrounding master protocol designs, the FDA recommends sponsors engage them early in study planning. The FDA also offers specific design considerations in master protocols, including, but not limited to, the use of a single common control arm in randomized umbrella trial designs, the determination of biomarker-defined subgroup assignment, and the addition and/or discontinuation of treatment arms.

Early communication with the FDA allows for key agreements on the design and conduct of the protocol to be reached, avoiding potential challenges anticipated by the FDA, which include:

- (i) Difficulty attributing adverse events to one or more investigational drugs when multiple drugs are administered within each investigational treatment arm and the trial lacks a single internal control for those drugs
- (ii) Difficulty assessing the safety profile of any given investigational drug may when multiple drugs are being studied at the same time
- (iii) Potential overinterpretation in findings, resulting in delays in drug development

In addition to following the FDA guidance for master protocol designs, master protocol trials that incorporate adaptive designs must also provide all information described in the guidances for industry Adaptive Designs for Clinical Trials of Drugs

Table 1: Comparison of Master Protocol Designs with Classical Trial Design

Characteristic	Basket Trial	Umbrella Trial	Platform Trial	Classical Trial
Objective	To study a single targeted therapy in the context of the multiple diseases or disease subtypes	To study multiple targeted therapies in the context of a single disease	To study multiple targeted therapies in the context of a single disease in a perpetual manner (therapies are allowed to enter or leave the study)	To study a single therapy in the context of a single disease
Applications	Oncology, hereditary periodic fevers, complement-mediated disorders	Oncology	Oncology, influenza, pneumonia, pre-operative surgery, Alzheimer's disease, COVID-19	Found in all areas of clinical trials
Null Hypothesis	k $H_0: \cap \theta_i = 0$ $i=1$	k $H_0: \cap \theta_i = 0$ $i=1$	$H_0: \{H_{01}: \theta_1 = 0, H_{02}: \theta_2 = 0, \dots\}$	$H_0: \theta = 0$
Alternative Hypothesis	p $H_1: \cap \theta_i \neq 0$ $i=1$ such that $p < k$	k $H_1: \cup \theta_i \neq 0$ $i=1$	$H_1: \{H_{11}: \theta_1 \neq 0, H_{12}: \theta_2 \neq 0, \dots\}$	$H_1: \theta \neq 0$
Sample Size (Median [Q1, Q3])	205 [90, 500]	346 [252, 565]	892 [255, 1835]	Phase I: 20-100 Phase II: several hundred Phase III: 300-3000
Duration in	52.3 [42.9, 74.1]	60.9 [46.9, 81.3]	58.9 [36.9, 101.3]	Phase I: several

and Biologics and Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products.

Recently, the FDA has encouraged the use of master protocols in the setting of a public health emergency such as the current COVID-19 pandemic, where there is a critical need for the efficient development of therapies. FDA expects master protocols to continue to play an important role in addressing the public health needs created by the current COVID-19 pandemic, as well as in future pandemics that might occur.

Current Challenging Issues and Possible Solutions

Sample Size and Power

In a standard comparative clinical trial, comprehensive research has been conducted regarding minimum sample size and power computations. In most cases for a standard clinical trial, multiple closed form sample calculations have been found and the respective advantages and disadvantages of each method have been compared. However, no such closed form exists for the sample size calculation in master protocol trials. As such, most trials rely on extensive simulation-based methods to conduct the power calculation for sample size, but these simulation-based methods quickly become complicated and require large amounts of computing power. Furthermore, lack of

methodological research conducted in this area has led to many power calculations that don't take advantage of borrowing information from the different subtrials within a master protocol trial.

Basket Trials – Basket trials have the potential to increase power by taking advantage of information-borrowing between different subgroups. However, the most widely implemented method is to calculate the minimum sample size for each subgroup treated separately, assigning the sum of each subgroup's sample size as the total minimum sample size. This method is largely implemented because of its simplicity and similarity to sample size and power computations in standard clinical trials. However, this does not take advantage of the ability to borrow information across subgroups.

Potential solutions for sample size determination in basket trials have largely been found by taking on a Bayesian approach. Recently, Zheng and others (2022) have developed a new Bayesian sample size closed-form formula for basket trials that permit borrowing of information between commensurate subsets.

Umbrella Trials – Umbrella trials can save on sample size by sharing the same control arm across the many treatment arms. However, for umbrella trials, the largest issue is the lack of a universal definition of study power in a multi-arm setting. Bennett and Mander (2020) defined the overall study power as the

probability of detecting all treatments that are better than the control, but Ren (2021) defined the overall study power to be the probability of detecting at least one effective treatment when all arms are active. Furthermore, Ouma and others (2022) found in a systematic review of umbrella trials that it was impossible to ascertain how the sample size was determined in a majority of trials (55.3%). Among the trials that did report sample size computation, most determined the sample size required for each subtrial separately.

Most potential solutions for sample size determination for umbrella trials utilize a Bayesian hierarchical approach. Kang and others (2021) developed a two-stage umbrella Phase II design with effect size clustering and information-borrowing across multiple biomarker-treatment pairs using a Bayesian hierarchical clustered design. Zang and others (2019) also proposed a Bayesian hierarchical model combined with an adaptive design. However, current literature on statistical methods for umbrella trials remains sparse and no agreed upon approach is taken as some trials opt for a pooled approach while others opt for regarding the treatment arms as separate. Consequently, the use of simulation is often required.

