

Review

Bayesian Methods for Information Borrowing in Basket Trials: An Overview

Tianjian Zhou ^{1,*}  and Yuan Ji ^{2,*} ¹ Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA² Department of Public Health Sciences, University of Chicago, Chicago, IL 60637, USA

* Correspondence: tianjian.zhou@colostate.edu (T.Z.); yji@health.bsd.uchicago.edu (Y.J.)

Simple Summary: This paper provides a review of statistical methods for tumor-agnostic clinical trials. In particular, the review focuses on basket trials and provides methodological insights into various Bayesian approaches. The key concept of borrowing information through Bayesian hierarchical models is emphasized, and some novel trial designs are introduced. The review is expected to provide oncology and biostatistics researchers with more exposure to powerful Bayesian methods for the design and analysis of tumor-agnostic clinical trials.

Abstract: Basket trials allow simultaneous evaluation of a single therapy across multiple cancer types or subtypes of the same cancer. Since the same treatment is tested across all baskets, it may be desirable to borrow information across them to improve the statistical precision and power in estimating and detecting the treatment effects in different baskets. We review recent developments in Bayesian methods for the design and analysis of basket trials, focusing on the mechanism of information borrowing. We explain the common components of these methods, such as a prior model for the treatment effects that embodies an assumption of exchangeability. We also discuss the distinct features of these methods that lead to different degrees of borrowing. Through simulation studies, we demonstrate the impact of information borrowing on the operating characteristics of these methods and discuss its broader implications for drug development. Examples of basket trials are presented in both phase I and phase II settings.

Keywords: basket trial; Bayesian method; borrow information; drug development; hierarchical model; oncology; shrinkage estimation; prior



Citation: Zhou, T.; Ji, Y. Bayesian Methods for Information Borrowing in Basket Trials: An Overview. *Cancers* **2024**, *16*, 251. <https://doi.org/10.3390/cancers16020251>

Academic Editors: Shivaani Kummar and Kyaw Thein

Received: 7 November 2023

Revised: 22 December 2023

Accepted: 3 January 2024

Published: 5 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The field of tissue-agnostic drug development has seen increasing interest due to recent advances in molecular genetics and biomarker-driven treatment strategies. Basket trials, a type of clinical trial, have gained particular attention in this area since they simultaneously evaluate a single therapy across multiple cancer types or subtypes of the same cancer [1–6]. The rationale behind basket trials is that treatments targeting specific molecular alterations can potentially treat tumors regardless of their origin in the body. By using master protocols, basket trials can enhance operational efficiency and increase patient participation. Examples of basket trials include a study of imatinib in multiple histological subtypes of advanced sarcoma [7], a study of vemurafenib in BRAF V600 mutation-positive non-melanoma cancers [8], and a study of larotrectinib in TRK fusion-positive cancers [9], among others.

The mechanism of a drug in a basket trial is based on modifying a cancer biomarker that is prevalent across different cancer types or subtypes. Instead of conducting one trial for one disease, a basket trial includes multiple baskets, each representing a disease, so that the drug efficacy can be tested at the same time across baskets in a single clinical trial. In a way, a basket trial can be seen as a collection of multiple single-arm subtrials, one for each disease. There are typically no control arms in a basket trial, and patients are enrolled in

parallel. An obvious benefit of basket trials is that only one study team and study protocol are needed for a trial. In a basket trial, the statistical analysis of each substudy can be independent of the others. This is known as stratified analysis. However, if a basket has a small sample size, e.g., for a rare cancer type, the stratified analysis often results in large uncertainty and lacks sufficient power for efficacy evaluation. To mitigate this issue, it may be desirable to borrow information across baskets since the same treatment is tested across all of them. This enables the treatment effect in one basket to be informed by the treatment effects in other baskets, leading to improved statistical precision and power. The Bayesian paradigm provides a natural way to achieve information borrowing. For example, by assuming a common prior distribution on the basket-specific response rates, their estimates are shrunk toward a common value and tend to have lower variances.

Numerous Bayesian methods [10–23] have been developed to facilitate information sharing for basket trials, or more broadly, in clinical trials involving multiple patient subpopulations. We review the common components of these methods, such as a sampling model for the numbers of responders which involves the response rates as parameters, a transformation applied to the response rates, a prior model for the transformed response rates which typically expresses an assumption of exchangeability, a criterion for selecting the promising baskets, and a possible interim analysis plan. With the same general components, these methods mainly differ by the transformation applied to the response rates, which reflects *what* information is borrowed, and the prior model for the transformed response rates, which determines *how* information is borrowed. For example, some methods directly borrow the raw response rates, while others borrow the response rate increments from the reference rates; some methods model the transformed response rates as a random sample from a unimodal distribution, while others utilize multimodal mixture distributions. We discuss the impact of these modeling choices, particularly those related to the degree of borrowing, on the operating characteristics of the methods.

The methods for information borrowing can be extended and applied to trials involving both multiple diseases and multiple doses. Such extensions are useful for dose optimization trials under the recent Project Optimus initiative launched by the U.S. Food and Drug Administration (FDA) [24,25]. For example, after a dose escalation stage, multiple doses may be considered for expansion in multiple disease indications [26,27]. In this case, the baskets are nested in the dose expansion cohorts. We review some recent developments in this area [28].

The remainder of this paper is structured as follows. In Section 2, we review the general components, possible modeling choices, and operating characteristics of Bayesian methods for basket trials in a phase II setting. In Section 3, we review some recent developments of Bayesian methods for basket trials in the context of phase I dose optimization, which accommodate both multiple diseases and multiple doses. In Section 4, we present a discussion on future directions. Finally, Section 5 encapsulates our conclusions.

2. Basket in Phase II Studies

2.1. Trial Examples

A large number of basket trials are conducted in an exploratory phase II setting. The primary endpoint is typically tumor response. If the drug is deemed promising in some cancer (sub)types, it would warrant further investigation in a confirmatory study or conditional marketing approval.

We review three such trials. The first trial was aimed at assessing the efficacy of imatinib in patients with one of 10 different subtypes of advanced sarcoma [7]. These included angiosarcoma, Ewing sarcoma, fibrosarcoma, leiomyosarcoma, liposarcoma, malignant fibrous histiocytoma (MFH), osteosarcoma, malignant peripheral-nerve sheath tumor (MPNST), rhabdomyosarcoma, and synovial sarcoma. The primary endpoint was tumor response, defined as complete response (CR) or partial response (PR) at 2 months, or stable disease, CR or PR at 4 months. The trial was designed based on a Bayesian hierarchical model [10]. Table 1 summarizes the patient responses by sarcoma subtype.

A total of 179 patients were available for analysis. By comparing the response rates to a reference rate of 30%, the authors concluded that imatinib was not an active agent in advanced sarcoma in these subtypes.

Table 1. Data from the imatinib trial. Here, y represents the number of responders, and n is the total number of patients by sarcoma subtype.

Subtype	y	n	%
1. Angiosarcoma	2	15	13.3
2. Ewing	0	13	0.0
3. Fibrosarcoma	1	12	8.3
4. Leiomyosarcoma	6	28	21.4
5. Liposarcoma	7	29	24.1
6. MFH	3	29	10.3
7. Osteosarcoma	5	26	19.2
8. MPNST	1	5	20.0
9. Rhabdomyosarcoma	0	2	0.0
10. Synovial	3	20	15.0
Total	28	179	15.6

The second trial was conducted to study vemurafenib in BRAF V600 mutation-positive non-melanoma cancers. The study included the following cancer cohorts that received vemurafenib monotherapy: non-small-cell lung cancer (NSCLC), cholangiocarcinoma (CCA), Erdheim–Chester disease or Langerhans’ cell histiocytosis (ECD/LCH), anaplastic thyroid cancer, breast cancer, ovarian cancer, multiple myeloma, colorectal cancer (CRC-V), and all others. An additional cohort of patients with colorectal cancer received vemurafenib combined with cetuximab (CRC-VC). The primary endpoint was tumor response at week 8, as assessed by the site investigators according to the Response Evaluation Criteria in Solid Tumors [29], or the criteria of the International Myeloma Working Group [30]. The trial was designed using Simon’s two-stage method [31,32], separately for each cohort. Due to insufficient accrual, patients in the breast cancer, multiple myeloma, and ovarian cancer cohorts were eventually included in the all-others cohort. Table 2 summarizes the patient responses by cohort (not including the all-others cohort). A total of 84 patients were available for analysis. By comparing the response rates to a reference rate of 15%, the authors concluded that BRAF V600 appeared to be a targetable oncogene in some, but not all, non-melanoma cancers. Specifically, preliminary vemurafenib activity was observed in NSCLC and in ECD/LCH.