Platform Trials – Platform trials share the same sample size advantage as umbrella trials by being able to share the same control arms across the many treatment arms. Platform trials also have the advantage of using various adaptive designs because of their ongoing nature. Many platform trials include adaptive patient allocation through response-adaptive randomization rules, which preferentially assigns patients to interventions that perform most favorably. However, platform trials vary incredibly as a result of the common practice of including adaptive design and randomization. Consequently, there is no closed-form expression for sample size or power calculations. Furthermore, due to the greater variability within platform trials, each new platform trial would be required to perform power calculation for sample size determination through extensive clinical trial simulations. In other words, sample size calculations for one platform trial aren't typically used to significantly inform sample size calculations for other platform trials.

Consequently, platform trials typically determine sample size and power as well as other parts of their statistical design through simulations of the trial across a wide range of scenarios. Tolles and others (2022) designed an adaptive platform trial for evaluating treatments in patients with life-threatening

hemorrhage and determined minimum sample size through simulations prior to trial implementation. Similarly, in the UPMC OPTIMISE-C19 trial, Huang et al. (2021) utilized a Bayesian adaptive design and response adaptive randomization to ensure ability to provide statistical significance despite variable sample size. Platform trials exclusively seem to determine sample size through simulations of their specific trial design.

Controlling Type I Error Rate

Each subset of clinical trials that utilize master protocol requires additional statistical considerations for controlling the overall Type I error rate. Basket trials largely need to consider controlling Type I error if pruning is included in the trial design. When basket trials prune indications that are doing poorly at interim endpoints, this can inflate the false positive rate for the pooled analysis after pruning. This error has been termed “random high bias”. The most common solution to controlling this type of bias has been to simply adjust the significance level α to control for Type I error.

Umbrella and platform trials, on the other hand, typically have greater reason to control for Type I error as a result of multiplicity concerns, which will be discussed further in the next section. Controlling the overall Type I error rate becomes difficult because there is no consensus on which Type I error rate should be controlled for (i.e., the family-wise error rate, the per-comparison error rate, etc.).

For multi-arm, response-adaptive, 2-stage designs, methods to control the family-wise Type I error rate were proposed by Gutjahr and others (2011). Robertson and Wason (2019) proposed procedures to control the family-wise Type I error rate in certain scenarios with response-adaptive randomization for normally distributed endpoints. Ghosh and others (2017) developed a frequentist approach for platform trials that guarantees strong control of the Type I error rate.

Multiplicity

Multiplicity concerns arise when multiple inferences are being tested simultaneously. Consequently, multiplicity concerns are largely present in umbrella and platform trials where multiple treatment effects are being compared. If each hypothesis is tested with the same significance level, testing more treatments increases the likelihood of encountering a false positive. For basket trials, on the other hand, if a pooled analysis is conducted, then there is no multiplicity concern present. However, if the subgroups are analyzed in isolation, then there would

Clinical Trails and Bioavailability Research

be multiplicity concerns present that could potentially increase the false positive error rate.

Conventionally, adjustments are made to the significance level when multiplicity concerns are present. However, Woodcock and LaVange (2017) note that as precision medicine focuses on smaller and smaller disease subtypes, traditional methods for multiplicity adjustment become impractical. Stallard and others (2019) have argued that since umbrella trials and platform trials compare treatment arms within subgroups, there is no need to adjust for multiplicity arising from the testing of multiple hypotheses. This is due to the view that the sub-studies, though run under a single protocol, are essentially independent trials with each subgroup receiving a different experimental treatment. As such, the efficacy of each treatment is only assessed once. However, Howard and others (2018), upon investigating multi-arm designs where the control arm is shared across multiple treatment arms, defined new Type I error rates of interest and determined which Type I error rates should be controlled for based on the nature of the hypothesis tested. Hlavin and others (2017) have also determined conditions under which the family-wise error rate can be controlled in multi-arm trials when performing many-to-one comparisons.

For basket trials where inferences are made among subgroups within treatment arms, Stallard and others (2019) propose making adjustments for multiplicity by using a Holm-Bonferroni correction or a method allowing for overlapping subgroups.

Pooled Analysis

Opportunities for pruning are created in basket trials when the study design includes interim surrogate endpoints. This mitigates the risk of ineffective indications being included in the final pool. However, for the remaining indications, the patients who contributed data at the interim analysis will also be the ones included in the final analysis. This may affect the false positive rate in pooled analysis because of the re-use of internal study information. After pruning, sample size readjustment is then required to maintain the power of the final pooled analysis, and the plan for the readjustment must be pre-specified.

Chen and others (2016) compared several sample size adjustment strategies and noted that the most aggressive strategy provided the best maintenance of reasonable power. This strategy consisted of increasing the numbers of patients in the remaining indications to keep the size of the final pooled analysis as originally planned. However, for pruning

to be effective in mitigating the risk of pooling, the bar for being included in the final pooled analysis is set relatively high. As such, pruning does not show that the indication is not worthy of further investigation, only that the indication is too high risk to remain in the basket.

Treatment Imbalance and Heterogeneity

Heterogeneity is also of concern in basket trials as the basket may contain a heterogeneous mixture of ineffective and effective indications where both move on to the final stage and the pooled analysis. In such a case, the ineffective indications included in the pooled analysis may dilute the overall pooled result, resulting in a false negative with respect to the effective indications. On the other hand, the effective indications could dominate the average result, leading to the pool of indications being approved and a false positive with respect to the ineffective indications. While a subgroup analysis for known predictive and prognostic factors is typically performed, these subgroup analyses are typically underpowered and provide only qualitative information.