Table 2. Data from the vemurafenib trial. Here, y represents the number of responders, and n is the total number of patients by cancer cohort.

Cohort	y	n	%
1. ATC	2	7	28.6
2. ECD/LCH	6	14	42.9
3. CCA	1	8	12.5
4. CRC-V	1	26	3.8
5. CRC-VC	0	10	0.0
6. NSCLC	8	19	42.1
Total	18	84	21.4

The data from the imatinib and vemurafenib trials have since then been reanalyzed multiple times [12,16,21].

A third example is the recent pivotal study of larotrectinib [9]. A family of genes called NTRK1, NTRK2, and NTRK3 encode a protein called tropomyosin receptor kinases (TRK). Mutation in NTRK genes results in TRK fusion proteins that lead to tissue-independent

oncogenic transformation [33–35]. TRK fusion proteins are found in more than 20 different tumor types. As a result, a phase II basket trial was conducted to evaluate the therapeutic effect of larotrectinib, a TRK inhibitor, in 55 patients diagnosed with 12 different cancer types. The overall response rate was 75% based on central assessment with a 95% confidence interval of (61%, 85%). Larotrectinib was well tolerated in both adult and child populations. Based on the efficacy and safety data, the drug has been approved for treating NTRK gene fusion-positive tumors in adult and pediatric patients across cancer types. Statistical analysis pooled all the patients enrolled in the trial regardless of their tumor types. Therefore, the baskets were not differentiated in the statistical inference of drug effects. This is a special case where the biomarker, NTRK gene fusion, is highly specific and causal to a small fraction of cancers, regardless of their tissue types. In general, a targeted therapy may work in some cancer types or subtypes, which requires more sophisticated statistical design and analysis.

2.2. Statistical Setup

Consider a basket trial with J baskets. Let n_j denote the number of patients enrolled in basket j . The number of responders in basket j , denoted by y_j , is typically modeled by a binomial distribution,

$$y_j | n_j, \pi_j \sim \text{Bin}(n_j, \pi_j), \quad j = 1, \dots, J. \quad (1)$$

Here, π_j represents the true but unknown response rate of the treatment in basket j . The efficacy of the treatment can be evaluated by comparing π_j to a prespecified reference rate π_{0j} via a hypothesis test,

$$H_{0j} : \pi_j \leq \pi_{0j} \quad \text{versus} \quad H_{1j} : \pi_j > \pi_{0j}. \quad (2)$$

This reference rate can vary across baskets due to different cancer (sub)types being considered. If the observed data show strong evidence in favor of H_{1j} , the null hypothesis H_{0j} is rejected and the treatment is determined to be efficacious in basket j . Under the Bayesian paradigm, one assigns a prior distribution to π_j and calculates its posterior distribution according to Bayes' rule. The treatment is deemed promising in basket j if the posterior probability of the alternative hypothesis exceeds a prespecified threshold q_j (e.g., 0.95),

$$\Pr(\pi_j > \pi_{0j} | \text{data}) > q_j.$$

As mentioned earlier, it is desirable to specify a prior for π_j s which allows information borrowing across baskets. In the following sections, we discuss several considerations involved in the prior specification.

2.3. Prior Specification and Exchangeability

Most existing methods start by transforming π_j into a real value using, for example, a logit transformation. We denote the transformation and the real-valued parameter as $\gamma_j = h(\pi_j)$. Note that each basket j is indexed by a different parameter γ_j . Then, the γ_j s are modeled as a random sample from a common population distribution G ,

$$\gamma_j | \boldsymbol{\theta} \stackrel{iid}{\sim} G(\boldsymbol{\theta}), \quad (3)$$

where $\boldsymbol{\theta}$ denotes the vector of hyperparameters that parameterize G . Figure 1 displays a graphical representation of the hierarchical model given by Equations (1) and (3). More discussions on the choices of h and $G(\boldsymbol{\theta})$ are deferred to Sections 2.3 and 2.4. Importantly, $\boldsymbol{\theta}$ is unknown and is estimated based on data from all baskets. As a result, the posterior of γ_j is informed by both the responses within basket j , through the likelihood (1), and those in other baskets, through the prior (3). Figure 2 illustrates the effect of information

borrowing through an analysis of the imatinib data in Table 1. The point estimates of π_j for individual baskets are shrunk towards the overall response rate. Additionally, the interval estimates of π_j have shorter lengths under borrowing compared to those under stratified analysis. More details of the analysis can be found in the caption of Figure 2.

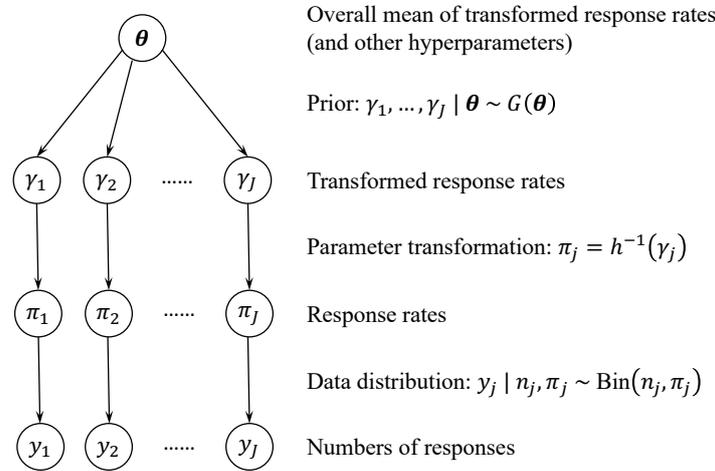


Figure 1. Graphical representation of the hierarchical model that allows information borrowing across baskets.

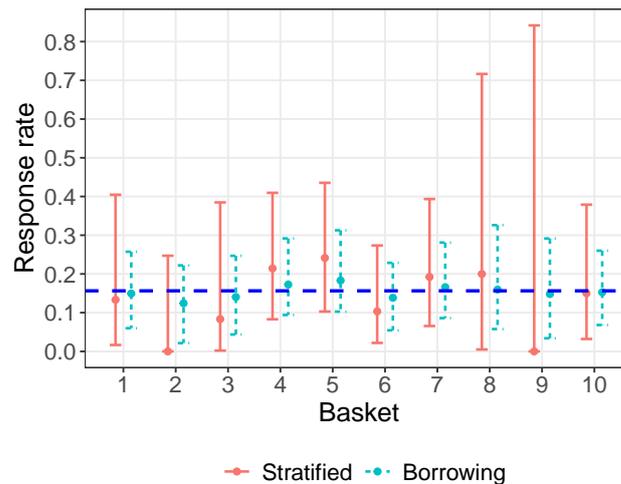


Figure 2. Illustration of the effect of information borrowing through an analysis of the imatinib data. Dots represent point estimates and error bars represent 95% confidence or credible intervals. In the stratified analysis, y_j/n_j is used as a point estimate for π_j , and the Clopper–Pearson exact method is used to construct a confidence interval. For information borrowing, the following hierarchical model is used: $\gamma_j = \text{logit}(\pi_j) - \text{logit}(\pi_{0j}), \gamma_j \mid \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2), \mu \sim N(0, 100^2), \sigma \sim \text{Half-N}(3)$. Then, the posterior mean of π_j is used as its point estimate, and the 2.5th and 97.5th percentiles of its posterior distribution are used to form a credible interval. The dashed horizontal line corresponds to the observed overall response rate, 15.6%.

Suppose a prior $p(\theta)$ is placed on θ . Implicit in model (3) is the (marginal) prior dependence among the γ_j s. Note that

$$p(\gamma_1, \dots, \gamma_J) = \int \left[\prod_{j=1}^J p(\gamma_j \mid \theta) \right] p(\theta) d\theta \neq \prod_{j=1}^J p(\gamma_j). \tag{4}$$

In fact, it can be shown that the γ_j s are positively correlated a priori [36], which enables information borrowing across baskets. Furthermore, model (3) implies an assumption of

prior exchangeability among the γ_j s. Mathematically, a sequence of random variables is called *exchangeable* if their joint distribution is invariant to permutations. From Equation (4), the joint density $p(\gamma_1, \dots, \gamma_J)$ is invariant to permutations of the indexes $(1, \dots, J)$. The assumption of prior exchangeability is reasonable when no information is available before the trial to claim that the treatment is more likely to be efficacious in certain baskets than others [37]. We note that modeling the γ_j s as independent draws from a common distribution is a stronger assumption than finite exchangeability: the former implies the latter, but not vice versa.