Because basket trials aim to examine therapies in varying populations, a natural consequence of this trial design is that patients may respond differently to the same treatment due to their distinct disease subtypes. Regarding the subgroups in isolation takes into account the potential heterogeneity in the subgroups, but it treats the analysis of each subgroup completely separately. Consequently, this strategy doesn't realize the full potential of treating the combined subgroups as a single study in basket trials while still facing new concerns of multiplicity and potential low power issues due to small sample sizes. More sophisticated analysis models, which feature borrowing of information between subgroups, have been proposed. Thall (2003) and Berry (2013) proposed fitting a Bayesian hierarchical random-effects model under the assumption that subgroup-specific treatment effects are exchangeable. In other words, that the treatment effects for each subgroup are random samples drawn from the same normal distribution with unknown mean and variance. This methodology has since been extended to a finite mixture of exchangeability distributions (Liu et al. 2017; Chu and Yuan 2018; Jin et al. 2020) and non-exchangeability distributions (Neuenschwander et al. 2016) to avoid over-representing extreme treatment effects. A common theme among proposed basket trial designs is the use of a Bayesian framework to mitigate the effects of heterogeneity.

Population Drift

Population drift, also known as temporal drift, arises as a consequence of platform trials' ongoing nature and the use of one shared control arm. The issue revolves around determining how to manage comparisons between different treatment groups and the control. Because the platform trial design allows for treatments to enter the study after the start of the study, the control arm may include both concurrent subjects (control subjects added when the treatment arm is in the platform) and non-concurrent subjects (control subjects added before the treatment had been added to the platform). Consequently, the unadjusted treatment-control comparisons are likely to be biased due to trends over time like changing inclusion/exclusion criteria or changing standard of care. When population drift is present and not considered, it can also result in Type I error rate inflation or deflation as well as biased estimates.

The most intuitive solution is to limit the extent to which each control patient can influence analysis based on their time of enrollment. In other words, interim analyses at some trial time t would only include control patients who were enrolled within some pre-determined interval of time from t . This would limit the amount of population drift between control subjects included in each analysis but would also limit the number of control subjects included in each analysis. Overbey et al. (2022) explored several methods to incorporate nonconcurrent control subjects, including test-then-pool, fixed power prior, dynamic power prior, and multi-source exchangeability model approaches. However, they concluded that compared to using concurrent control subjects only, these approaches could not guarantee type I error control or unbiased estimates. Saville and others (2022) proposed a "Bayesian Time Machine" that models potential population drift and smooths estimates across time using a second-order normal dynamic linear model.

Cases Studies

Since master protocol trials were initially designed for oncology due to the prevalence of targeted therapies in cancer research, master protocols continue to still be almost exclusively used in oncology. Furthermore, most master protocol study designs are still based off early master protocol studies involved in cancer therapeutics like ISPY-2 (platform trial for breast cancer patients) and IMPACT-2 (basket trial that studied personalized medicine across different tumour types). However, as the shift towards precision medicine and personalized healthcare becomes increasingly popular, the use of master

protocols in clinical trials is becoming more prevalent in other therapeutic areas.

ROAR: Basket Trial Example

The Rare Oncology Agnostic Research (ROAR) trial was a Phase II, open-label, single-arm, multi-center basket trial that aimed to evaluate the activity and safety of dabrafenib and trametinib combination therapy in patients with BRAF-mutated biliary tract cancer. The study is currently still active and have obtained results from an interim analysis.

A total of 43 patients were enrolled to the study and evaluated. They had planned to enroll a maximum of 25 participants for every histological cohort in the primary analysis group. They determined the minimum sample size by evaluating the performance of the design through simulation studies that considered various assumptions for the distribution of the true overall response rates across the histological cohorts and accounted for anticipated small sample sizes due to low prevalence.

The design maintains 84-98% power and a Type I error rate of no more than 0.04 when treatment effects are similar across all histological cohorts. To increase the precision of the outcome estimates in the small sample size, a Bayesian hierarchical model that borrowed overall response rate information across histological cohorts was used. More borrowing was allowed if the response rates were similar. The Bayesian estimate of overall response was calculated from the primary analysis cohort and used the posterior mean and corresponding 95% credibility interval.

The study seemed to skillfully take advantage of the ability to borrow information across various histological cohorts in order to increase the precision of the overall response rate estimates, allowing for lower minimum sample size per cohort. This is particularly important in this study as it studies a rare disease. Such a strategy also allowed greater flexibility in the amount of information borrowed across certain subsets of cohorts. However, the study protocol did not include central confirmation of histology, which resulted in similar histological subtypes being reported with different names. This would affect the measured treatment effects in each cohort. Furthermore, while the study design was evaluated to maintain sufficient power and control of Type I error rate, it only holds true under the assumption that the treatment effects are similar across all histological cohorts, which may not hold. A more conservative approach may be to only rely on the assumption of exchangeability in a subset of

Clinical Trials and Bioavailability Research

histological cohorts, which could also help to control for potential heterogeneity across cohorts.

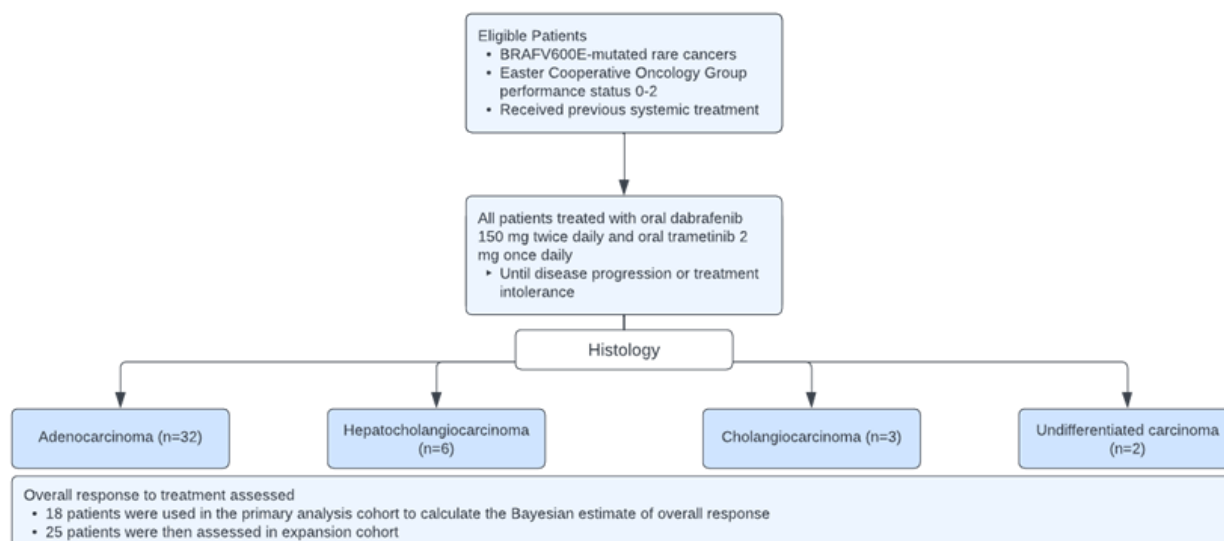


Fig 1: Trial Schema for ROAR1

[1] Figure was reproduced from figure found in “Dabrafenib plus trametinib in patients with BRAF V600E-mutated biliary tract cancer (ROAR): a phase 2, open-label, single-arm, multicentre basket trial” by V. Subbiah, U. Lassen, et al., 2020, *Lancet Oncology*, 21, p. 1234-1243, Copyright 2020 by Elsevier Ltd.

FOCUS4: Umbrella Trial Example

FOCUS4 is a multistage umbrella trial that sought to evaluate a number of treatments and biomarkers for advanced colorectal cancer. The proposal aimed to answer if targeted therapies provide activity signals in different biomarker-defined populations and if these definitively improve outcomes. The protocol randomized novel agents against placebo concurrently across a number of different biomarker-defined population-enriched cohorts, including but not limited to BRAF mutation, activated AKT pathways, and KRAS and NRAS mutations.

In this trial, each treatment is evaluated first in the cohort of patients for whom the biomarker is hypothesized to be predictive of response. Then, if appropriate, the hypothesis of the predictive ability of the biomarker is then tested by evaluating the agent in the biomarker-negative patients.

Sample size calculations were performed using the nstage program in STATA software published by Barthel and others (2009). For each of the subtrials within FOCUS4, the overall power is maintained at 80%, allowing for multiple interim data looks, with a maximum 5% Type I error rate. To maintain a power of 80%, the power of each subtrial for the primary analysis varies from 85-95%. Each biomarker-defined subtrial was considered separately. Furthermore, an

independent data monitoring committee can advise early closure of a trial in the event of overwhelming evidence of efficacy at a significance level of 0.001. The trials adopted this approach to preserve the overall Type I error rate at the end of the trial.

This trial proposed an efficient approach that allows for multiple treatments and multiple biomarkers to be evaluated while including measures to control Type I error rate and achieve sufficient power. In particular, the trial included adaptive design components such as early discontinuation due to either the futility rule or success criteria as well as both phase II and phase III components allowing treatments to move seamlessly between phases. However, the nstage program in STATA seems designed for multi-arm multi-stage (MAMS) studies, which feature a few differences when compared to this trial: (1) MAMS typically have 1 common control arm whereas FOCUS4 proposes a control arm for each different biomarker group, and (2) MAMS assumes randomized assignment to each of the treatment arms whereas FOCUS4 is biomarker-dependent and randomizes between placebo and treatment within each biomarker sub-population. Furthermore, this method of conducting sample size calculations does not take advantage of information borrowing across the biomarker sub-populations. A potential approach could be that proposed by Kang et al. (2021), which features information borrowing across multiple biomarker-treatment pairs using a Bayesian hierarchical design, or that proposed by Zang et al. (2019) that combines a Bayesian hierarchical model with an adaptive design.