If there is prior knowledge to distinguish some γ_j s from others, one may incorporate an expanded notion of exchangeability in the prior construction. For example, historical clinical trials may suggest that the baskets can be divided into several subgroups. Each subgroup consists of baskets with similar historical success rates. Then, one may specify a separate prior model for the γ_j s within each subgroup. While the parameters within the same subgroup are exchangeable, those across different subgroups are not. This is known as partial exchangeability. For another example, patient responses are often associated with basket-level and patient-level covariates. If these covariates are available, they may be used to construct a regression model with an underlying assumption of conditional exchangeability. For the rest of this paper, we will restrict our attention to the exchangeable model given by Equation (3), which is employed by most existing methods.

2.4. What Information to Borrow?

The transformation $\gamma_j = h(\pi_j)$ reflects what information is borrowed across baskets. A straightforward choice is to directly borrow the response rates by assuming $\pi_j \mid \boldsymbol{\theta} \stackrel{iid}{\sim} G(\boldsymbol{\theta})$, where $G(\boldsymbol{\theta})$ is a distribution on the unit interval, e.g., a beta distribution. In this case, $h(\pi_j) = \pi_j$ is the identity transformation, and the underlying assumption is that the treatment has similar response rates across baskets. A variation in this choice is to consider a logit transformation, $h(\pi_j) = \text{logit}(\pi_j) = \log[\pi_j/(1 - \pi_j)]$. This can simplify posterior computation by allowing $G(\boldsymbol{\theta})$ to be a distribution over the real line, e.g., a normal distribution.

An alternative choice of h incorporates an adjustment for the reference rate π_{0j} . Typically, the reference rate for each basket is determined based on how well the cancer (sub)type responds to the standard of care. If there are substantial differences in the reference rates across baskets, it may be implausible to assume that the π_j s are similar. This is because baskets with lower (or higher) reference rates are also more likely to respond poorly (or positively) to the new treatment. To account for the differential reference rates, it may be more appropriate to model the response rate increments from the reference rates. For example, Berry et al. [11] considered borrowing the increments of the logit response rates, $h(\pi_j) = \text{logit}(\pi_j) - \text{logit}(\pi_{0j})$.

Lastly, a different strategy is to borrow information at the hypothesis level by letting $\gamma_j = h(\pi_j) = \mathbb{1}(\pi_j > \pi_{0j})$. See, e.g., Zhou and Ji [21]. Here, $\gamma_j = 1$ (or 0) represents H_{1j} is true (or false), indicating that the treatment is efficacious (or inefficacious) in basket j . The prior $G(\boldsymbol{\theta})$ for γ_j s can be a Bernoulli distribution. Borrowing across γ_j s reflects the assumption that if the treatment is promising, it is likely to be efficacious across multiple baskets simultaneously. This is a more general assumption than assuming the response rates are similar. For example, π_j and $\pi_{j'}$ may be quite different, but as long as they are larger than π_{0j} and $\pi_{0j'}$, respectively, the treatment is efficacious in both baskets j and j' . An additional complexity of this approach is that h is a many-to-one transformation, and the value of π_j cannot be uniquely determined by γ_j through $\pi_j = h^{-1}(\gamma_j)$. Instead, one needs to construct a prior for π_j conditional on the value of γ_j . For example, $\pi_j \mid \gamma_j = 0$ can be a beta distribution truncated to the interval $[0, \pi_{0j}]$, and $\pi_j \mid \gamma_j = 1$ can be a beta distribution truncated to the interval $(\pi_{0j}, 1]$. The prior $p(\pi_j \mid \gamma_j)$ establishes the connection between γ_j and π_j in Figure 1.

2.5. How Is Information Borrowed?

The choice of the prior $G(\theta)$ determines how information is borrowed across baskets. To illustrate ideas, suppose γ_j is real-valued, e.g., $\gamma_j = \text{logit}(\pi_j) - \text{logit}(\pi_{0j})$. A natural choice of $G(\theta)$ is then a normal distribution,

$$\gamma_j \mid \mu, \sigma^2 \stackrel{iid}{\sim} \text{N}(\mu, \sigma^2), \quad (5)$$

where the hyperparameter vector $\theta = (\mu, \sigma^2)$. The mean parameter μ represents a transformed version of the overall response rate of the treatment across all baskets. The basket-specific γ_j s are shrunk toward the common μ . The variance parameter σ^2 controls the degree of borrowing, with smaller values implying stronger shrinkage effects. At one extreme, when $\sigma^2 = 0$, all γ_j values must be equal. At the other extreme, when σ^2 approaches infinity, the shrinkage effects become negligible. The estimation of σ^2 plays a crucial role in the statistical analysis. On the one hand, overestimating σ^2 may lead to inadequate borrowing, diminishing the benefits of shrinkage estimation. On the other hand, underestimating σ^2 may result in excessive borrowing, leading to inflated type I error rates and potential failures in drug development (more on this point in Section 2.6). Yet, due to the typically limited number of baskets in a basket trial, accurate estimation of σ^2 is a challenging task.

Taking a full Bayesian approach, a hyperprior is assigned to σ^2 . A computationally convenient choice is the inverse-gamma prior, $\sigma^2 \sim \text{IG}(\alpha, \beta)$. See, e.g., Thall et al. [10] and Berry et al. [11]. It is commonly thought that small values of α and β produce a noninformative prior for σ^2 . However, Gelman [36] showed that even with small values of α and β , the $\text{IG}(\alpha, \beta)$ prior could still be quite informative and might lead to underestimation of σ^2 . Instead, the author advocated the use of a half- t prior as a less informative choice for the hierarchical standard deviation parameter, $\sigma \sim \text{Half-}t_\nu(A)$, with small ν and large A . Here, ν is the number of degrees of freedom, and A is the scale parameter. Special cases of the half- t prior include the half-Cauchy (when $\nu = 1$) and half-normal (when $\nu = \infty$) priors. The half- t prior was used by Neuenschwander et al. [12] and Zhou and Ji [21].

Alternatively, Chu and Yuan [14] proposed an empirical Bayesian approach to specify the value of σ^2 based on a measure of homogeneity among the baskets. The relationship between σ^2 and the homogeneity measure is determined through a simulation-based calibration procedure.

To further reduce the risk of excessive borrowing, the normal distribution prior in Equation (5) may be replaced by a distribution with heavier tails, e.g., a t -distribution. Such a prior accommodates occasional extreme parameters. In a basket trial, the response rates in a few baskets may be quite different from the others. A heavy-tailed prior still shrinks these extreme response rates toward the overall mean but avoids pulling them too much [37].

2.5.1. Mixture Models

In some basket trials, patient responses across baskets exhibit a clustering structure. For example, in the vemurafenib trial (Table 2), the ECD/LCH and NSCLC cohorts have similar proportions of responses, suggesting they can be clustered together. The same applies to the CRC-V and CRC-VC cohorts. To exploit such a clustering structure, a multimodal mixture prior can be placed on γ_j [12,15,18,21]. For example, consider $G(\theta)$ to be a mixture of normal distributions,

$$\gamma_j \mid \mu, \sigma^2, \mathbf{w}, K \stackrel{iid}{\sim} \sum_{k=1}^K w_k \cdot \text{N}(\mu_k, \sigma_k^2). \quad (6)$$

In this case, the hyperparameter vector $\theta = (\mu, \sigma^2, \mathbf{w}, K)$ with $\mu = (\mu_1, \dots, \mu_K)$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)$, and $\mathbf{w} = (w_1, \dots, w_K)$. Here, K is the number of mixture components, and w_k ,

μ_k , and σ_k^2 are the weight, mean, and variance of mixture component k , respectively. The weights satisfy $\sum_{k=1}^K w_k = 1$.

To facilitate interpretation, observe that the mixture prior in Equation (6) is equivalent to the following hierarchical prior,

$$\gamma_j \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2, K, s_j = k \sim N(\mu_k, \sigma_k^2), \quad \Pr(s_j = k \mid \boldsymbol{w}, K) = w_k. \quad (7)$$

In other words, each basket can be thought of as belonging to one of K latent subgroups. The indicator $s_j \in \{1, \dots, K\}$ denotes the subgroup membership for basket j , and w_k represents the prevalence of subgroup k . Conditional on the subgroup memberships, information borrowing only occurs within each subgroup. Therefore, compared to the simple normal prior, the normal mixture prior allows for more judicious information borrowing. Specifically, in the presence of substantial heterogeneity among baskets, the normal mixture prior usually leads to less borrowing, reducing the risk of type I error rate inflation. Note that the subgroup memberships are unknown a priori, and all baskets share the same prior probability of belonging to any given subgroup. As a result, prior exchangeability of the γ_j s still holds under model (6) or (7). This differs from the situation where prior knowledge exists to distinguish some baskets from others, thereby breaking the prior exchangeability assumption as discussed in Section 2.3.

The estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ follows a similar logic as in the simple normal prior case. The number and weights of mixture components, K and \boldsymbol{w} , may be prespecified or estimated from the data. Standard prior choices include a symmetric Dirichlet distribution prior for \boldsymbol{w} conditional on K , and a zero-truncated Poisson distribution prior for K [38]. Since the dimensions of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$, and \boldsymbol{w} depend on K , posterior computation under this approach typically requires trans-dimensional Markov chain Monte Carlo [39]. To avoid such computational complexity, an alternative strategy is to fit multiple models with different values of K and select the most appropriate K based on a model selection criterion such as the deviance information criterion [15,40]. From a nonparametric Bayesian modeling perspective, one may set $K = \infty$ to allow for flexibility. By further letting $w_k = v_k \prod_{l=1}^{k-1} (1 - v_l)$ and $v_l \sim \text{Beta}(1, \zeta)$, $G(\boldsymbol{\theta})$ becomes a Dirichlet process mixture model [21,41].

2.5.2. Bayesian Model Averaging

When it is reasonable to assume that the treatment has the same (transformed) response rate across multiple baskets, Bayesian model averaging can be utilized to facilitate information borrowing [16,22]. Let M_ℓ denote a partition of the J baskets into subsets. For example, when $J = 3$ baskets, the possible partitions include $M_1 = \{\{1, 2, 3\}\}$, $M_2 = \{\{1, 2\}, \{3\}\}$, $M_3 = \{\{1, 3\}, \{2\}\}$, $M_4 = \{\{2, 3\}, \{1\}\}$, and $M_5 = \{\{1\}, \{2\}, \{3\}\}$. Each partition constrains different subsets of the (transformed) basket-specific response rates to be equal. For example, under partition M_2 , γ_1 must be equal to γ_2 , but there is no constraint on γ_3 . To perform Bayesian inference, one may specify a prior for the distinct values of the (transformed) response rates conditional on the partition, as well as an additional prior for the partition itself.

Conditional on a given partition, information is pooled among baskets belonging to the same subset, while no information is borrowed between baskets in different subsets. The (marginal) posterior distribution of γ_j is a weighted average of its posteriors under different partitions,

$$p(\gamma_j \mid \text{data}) = \sum_{\ell} p(\gamma_j \mid M_\ell, \text{data}) \cdot P(M_\ell \mid \text{data}),$$

which represents a compromise between complete pooling and stratified analysis. The weights in this average correspond to the posterior probabilities of the partitions.

2.6. Operating Characteristics

The likelihood (1) and prior (3) on π_j (or a transformation of π_j) allow one to compute the posterior distribution of π_j . In most cases, the posterior is not analytically available, and Monte Carlo methods are used to approximate the posterior [42,43]. For the hypothesis test in Equation (2), a commonly used criterion to reject the null hypothesis H_{0j} is when the posterior probability $\Pr(\pi_j > \pi_{0j} \mid \text{data}) > q_j$, where q_j is a prespecified threshold that may differ across baskets.

It is common practice to evaluate the operating characteristics of a Bayesian procedure under the frequentist paradigm, as it provides insights into the procedure's long-run average behavior in repeated practical use [44]. In the context of basket trials, such evaluations are useful for understanding the practical implications of different prior choices for π_j s. Often, a set of scenarios is considered in which the true response rates are specified for the baskets, hypothetical response data are generated under each scenario, and relevant operating characteristics are recorded over repeated simulations. Table 3 provides an illustration of some possible response rate scenarios with four baskets and a reference rate of 20% for every basket. The scenarios encompass different combinations of promising and nonpromising baskets. The treatment response rates may also vary across the promising (or nonpromising) baskets. In Table 3, Scenario 1 is a global null scenario in which the treatment is inefficacious in all baskets, Scenario 2 is a global alternative scenario in which the treatment is efficacious in all baskets, and Scenarios 3–6 are mixed scenarios in which the treatment is efficacious in some but not all baskets.

Table 3. Examples of response rate scenarios used in simulations to evaluate methodologies for analyzing basket trials.

Scenario	Response Rates				Reference Rates			
	π_1	π_2	π_3	π_4	π_{01}	π_{02}	π_{03}	π_{04}
1 global null	20%	20%	20%	20%	20%	20%	20%	20%
2 global alternative	35%	35%	35%	35%	20%	20%	20%	20%
3 mixed	20%	35%	35%	35%	20%	20%	20%	20%
4 mixed	20%	20%	35%	35%	20%	20%	20%	20%
5 mixed	10%	20%	30%	40%	20%	20%	20%	20%
6 mixed	20%	20%	20%	35%	20%	20%	20%	20%

The values highlighted in bold represent the promising baskets.

The type I error rate and power are the most pertinent operating characteristics for basket trials [45]. A type I error refers to the incorrect rejection of a true null hypothesis, which, for basket trials, means to select a nonpromising basket for further investigation in a large-scale phase III study. The *basket-specific type I error rate* refers to the probability of committing a type I error in a specific basket, whereas the *family wise type I error rate* (FWER) is the probability of committing a type I error in any of the baskets. Using computer simulations, these error rates can be approximated by the relative frequencies of making the corresponding errors in a large number of simulated trials. When the null hypothesis is false, the correct action is to reject the null and select a truly promising basket for further investigation. The *basket-specific power* refers to the probability of correctly selecting a promising basket. The *family wise power* (FWP) is defined in a few different ways. For example, the *disjunctive power* (FWP-D) is the probability of correctly selecting any promising baskets, while the *conjunctive power* (FWP-C) is the probability of correctly selecting all promising baskets [46]. For a quick summary of the statistical concepts pertaining to type I error rate and power, refer to Table 4.

In exploratory basket trials, strict type I error rate control is not enforced by the regulators and is often at the discretion of the sponsors. While a more lenient type I error rate is linked to increased power, it implies a higher chance of selecting a nonpromising basket for further development, increasing the cost associated with a drug development program that will ultimately fail. On the other hand, a more stringent type I error rate

is associated with reduced power, which leads to an increased chance of missing a truly promising basket. Sponsors should carefully navigate the tradeoff between risk and benefit, determining appropriate decision criteria under limited sample size that align with their specific needs and objectives.

Table 4. Summary of concepts pertaining to type I error rate and power in basket trials.

Concept	Description
Type I error rate	
Basket-specific	Probability that a specific nonpromising basket is falsely identified as promising
Family wise (FWER)	Probability that any nonpromising basket is falsely identified as promising
Power	
Basket-specific	Probability that a specific promising basket is correctly identified as promising
Family wise & disjunctive (FWP-D)	Probability that any promising basket is correctly identified as promising
Family wise & conjunctive (FWP-C)	Probability that all promising baskets are correctly identified as promising
FWER control	
Weak control	FWER is controlled when all baskets are nonpromising
Strong control	FWER is controlled regardless of which and how many baskets are nonpromising

We illustrate the impact of information borrowing on the type I error rate and power through a simulation study based on the six scenarios in Table 3. Under each scenario, 1000 sets of hypothetical response data are generated with sample size of 20 patients for every basket. Suppose that the borrowing occurs at the logit response rate level with an adjustment for the reference rate, i.e., we let $\gamma_j = \text{logit}(\pi_j) - \text{logit}(\pi_{0j})$. Three prior choices are considered for the γ_j s that lead to different degrees of borrowing:

- (I): $\gamma_j \stackrel{iid}{\sim} N(0, 100^2)$;
- (II): $\gamma_j \mid \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2), \mu \sim N(0, 100^2), \sigma \sim \text{Half-N}(3)$;
- (III): $\gamma_j \mid \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2), \mu \sim N(0, 100^2), \sigma \sim \text{Half-N}(0.3)$.

Here, Half-N(*A*) represents a half-normal distribution with scale parameter *A*, which belongs to the half-*t* prior family discussed by Gelman [36]. Priors I, II and III correspond to no, moderate and strong borrowing, respectively.