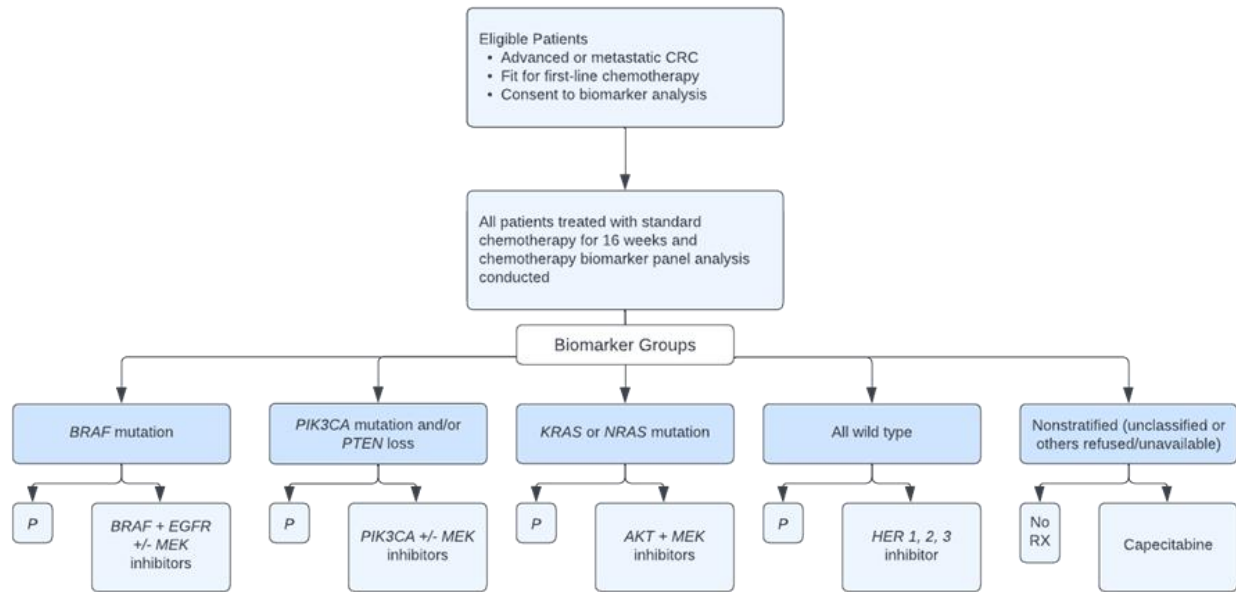


Fig 2: Trial Schema for FOCUS2

2 Figure was reproduced from figure found in “Evaluating many treatments and biomarkers in oncology: a new design” by R. Kaplan, T. Maughan, et al., 2013, *Journal of Clinical Oncology*, 31, p. 4562-4568, Copyright 2013 by the American Society of Clinical Oncology.

TICO MAMS is a global multi-arm, multi-stage platform master protocol trial that evaluates the safety and efficacy of anti-viral therapeutic agents for adults hospitalized with COVID-19. TICO planned to enroll 300 patients at Stage 1 with subsequent expansion to full sample size and an expanded target population in Stage 2 if the agent shows an acceptable safety profile and evidence of efficacy. The trials within TICO are randomized, double blinded, placebo-controlled and phase III. Each randomized subject could potentially receive any of the active agents or matching placebo for which they are eligible. The placebo group is then pooled such that those randomized to the placebo of one agent will be part of the control group for other agents to which the participant could have been allocated. This allows for a greater probability that a participant will receive an active agent when there are more agents in the platform.

In Stage 1, the planned sample size for each investigational agent and concurrently randomized pooled placebo arm is 300 patients under a 1:1 randomization, using proportional odds models and the estimated summary odds ratio to compare the investigational agent and the placebo. This sample size was determined to be sufficient to detect an OR of 1.60 or greater with 95% power. This was done with a relatively high Type I error rate of 0.30, which was based on previous work in MAMS cancer trials to

avoid premature futility declarations. The final sample size in Stage 2 will be event driven and was estimated at 1,000 patients. No adjustment will be made for the number of other agents being tested in TICO trials. Since the time of conception, TICO has studied 3 different investigational drugs.

TICO seems to be a successfully implemented platform trial, generating results for three novel agents so far with two other agents still under study and future agents ready to enter. Type I error rate was controlled using the Lan-DeMets spending function analogue of the O’Brien-Fleming Boundaries. Sample size considerations were made for both the initial futility assessment and the final assessment of efficacy with power aimed at 90% and an effect size of 25% increase in the sustained recovery rate for the treatment compared to the placebo. Furthermore, TICO was able to take advantage of the use of shared control arm for each of the different treatment arms, allowing for both a smaller minimum sample size requirement and a greater likelihood of being randomized into a treatment arm. To avoid the issue of population drift, the study chose to only share placebo across agents that were studied at the same time – in other words, only concurrent control subjects were used. This approach is reasonable as most proposed methods of including non-concurrent control subjects still fall short in controlling for the inflated Type I error rate. However, the “Bayesian Time Machine” proposed by Saville et al. (2022) models potential population drift using a second-order normal dynamic linear model, which may pose as a viable alternative that allows for the use of non-concurrent controls.