Recall that the null hypothesis associated with basket *j*, H_{0j} , is rejected when $\Pr(\pi_j > \pi_{0j} \mid \text{data}) > q_j$. These posterior probability thresholds are typically chosen to achieve certain desirable type I error rate. Since multiple hypotheses are tested simultaneously, it may be desirable to incorporate a notion of FWER control, which limits the chance of falsely selecting any nonpromising baskets for further investigation [46]. The first type of FWER control, called *weak control*, requires that the FWER is controlled when all of the *J* null hypotheses are simultaneously true. For the six scenarios considered in Table 3, weak control of the FWER requires that the FWER is controlled under Scenario 1, the global null scenario. Suppose for simplicity the same posterior probability threshold is used across all baskets. To achieve a FWER of 5% under Scenario 1, the threshold values are 0.982, 0.946 and 0.964 under Priors I, II and III, respectively. The second type of FWER control, called *strong control*, is more stringent. It requires the control of the FWER regardless of which and how many null hypotheses are true. For the scenarios considered, strong control of the FWER requires that the FWER is controlled under all six scenarios including the mixed scenarios. Note that this does not guarantee FWER control beyond these six scenarios, but we restrict our attention to the six scenarios for simplicity. To achieve a FWER of below 5% under all six scenarios, the required threshold values under Priors I, II and III are 0.982, 0.996 and 0.984, respectively.

Table 5 shows the simulation results with weak FWER control. From Table 5, information borrowing is beneficial when the treatment response rates are homogeneous

across baskets. For example, in Scenario 2, borrowing leads to substantially increased basket-specific and family wise power. In this case, the stronger the borrowing, the larger the increase in power. When the response rates are heterogeneous, the performance of borrowing does not always compare favorably with that of no borrowing. For example, in Scenario 3, borrowing results in inflated type I error rate. In Scenario 6, strong borrowing results in lower power compared to no borrowing.

Table 5. Operating characteristics under the six simulation scenarios with different degrees of borrowing. The posterior probability thresholds are calibrated to achieve weak control of the FWER under the global null scenario.

Scenario	Borrowing	% Reject				FWER	FWP-D	FWP-C
1		0.2	0.2	0.2	0.2			
	No	1.2	1.3	1.3	1.3	5.0	0.0	0.0
	Moderate	1.5	1.5	1.3	1.4	5.0	0.0	0.0
	Strong	2.2	2.3	2.3	2.2	5.0	0.0	0.0
2		0.35	0.35	0.35	0.35			
	No	25.6	25.4	25.6	26.0	0.0	69.6	0.2
	Moderate	54.0	52.9	54.7	55.5	0.0	87.0	18.9
	Strong	79.9	78.1	78.8	79.8	0.0	90.3	63.6
3		0.2	0.35	0.35	0.35			
	No	1.1	25.2	25.2	27.0	1.1	59.2	1.7
	Moderate	8.7	43.1	44.1	45.5	8.7	73.5	15.6
	Strong	33.9	61.6	62.7	62.1	33.9	74.9	47.7
4		0.2	0.2	0.35	0.35			
	No	1.2	1.6	25.3	26.4	2.8	45.6	6.1
	Moderate	5.5	4.9	34.6	37.0	8.8	54.1	17.5
	Strong	18.6	18.1	43.7	44.1	25.3	53.1	34.7
5		0.1	0.2	0.3	0.4			
	No	0.0	1.5	12.2	44.9	1.5	51.9	5.2
	Moderate	0.0	2.4	17.8	46.6	2.4	52.6	11.8
	Strong	1.9	8.7	21.5	38.4	8.8	41.7	18.2
6		0.2	0.2	0.2	0.35			
	No	1.0	1.5	1.2	27.0	3.7	27.0	27.0
	Moderate	3.2	2.8	2.4	27.0	7.1	27.0	27.0
	Strong	8.2	7.9	7.6	24.7	14.0	24.7	24.7

The values highlighted in bold represent the promising baskets.

Table 6 reports the simulation results with strong FWER control. The issue of inflated type I error rate due to borrowing is mitigated by increasing the posterior probability thresholds. In the global alternative scenario, although less substantial, borrowing still leads to increased power. In the mixed scenarios, however, borrowing (especially strong borrowing) usually results in lower power compared to no borrowing.

In summary, in terms of operating characteristics, borrowing is beneficial when the response rates are homogeneous but may be unfavorable when the response rates are heterogeneous. For this reason, there has been some controversy about the usefulness of information borrowing in basket trials [47]. Our opinion is that borrowing is still useful. First, in the Bayesian framework, a prior serves as an expression of belief regarding which parameter values are deemed more plausible. When the prior is designed to encourage information borrowing, it implies a belief that the response rates are more likely to be homogeneous across baskets. Consequently, the performance in scenarios with homogeneous response rates should be given greater weight compared to that in scenarios with heterogeneous response rates. Second, Table 5 shows that under weak FWER control, moderate borrowing leads to considerable gain in power in the global alternative scenario without compromising much of the type I error rate and power in the mixed scenarios. In fact, with more sophisticated Bayesian modeling and judicious information borrowing, many recent methods achieve even

more improvements in power while maintaining type I error rates at reasonable levels, even in the mixed scenarios [12,14,21]. To this end, we recommend setting up the statistical analysis to borrow information across baskets where the treatment is expected to exhibit similar behavior based on the drug mechanism. If there is uncertainty about the homogeneity of the true response rates, it is recommended to borrow information in a judicious manner.

Table 6. Operating characteristics under the six simulation scenarios with different degrees of borrowing. The posterior probability thresholds are calibrated to achieve strong control of the FWER under all six scenarios.

Scenario	Borrowing	% Reject				FWER	FWP-D	FWP-C
1		0.2	0.2	0.2	0.2			
	No	1.2	1.3	1.3	1.3	5.0	0.0	0.0
	Moderate	0.6	0.5	0.7	0.5	2.0	0.0	0.0
	Strong	0.1	0.1	0.1	0.1	0.3	0.0	0.0
2		0.35	0.35	0.35	0.35			
	No	25.6	25.4	25.6	26.0	0.0	69.6	0.2
	Moderate	40.0	39.4	40.4	42.1	0.0	75.8	9.3
	Strong	35.4	34.9	36.5	37.5	0.0	60.4	12.8
3		0.2	0.35	0.35	0.35			
	No	1.1	25.2	25.2	27.0	1.1	59.2	1.7
	Moderate	5.0	30.2	30.1	32.0	5.0	58.6	7.2
	Strong	5.0	21.3	20.5	21.7	5.0	37.5	7.4
4		0.2	0.2	0.35	0.35			
	No	1.2	1.6	25.3	26.4	2.8	45.6	6.1
	Moderate	2.7	2.0	23.3	23.4	4.5	38.6	8.1
	Strong	2.2	1.3	10.1	10.6	3.0	16.0	4.7
5		0.1	0.2	0.3	0.4			
	No	0.0	1.5	12.2	44.9	1.5	51.9	5.2
	Moderate	0.0	1.5	10.6	33.3	1.5	39.3	4.6
	Strong	0.0	0.4	3.0	9.9	0.4	11.5	1.4
6		0.2	0.2	0.2	0.35			
	No	1.0	1.5	1.2	27.0	3.7	27.0	27.0
	Moderate	1.4	1.1	1.1	16.9	3.3	16.9	16.9
	Strong	0.5	0.4	0.4	4.0	1.2	4.0	4.0

The values highlighted in bold represent the promising baskets.

2.7. Interim Analysis

Patient enrollment in clinical trials typically occurs sequentially. Therefore, when designing a clinical trial, it may be desirable to incorporate provisions for interim analyses of accumulating data, allowing for the possibility of early termination of the trial [48]. Oftentimes, basket trial designs include interim monitoring for futility [11,14,16]. At the r th interim analysis, if

$$\Pr(\pi_j > \tilde{\pi}_{0j} \mid \text{data at } r\text{th interim}) < c_{jr},$$

patient accrual in basket j is halted, as the treatment is deemed inefficacious in this basket. Here, c_{jr} is a prespecified threshold (e.g., 0.05), and $\tilde{\pi}_{0j}$ may be chosen based on both the reference rate π_{0j} and a prespecified target response rate π_{1j} (e.g., $\tilde{\pi}_{0j} = (\pi_{0j} + \pi_{1j})/2$). Alternatively, the futility stopping rule can be based on the posterior predictive probability of success [48,49]. Early stopping rules have an impact on the operating characteristics of a design. For example, futility stopping rules reduce the expected number of patients enrolled and type I error rate, which can help avoid devoting too much resources to nonpromising baskets. However, they also result in a decrease in the power of finding the promising baskets.

Interim analyses can also be used to serve the purpose of adjusting the extent of information borrowing as the trial progresses. In Cunanan et al. [50], the authors proposed to assess the homogeneity of treatment effects across baskets in an interim analysis via Fisher's exact test [51]. If homogeneity is not rejected, data across all baskets are pooled into one group in the final analysis, whereas otherwise each basket is analyzed individually. The critical value of the Fisher's exact test statistic is a tuning parameter and is prespecified. As another example, Liu et al. [13] proposed to evaluate response rate heterogeneity in an interim analysis using Cochran's Q test [52]. If homogeneity is not rejected, a Bayesian hierarchical mixture model is used to borrow information across baskets in the final analysis. Otherwise, each basket is investigated independently.

2.8. Non-Technical Summary

This section discusses several aspects of Bayesian methods for information borrowing in phase II basket trials. Key statistical considerations include setting up appropriate prior distributions for quantities that are deemed homogeneous across baskets, such as response rates of the investigational drug or their increments over the reference rates. We demonstrate the benefits of information borrowing through simple simulations and advocate for the use of Bayesian methods that lead to increased statistical power despite potential type I error rate inflation.

3. Basket in Phase I for Dose Optimization

Traditional cytotoxic oncology drugs (e.g., chemo-therapies) exert their efficacy through mechanisms that directly induce cell death, cancerous or not. Therefore, a higher dose leads to more cell death, which then leads to higher efficacy and toxicity. For this reason, the maximum tolerated dose (MTD) is considered optimal for patient care since it produces the highest efficacy among all the doses that can be tolerated. In a phase I oncology trial, simple statistical designs like 3+3 [53] and i3+3 [54] are routinely used to identify a single dose as the MTD at the end of the trial. However, this MTD-centric paradigm is now being challenged.

Due to the explosive advancement in biological and genomics research since the human genome was sequenced in the early 2000s [55,56], oncology drugs have switched from directly eradicating cancer cells based on cytotoxic means to precisely targeting biological processes at the molecular level such as genetic and immune pathways. The vast success in PD-1 inhibitors [57] highlights the paradigm shift in oncology drug development. As a consequence, the design and conduct of phase I oncology trials are being transformed with the launch of the U.S. FDA's Project Optimus [25], which aims to adapt the approach of clinical trials to the new realities of cancer treatment. Under this initiative, the FDA encourages the development and application of novel trial designs and statistical methods that attempt to identify the optimal dose of oncology drugs instead of the MTD. Several publications [58–60] and an FDA draft guidance [24] have called for changes to early phase clinical trial designs. See Figure 3 for a summary of the draft guidance.

Ji and Bi [61] proposed a new platform trial design for early phase dose optimization. The design, called ADOPT, standing for Adaptive Dose Optimization Platform Trial, is structured as a phase I trial consisting of three seamless sub-phases, Ia, Ib, and Ic. Two versions of ADOPT are presented in Figure 4, denoted as ADOPT-V1 and ADOPT-V2. In both versions, phase Ia represents an improved dose escalation study highlighted by novel features like patient backfill and the use of PK/PD data. At the end of phase Ia, doses 10 mg (the MTD) and 3 mg (the dose below MTD) are selected and sent to phases Ib and Ic for testing of efficacy. ADOPT-V1 (Figure 4a) applies the multi-arm two-stage (MATS) design [28] to phases Ib and Ic. Specifically, phase Ib expands the higher dose 10 mg in three indications, making it a basket-like study. At the end of phase Ib, an interim analysis is performed for each indication to determine whether the higher dose 10 mg shows promising efficacy in that indication. If yes, the indication is selected for a randomized comparison between 10mg and 3 mg in the subsequent phase Ic. Multiple indications may be selected for phase Ic, making it another basket-like study that also

involves multiple doses. In other words, phases Ib/Ic constitute a double-basket trial. ADOPT-V2 (Figure 4b) reverses the order of dose expansion and randomized comparison. The two versions of ADOPT may be suitable for different drug development programs and mechanisms of action. For example, if it is strongly believed that the higher dose is more efficacious than the lower dose, ADOPT-V1 might be a better design since it only tests the lower dose (in phase Ic) when the higher dose demonstrates promising efficacy. Otherwise, ADOPT-V2 might be preferred, which allows randomized comparison between the two doses immediately after dose escalation in phase Ia.

Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to <https://www.regulations.gov>. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Room 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact Mirat Shah at 301-796-8547 or Stacy Shord at 301-796-6261.

U.S. Department of Health and Human Services
Food and Drug Administration
Oncology Center of Excellence (OCE)
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
January 2023
Clinical/Medical

	Collection and Interpretation of Clinical Pharmacokinetic, Pharmacodynamic, and Pharmacogenomic Data	Integrate PK/PD/PG data with clinical data (safety and efficacy) Investigate effects in multiple populations when possible
	Trial Designs to Compare Multiple Dosages	Backfill patients on multiple doses before dose comparison Randomized dose comparison (no need to power the study for superiority or non-inferiority)
	Safety and Tolerability -- Endpoints	DLT and low grade toxicity should be considered -- Toxicity burden PRO
	Subsequent Indications and Usages	Different doses for different diseases should be considered

Figure 3. Summary of the FDA draft guidance on dose optimization.

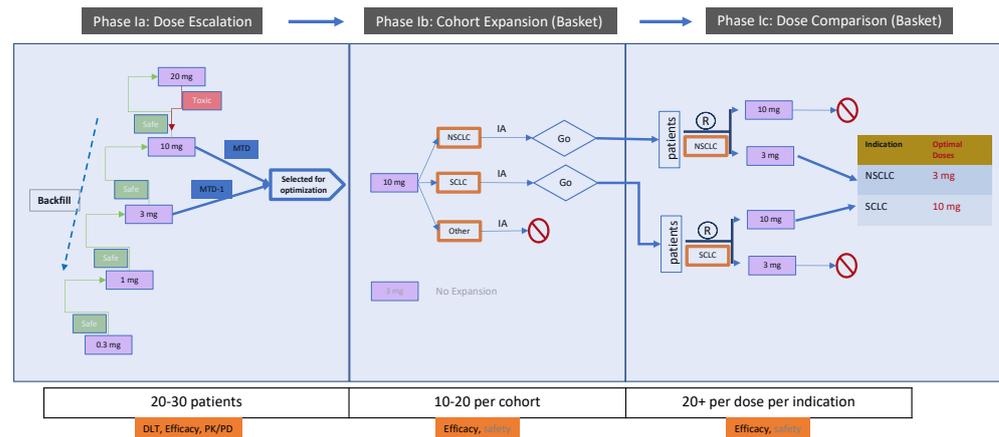
The double-basket phases Ib and Ic in ADOPT offer opportunities for employing statistical methods that facilitate information borrowing across indications. Take ADOPT-V1 as an example, which utilizes the MATS design [28] for the double-basket phases. Let $i = 1$ and 2 denote the higher and lower doses, 10 mg and 3 mg, respectively. Furthermore, let $k = 1$ and 2 denote the two stages corresponding to phases Ib and Ic. Finally, let $j = 1, \dots, J$ denote the indications. In Figure 4a, $J = 3$ corresponding to NSCLC, SCLC and other. The tuple (i, j, k) uniquely identifies an “arm” in the trial. For each arm (i, j, k) , denote by n_{ijk} the number of patients treated and y_{ijk} the number of responders. Then, assume the following sampling model,

$$\begin{aligned} \text{Stage 1: } & y_{1j1} \mid n_{1j1}, \pi_{1j} \sim \text{Bin}(n_{1j1}, \pi_{1j}), \\ \text{Stage 2: } & y_{ij2} \mid n_{ij2}, \pi_{ij}, D_{j1} = 1 \sim \text{Bin}(n_{ij2} \cdot D_{j1}, \pi_{ij}), \end{aligned}$$

where π_{ij} represents the true but unknown response rate of dose i in indication j , and $D_{j1} = 1$ (or 0) represents that indication j is selected (or not selected) for further testing in stage 2. The goal of the double-basket phases is twofold: comparing π_{1j} to an indication-specific reference rate π_{0j} for both doses (proof of concept), and comparing π_{1j} to π_{2j} between the two dose levels (dose optimization). The MATS design utilizes the following Bayesian hierarchical model to borrow information across indications,