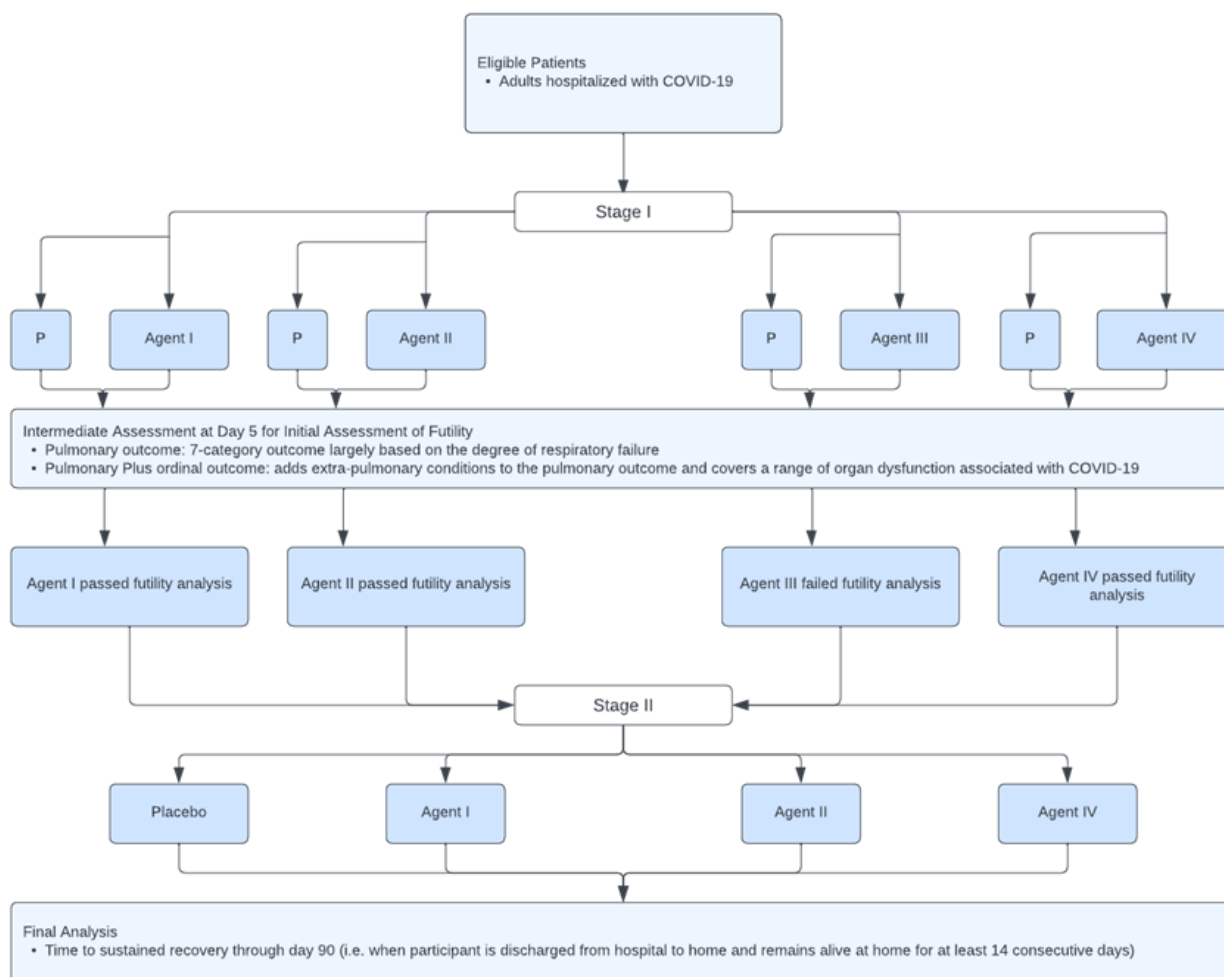


Fig 3: Agent Entry and Progression through TICO Study3

3 Figure was reproduced from figure found in “Design and implementation of an international, multi-arm, multi-stage platform master protocol for trials of novel SARS-CoV-2 antiviral agents: Therapeutics for Inpatients with COVID-19 (TICO/ACTIV-3)” by D. Murray, A. Babiker, et al., 2021, *Clinical Trials*, 19, p. 52-61, Copyright 2022 by SAGE Publications.

Concluding Remarks

As indicated in the previous sections, the use of master protocol design in clinical trials have the ability to investigate either multiple therapies or multiple populations at the same time in one study. However, statistical methodologies for data analysis are not fully developed. In practice, it is suggested that possible solutions for addressing the challenging issues such as insufficient sample size, heterogeneity, multiplicity, possible population drift, and the control of overall type I error rate should be taken into consideration for an accurate and reliable assessment of the treatment effect under investigation.

In general, there is no one preferred approach which is superior to others when it comes to addressing these issues. Rather, the most appropriate approach

will likely depend on the study objectives, study design, and the nature of the research question asked. However, some common trends do seem to arise. In particular, many proposed solutions seem to take on a Bayesian approach such as using a Bayesian hierarchical model for either sample size determination or to control for potential heterogeneity between treatment groups.

Possible approaches range from more conservative to less conservative methods. More conservative methods rely on fewer assumptions that may not hold due to the issues discussed in this paper. Less conservative methods take greater advantage of the potential ability to share information across the multiple subtrials. While the less conservative methods have the potential to increase study power, bias and inflated Type I error will be exacerbated when their assumptions do not hold. Consequently, it is important to comprehensively assess the assumptions that are made, especially as they pertain to exchangeability of treatment groups.

In practice, for a more efficient and reliable assessment of the test treatment under investigation, it is also suggested that the concept of master

protocol design should be applied in conjunction with some complex innovative designs (CIDs) such as multi-stage (e.g., two-stage) seamless adaptive trial design, the innovative n-of-1 trial design, and Bayesian sequential adaptive design with appropriate selection of a prior. The use of the concept of master protocol in tandem with an innovative complex design cannot only provide an efficient, accurate, and reliable assessment of the test treatment under study, but also shorten the drug development process and increase the probability of success of the test treatment under investigation.

Conflict of interest

None.