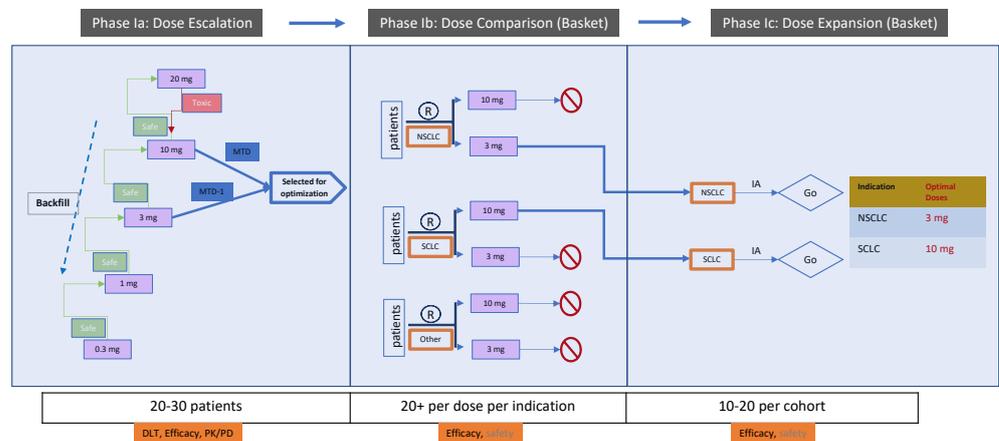
$$\begin{aligned} \text{Transformation: } & \gamma_{1j} = \text{logit}(\pi_{1j}) - \text{logit}(\pi_{0j}); \\ & \gamma_{2j} = \text{logit}(\pi_{1j}) - \text{logit}(\pi_{2j}); \\ \text{Prior for } \gamma_{1j}: & \gamma_{1j} \mid \mu_1, \sigma_1^2 \stackrel{iid}{\sim} \text{N}(\mu_1, \sigma_1^2); \\ \text{Prior for } \gamma_{2j}: & \gamma_{2j} \mid \mu_2, \sigma_2^2 \stackrel{iid}{\sim} \text{LogNormal}(\mu_2, \sigma_2^2); \\ \text{Hyperpriors: } & \mu_1 \sim \text{Normal}, \sigma_1^2 \sim \text{Inv-Gamma}; \\ & \mu_2 \sim \text{Normal}, \sigma_2^2 \sim \text{Inv-Gamma}. \end{aligned}$$

Here, γ_{1j} represents the response rate increment (on the logit scale) of the high dose over the reference rate in indication j . Modeling the γ_{1j} s as a random sample from a common normal distribution allows information borrowing of the response rate increments across indications. This is analogous to the idea illustrated in Equation (3) and Figure 1. Similarly, γ_{2j} represents the response rate difference (on the logit scale) between the high and low doses in indication j . Again, the common log-normal prior for the γ_{2j} s facilitates information borrowing of the response rate differences. It is assumed that the response rate is increasing with the dose level, and thus the γ_{2j} s are restricted to be positive. However, when the assumption is unlikely to hold, a more neutral prior, such as a normal distribution allowing γ_{2j} s to be negative, may be considered.



ADOPT: Adaptive Dose Optimization Platform Trial

(a) ADOPT-V1



ADOPT: Adaptive Dose Optimization Platform Trial

(b) ADOPT-V2

Figure 4. A stylized illustration of the Adaptive Dose Optimization Platform Trial (ADOPT). It consists of three seamless phases, Ia, Ib, and Ic. Phase Ia is for dose escalation. Phases Ib and Ic are basket trials for expansion and randomized dose comparison. IA stands for interim analysis. Novel features like backfill and integration of PK/PD data can be considered in phase Ia. The order of phases Ib and Ic may change depending on specific settings in practice, shown as the two versions V1 in (a) and V2 in (b). In the end, different indications may have different optimal doses. For example, 3mg for NSCLC and 10mg for SCLC are selected as the optimal doses.

Non-Technical Summary

This section discusses the application of Bayesian methods for information borrowing in phase I dose optimization trials. Through reviewing the MATS design, we demonstrate how information borrowing can be incorporated in trials that involve multiple doses, indications, and stages.

4. Discussion and Future Directions

We have provided an overview of Bayesian methods for information borrowing in basket trials and have summarized the general components of these methods. For other aspects of basket trials, we refer interested readers to [62–66]. For example, Park et al. [62] performed a systematic literature search to identify clinical trials that had been proposed and conducted with a basket design. Kaizer et al. [63] offered more insights into the statistical considerations, in particular those related to the type I error rate. Pohl et al. [64] covered both Bayesian and frequentist methods with more emphasis on the variety of statistical models.

Statistical software that implements Bayesian methods for basket trials is scarce. Table 7 lists a few notable ones with the most comprehensive software being commercial.

Table 7. Selected software packages that implement Bayesian methods for basket trials.

Software Name	Type	Description
R Package: <code>basket</code> (Ver. 0.10.11)	Open source	Implements the multi-source exchangeability model in Hobbs and Landin [16] and Kaizer et al. [67]
R Package: <code>bmabasket</code> (Ver. 0.1.2)	Open source	Implements the Bayesian model averaging approach in Psioda et al. [22]
R Package: <code>bhmbasket</code> (Ver. 0.9.5)	Open source	Implements the Bayesian hierarchical modeling approaches in Berry et al. [11] and Neuenschwander et al. [12]
Website: https://trialsdesign.org/ (accessed on 22 December 2023)	Free	Implements the calibrated Bayesian hierarchical model in Chu and Yuan [14]
East Bayes (https://www.cytel.com/software/east-bayes/ (accessed on 22 December 2023))	Commercial software carried by Cytel, Inc. (Cambridge, MA, USA)	Compares up to four different basket trial designs

To date, we have discussed the applications of basket designs in exploratory phase I and phase II trials. Confirmatory basket trials, on the other hand, require additional statistical considerations. For example, whether it is still appropriate to borrow information across baskets, and whether it is necessary to impose stringent control of the FWER [68,69]. Recent novel basket trial designs [70–72] shed some lights on the potential efficiency gain of confirmatory basket trials by adding a “pruning” step using external data and interim trial data to weed out unpromising indications and by performing a post-individual check after the final pooled analysis of data from all indications. In He et al. [72], the authors showed that such a design could improve the efficiency of the trial while still controlling the FWER. While the proposed design was not based on Bayesian models, the authors suggested that Bayesian techniques devised for exploratory basket trials may further improve the performance of their design.

None of the methods reviewed in this article consider borrowing on the basis of similarities between patients. In other words, if patient populations across two baskets are “similar”, it is more likely they will respond to the treatment similarly. The similarity of patients can be measured by the distance of their covariate distributions, which sets up a model framework for dependent distributions of covariates. This might be a future direction of statistical research for basket trials.

5. Conclusions

Basket trials allow simultaneous evaluation of an investigational drug in multiple patient subpopulations within a single study. Since patients across baskets receive the same treatment, it is sensible to borrow information across them to improve estimation of treatment effects in different baskets. Bayesian methods provide a natural choice to achieve information borrowing and are the focus of our review. An overarching theme across the reviewed methods is to assume the (transformed) response rates for different baskets arise from a common population distribution. This provides opportunities for Bayesian statisticians to set up priors for the transformed response rates that are essentially exchangeable and therefore enable information sharing in the estimation procedure. Simulation studies can be used to calibrate the decision criteria for efficacy evaluation to achieve desirable operating characteristics under information borrowing.

Author Contributions: Conceptualization, T.Z. and Y.J.; investigation, T.Z. and Y.J.; methodology, T.Z. and Y.J.; resources, T.Z. and Y.J.; visualization, T.Z. and Y.J.; writing—original draft preparation, T.Z. and Y.J.; writing—review and editing, T.Z. and Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: Yuan Ji declares the following conflict of interests: 1. Co-Founder of Bayesoft Inc., a statistical company providing consultation and software service for pharmaceutical and biotech industry. 2. IDMC member for Astellas, Boehringer Ingelheim and Lyell. 3. Research contract with Sanofi. 4. Executive Advisor for Cytel Inc.