References

- Bennett, M., & Mander, A. P. (2020). Designs for adding a treatment arm to an ongoing clinical trial. *Trials*, 21(1), 251.
- Chow, S. C. (2020). *Innovative Methods for Rare Disease Drug Development*. Chapman and Hall/CRC Press, Taylor & Francis, New York.
- FDA (2021). *Guidance for Industry – COVID-19: Master Protocols Evaluating Drugs and Biological Products for Treatment or Prevention*. The United States Food and Drug Administration, Silver Spring, Maryland.
- FDA (2018). *Guidance for Industry – Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics*. The United States Food and Drug Administration, Silver Spring, Maryland.
- Ouma, L. O., Wason, J. M. S., Zheng, H., Wilson, N., & Grayling, M. (2022). Design and analysis of umbrella trials: Where do we stand? *Frontiers in medicine*, 9, 1037439.
- Park, J. J. H., Siden, E., Zoratti, M. J. et al. (2019). Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. *Trials* 20, 572.
- Renfro, L. A., Sargent, D. J. (2017). Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocol: a review and examples. *Annals of Oncology*, 28(1).
- Woodcock, J. and LaVange, L.M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *N. Eng. J Med*, 377, 62-70.
- Ren, Y., Li, X., Chen, C. (2021). Statistical considerations of phase 3 umbrella trials allowing adding one treatment arm mid-trial. *Contemporary Clinical Trials*, 109.
- W. Brannath, G. Gutjahr, P. Bauer. (2012) Probabilistic Foundation of Confirmatory Adaptive Designs. *Journal of the American Statistical Association* 107:498, pages 824-832.
- Howard, D. R., Brown, J. M., Todd, S., & Gregory, W. M. (2018). Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Statistical methods in medical research*, 27(5), 1513–1530.
- Robertson, D. S., & Wason, J. M. S. (2019). Familywise error control in multi-armed response-adaptive trials. *Biometrics*, 75(3), 885–894.
- Hlavin, G., Hampson, L. V., & Koenig, F. (2017). Many-to-one comparisons after safety selection in multi-arm clinical trials. *PloS one*, 12(6), e0180131.
- Chen, C., Li, X., Yuan, S. (2016). Statistical Design and Considerations of a Phase 3 Basket Trial for Simultaneous Investigation of Multiple Tumor Types in One Study. *Statistics in Biopharmaceutical Research*, 8(3), 248-257.
- Zheng, H., Grayling, M. J., Mozgunov, P., et al. (2022). Bayesian sample size determination in basket trials borrowing information between subsets. *Biostatistics*.
- Tolles, J., Beiling, M., Schreiber, M. A., Del Junco, D. J., McMullan, J. T., Guyette, F. X., Wang, H., Jansen, J. O., Meurer, W. J., Mainali, S., Yadav, K., & Lewis, R. J. (2022). An adaptive platform trial for evaluating treatments in patients with life-threatening hemorrhage from traumatic injuries: Rationale and proposal. *Transfusion*, 62 Suppl 1, S231–S241.
- Huang, D. T., McCreary, E. K., Bariola, J. R. et al. (2021). The UPMC OPTIMISE-C19 (OPTimizing Treatment and Impact of Monoclonal antibodies through Evaluation for COVID-19) trial: a structured summary of a study protocol for an open-label, pragmatic, comparative effectiveness platform trial with response-adaptive randomization. *Trials*, 22(363).
- Berry, S. M., Petzold, E. A., Dull, P., Thielman, N. M., Cunningham, C. K., Corey, G. R., McClain, M. T., Hoover, D. L., Russell, J., Griffiss, J. M., & Woods, C. W. (2016). A response adaptive randomization platform trial for efficient evaluation of Ebola virus treatments: A model for pandemic response. *Clinical trials (London, England)*, 13(1), 22–30.
- Noor, N. M., Pett, S. L., Esmail, H., Crook, A. M., Vale, C. L., Sydes, M. R., & Parmar, M. K. B. (2020). Adaptive platform trials using multi-arm,

Clinical Trials and Bioavailability Research

- multi-stage protocols: getting fast answers in pandemic settings. *F1000Research*, 9, 1109.
20. Longini, I. M., Yang, Y., Fleming, T. R., Muñoz-Fontela, C., Wang, R., Ellenberg, S. S., Qian, G., Halloran, M. E., Nason, M., Gruttola, V., Mulangu, S., Huang, Y., Donnelly, C. A., & Henao Restrepo, A. M. (2022). A platform trial design for preventive vaccines against Marburg virus and other emerging infectious disease threats. *Clinical trials (London, England)*, 19(6), 647–654.
 21. Kang, D., S Coffey, C., J Smith, B., Yuan, Y., Shi, Q., & Yin, J. (2021). Hierarchical Bayesian clustering design of multiple biomarker subgroups (HCOMBS). *Statistics in medicine*, 40(12), 2893–2921.
 22. Zang, Y., Guo, B., Han, Y., Cao, S., & Zhang, C. (2019). A Bayesian adaptive marker-stratified design for molecularly targeted agents with customized hierarchical modeling. *Statistics in medicine*, 38(15), 2883–2896.
 23. Cunanan, K. M., Iasonos, A., Shen, R., Begg, C. B., & Gönen, M. (2017). An efficient basket trial design. *Statistics in medicine*, 36(10), 1568–1579.
 24. Krajewska, M., & Rauch, G. (2021). A new basket trial design based on clustering of homogeneous subpopulations. *Journal of biopharmaceutical statistics*, 31(4), 425–447.
 25. Meyer, E. L., Mesenbrink, P., Dunger-Baldauf, C., Fülle, H. J., Glimm, E., Li, Y., Posch, M., & König, F. (2020). The Evolution of Master Protocol Clinical Trial Designs: A Systematic Literature Review. *Clinical therapeutics*, 42(7), 1330–1360.
 26. Overbey, J. R., Cheung, Y. K., & Bagiella, E. (2022). Integrating non-concurrent controls in the analyses of late-entry experimental arms in multi-arm trials with a shared control group in the presence of parameter drift. *Contemporary clinical trials*, 123, 106972.
 27. Korn, E. L., & Freidlin, B. (2017). Adaptive Clinical Trials: Advantages and Disadvantages of Various Adaptive Design Elements. *Journal of the National Cancer Institute*, 109(6), dxj013.
 28. Kaplan, R., Maughan, T., Crook, A., Fisher, D., Wilson, R., Brown, L., & Parmar, M. (2013). Evaluating many treatments and biomarkers in oncology: a new design. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 31(36), 4562–4568.
 29. Barthel, F. M.-S., Royston, P., & Parmar, M. K. B. (2009). A Menu-Driven Facility for Sample-Size Calculation in Novel Multiarm, Multistage Randomized Controlled Trials with a Time-to-Event Outcome. *The Stata Journal*, 9(4), 505–523.
 30. Blenkinsop, A., & Choodari-Oskooei, B. (2019). Multiarm, multistage randomized controlled trials with stopping boundaries for efficacy and lack of benefit: An update to nstage. *The Stata Journal*, 19(4), 782–802.
 31. Murray, D. D., Babiker, A. G., Baker, J. V., Barkauskas, C. E., Brown, S. M., Chang, C. C., Davey, V. J., Gelijns, A. C., Ginde, A. A., Grund, B., Higgs, E., Hudson, F., Kan, V. L., Lane, H. C., Murray, T. A., Paredes, R., Parmar, M. K., Pett, S., Phillips, A. N., Polizzotto, M. N., ... Lundgren, J. D. (2022). Design and implementation of an international, multi-arm, multi-stage platform master protocol for trials of novel SARS-CoV-2 antiviral agents: Therapeutics for Inpatients with COVID-19 (TICO/ACTIV-3). *Clinical trials (London, England)*, 19(1), 52–61.
 32. Subbiah, V., Lassen, U., Élez, E., Italiano, A., Curigliano, G., Javle, M., ... & Wainberg, Z. A. (2020). Dabrafenib plus trametinib in patients with BRAFV600E-mutated biliary tract cancer (ROAR): A phase 2, open-label, single-arm, multicentre basket trial. *The Lancet Oncology*, 21(9), 1234-1243.
 33. Liu, M., Li, Q., Lin, J., Lin, Y., & Hoffman, E. (2021). Innovative trial designs and analyses for vaccine clinical development. *Contemporary clinical trials*, 100, 106225.
 34. Said, R., & Tsimberidou, A. M. (2019). Basket Trials and the MD Anderson Precision Medicine Clinical Trials Platform. *Cancer journal (Sudbury, Mass.)*, 25(4), 282–286.
 35. Park, J. J. H., Hsu, G., Siden, E. G., Thorlund, K., & Mills, E. J. (2020). An overview of precision oncology basket and umbrella trials for clinicians. *CA: a cancer journal for clinicians*, 70(2), 125–137.
 36. Lu, C. C., Li, X. N., Broglio, K., Bycott, P., Jiang, Q., Li, X., McGlothlin, A., Tian, H., & Ye, J. (2021). Practical Considerations and Recommendations for Master Protocol Framework: Basket, Umbrella and Platform Trials. *Therapeutic innovation & regulatory science*, 55(6), 1145–1154.
 37. Kaizer, A. M., Koopmeiners, J. S., Kane, M. J., Roychoudhury, S., Hong, D. S., & Hobbs, B. P. (2019). Basket Designs: Statistical Considerations for Oncology Trials. *JCO precision oncology*, 3, 1–9.
 38. Hirakawa, A., Asano, J., Sato, H., & Teramukai,

- S. (2018). Master protocol trials in oncology: Review and new trial designs. *Contemporary clinical trials communications*, 12, 1–8.
39. Kaizer, A. M., Hobbs, B. P., & Koopmeiners, J. S. (2018). A multi-source adaptive platform design for testing sequential combinatorial therapeutic strategies. *Biometrics*, 74(3), 1082–1094.
 40. Saville, B. R., Berry, D. A., Berry, N. S., Viele, K., & Berry, S. M. (2022). The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials. *Clinical trials (London, England)*, 19(5), 490–501.
 41. Simon, R. (2017), Critical Review of Umbrella, Basket, and Platform Designs for Oncology Clinical Trials. *Clin. Pharmacol. Ther.*, 102, 934-941.
 42. Hobbs B. P., Chen N., Lee J. J. (2016). Controlled multi-arm platform design using predictive probability. *Statistical Methods in Medical Research*, 27(1), 65-78.
 43. Ghosh, P., Liu, L., Senchaudhuri, P., Gao, P., & Mehta, C. (2017). Design and monitoring of multi-arm multi-stage clinical trials. *Biometrics*, 73(4), 1289–1299.
 44. Stallard, N., Todd, S., Parashar, D., Kimani, P. K., & Renfro, L. A. (2019). On the need to adjust for multiplicity in confirmatory clinical trials with master protocols. *Annals of oncology : official journal of the European Society for Medical Oncology*, 30(4), 506–509.
 45. Thall, P. F., Wathen, J. K., Bekele, B. N., Champalin, R. E., Baker, L. H., & Benjamin, R. S. (2003). Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Statistics in medicine*, 22(5), 763–780.
 46. Berry, S. M., Broglio, K. R., Groshen, S., & Berry, D. A. (2013). Bayesian hierarchical modeling of patient subpopulations: efficient designs of Phase II oncology clinical trials. *Clinical trials (London, England)*, 10(5), 720–734.
 47. Liu, R., Liu, Z., Ghadessi, M., & Vonk, R. (2017). Increasing the efficiency of oncology basket trials using a Bayesian approach. *Contemporary clinical trials*, 63, 67–72.
 48. Chu, Y. and Yuan, Y. (2018). BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity. *Journal of the Royal Statistical Society Series C* 67, 723–740.
 49. Jin, J., Liu, Q., Zheng, W., Lin, T., Gao, L. and Dong, Y. (2020). A Bayesian method for the detection of proof of concept in early phase oncology studies with a basket design. *Statistics in Biosciences* 12, 167–179.
 50. Neuenschwander, B., Wandel, S., Roychoudhury, S., & Bailey, S. (2016). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical statistics*, 15(2), 123–134.
 51. Chen, N., & Lee, J. J. (2020). Bayesian cluster hierarchical model for subgroup borrowing in the design and analysis of basket trials with binary endpoints. *Statistical methods in medical research*, 29(9), 2717–2732.