References

1. Redig, A.; Jänne, P. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J. Clin. Oncol.* **2015**, *33*, 975–977. [[CrossRef](#)] [[PubMed](#)]
2. Renfro, L.; Sargent, D. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: A review and examples. *Ann. Oncol.* **2016**, *28*, 34–43. [[CrossRef](#)] [[PubMed](#)]
3. Woodcock, J.; LaVange, L.M. Master protocols to study multiple therapies, multiple diseases, or both. *N. Engl. J. Med.* **2017**, *377*, 62–70. [[CrossRef](#)] [[PubMed](#)]
4. Cunanan, K.M.; Gonen, M.; Shen, R.; Hyman, D.M.; Riely, G.J.; Begg, C.B.; Iasonos, A. Basket trials in oncology: A trade-off between complexity and efficiency. *J. Clin. Oncol.* **2017**, *35*, 271. [[CrossRef](#)] [[PubMed](#)]
5. Hobbs, B.P.; Pestana, R.C.; Zabor, E.C.; Kaizer, A.M.; Hong, D.S. Basket trials: Review of current practice and innovations for future trials. *J. Clin. Oncol.* **2022**, *40*, 3520–3528. [[CrossRef](#)] [[PubMed](#)]
6. Food and Drug Administration. Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics (Guidance for Industry). 2022. Available online: <https://www.fda.gov/media/120721/download> (accessed on 22 December 2023).
7. Chugh, R.; Wathen, J.K.; Maki, R.G.; Benjamin, R.S.; Patel, S.R.; Meyers, P.A.; Priebat, D.A.; Reinke, D.K.; Thomas, D.G.; Keohan, M.L.; et al. Phase II multicenter trial of imatinib in 10 histologic subtypes of sarcoma using a Bayesian hierarchical statistical model. *J. Clin. Oncol.* **2009**, *27*, 3148–3153. [[CrossRef](#)] [[PubMed](#)]
8. Hyman, D.M.; Puzanov, I.; Subbiah, V.; Faris, J.E.; Chau, I.; Blay, J.Y.; Wolf, J.; Raje, N.S.; Diamond, E.L.; Hollebecque, A.; et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N. Engl. J. Med.* **2015**, *373*, 726–736. [[CrossRef](#)] [[PubMed](#)]
9. Drilon, A.; Laetsch, T.W.; Kummar, S.; DuBois, S.G.; Lassen, U.N.; Demetri, G.D.; Nathanson, M.; Doebele, R.C.; Farago, A.F.; Pappo, A.S.; et al. Efficacy of larotrectinib in TRK fusion–positive cancers in adults and children. *N. Engl. J. Med.* **2018**, *378*, 731–739. [[CrossRef](#)]
10. Thall, P.F.; Wathen, J.K.; Bekele, B.N.; Champlin, R.E.; Baker, L.H.; Benjamin, R.S. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat. Med.* **2003**, *22*, 763–780. [[CrossRef](#)]
11. Berry, S.M.; Broglio, K.R.; Groshen, S.; Berry, D.A. Bayesian hierarchical modeling of patient subpopulations: Efficient designs of phase II oncology clinical trials. *Clin. Trials* **2013**, *10*, 720–734. [[CrossRef](#)]
12. Neuenschwander, B.; Wandel, S.; Roychoudhury, S.; Bailey, S. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharm. Stat.* **2016**, *15*, 123–134. [[CrossRef](#)] [[PubMed](#)]
13. Liu, R.; Liu, Z.; Ghadessi, M.; Vonk, R. Increasing the efficiency of oncology basket trials using a Bayesian approach. *Contemp. Clin. Trials* **2017**, *63*, 67–72. [[CrossRef](#)]
14. Chu, Y.; Yuan, Y. A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clin. Trials* **2018**, *15*, 149–158. [[CrossRef](#)] [[PubMed](#)]

15. Chu, Y.; Yuan, Y. BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2018**, *67*, 723–740. [[CrossRef](#)]
16. Hobbs, B.P.; Landin, R. Bayesian basket trial design with exchangeability monitoring. *Stat. Med.* **2018**, *37*, 3557–3572. [[CrossRef](#)] [[PubMed](#)]
17. Xu, Y.; Müller, P.; Tsimberidou, A.M.; Berry, D. A nonparametric Bayesian basket trial design. *Biom. J.* **2019**, *61*, 1160–1174. [[CrossRef](#)] [[PubMed](#)]
18. Chen, N.; Lee, J.J. Bayesian cluster hierarchical model for subgroup borrowing in the design and analysis of basket trials with binary endpoints. *Stat. Methods Med. Res.* **2020**, *29*, 2717–2732. [[CrossRef](#)]
19. Fujikawa, K.; Teramukai, S.; Yokota, I.; Daimon, T. A Bayesian basket trial design that borrows information across strata based on the similarity between the posterior distributions of the response probability. *Biom. J.* **2020**, *62*, 330–338. [[CrossRef](#)]
20. Jin, J.; Riviere, M.K.; Luo, X.; Dong, Y. Bayesian methods for the analysis of early phase oncology basket trials with information borrowing across cancer types. *Stat. Med.* **2020**, *39*, 3459–3475. [[CrossRef](#)]
21. Zhou, T.; Ji, Y. RoBoT: A robust Bayesian hypothesis testing method for basket trials. *Biostatistics* **2021**, *22*, 897–912. [[CrossRef](#)]
22. Psioda, M.A.; Xu, J.; Jiang, Q.; Ke, C.; Yang, Z.; Ibrahim, J.G. Bayesian adaptive basket trial design using model averaging. *Biostatistics* **2021**, *22*, 19–34. [[CrossRef](#)]
23. Zheng, H.; Wason, J.M. Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. *Biostatistics* **2022**, *23*, 120–135. [[CrossRef](#)] [[PubMed](#)]
24. Food and Drug Administration. Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases (Guidance for Industry). 2023. Available online: <https://www.fda.gov/media/164555/download> (accessed on 22 December 2023).
25. Food and Drug Administration. Project Optimus: Reforming the Dose Optimization and Dose Selection Paradigm in Oncology. 2023. Available online: <https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus> (accessed on 22 December 2023).
26. Food and Drug Administration. Expansion Cohorts: Use in First-In-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics (Guidance for Industry). 2022. Available online: <https://www.fda.gov/media/115172/download> (accessed on 22 December 2023).
27. Lyu, J.; Zhou, T.; Yuan, S.; Guo, W.; Ji, Y. MUCE: Bayesian hierarchical modelling for the design and analysis of phase 1b multiple expansion cohort trials. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2023**, *72*, 649–669. [[CrossRef](#)]
28. Jiang, Z.; Mi, G.; Lin, J.; Lorenzato, C.; Ji, Y. A multi-arm two-stage (MATS) design for proof-of-concept and dose optimization in early phase oncology trials. *Contemp. Clin. Trials* **2023**, *132*, 107278. [[CrossRef](#)] [[PubMed](#)]
29. Eisenhauer, E.A.; Therasse, P.; Bogaerts, J.; Schwartz, L.H.; Sargent, D.; Ford, R.; Dancey, J.; Arbuck, S.; Gwyther, S.; Mooney, M.; et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **2009**, *45*, 228–247. [[CrossRef](#)] [[PubMed](#)]
30. Durie, B.G.; Harousseau, J.; Miguel, J.; Blade, J.; Barlogie, B.; Anderson, K.; Gertz, M.; Dimopoulos, M.; Westin, J.; Sonneveld, P.; et al. International uniform response criteria for multiple myeloma. *Leukemia* **2006**, *20*, 1467–1473. [[CrossRef](#)] [[PubMed](#)]
31. Simon, R. Optimal two-stage designs for phase II clinical trials. *Control. Clin. Trials* **1989**, *10*, 1–10. [[CrossRef](#)]
32. Lin, Y.; Shih, W.J. Adaptive two-stage designs for single-arm phase IIA cancer clinical trials. *Biometrics* **2004**, *60*, 482–490. [[CrossRef](#)]
33. Martin-Zanca, D.; Hughes, S.H.; Barbacid, M. A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. *Nature* **1986**, *319*, 743–748. [[CrossRef](#)]
34. Vaishnavi, A.; Le, A.T.; Doebele, R.C. TRKking down an old oncogene in a new era of targeted therapy. *Cancer Discov.* **2015**, *5*, 25–34. [[CrossRef](#)]
35. Amatu, A.; Sartore-Bianchi, A.; Siena, S. NTRK gene fusions as novel targets of cancer therapy across multiple tumour types. *ESMO Open* **2016**, *1*, e000023. [[CrossRef](#)] [[PubMed](#)]
36. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534. [[CrossRef](#)]
37. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*, 3rd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014.
38. Miller, J.W.; Harrison, M.T. Mixture models with a prior on the number of components. *J. Am. Stat. Assoc.* **2018**, *113*, 340–356. [[CrossRef](#)] [[PubMed](#)]
39. Green, P.J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **1995**, *82*, 711–732. [[CrossRef](#)]
40. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2002**, *64*, 583–639. [[CrossRef](#)]
41. Sethuraman, J. A constructive definition of Dirichlet priors. *Stat. Sin.* **1994**, *4*, 639–650.
42. Robert, C.P.; Casella, G. *Monte Carlo Statistical Methods*; Springer: New York, NY, USA, 1999; Volume 2.
43. Liu, J.S. *Monte Carlo Strategies in Scientific Computing*; Springer: New York, NY, USA, 2001; Volume 75.
44. Rubin, D.B. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Stat.* **1984**, *12*, 1151–1172. [[CrossRef](#)]

45. Cunanan, K.M.; Iasonos, A.; Shen, R.; Hyman, D.M.; Riely, G.J.; Gönen, M.; Begg, C.B. Specifying the true- and false-positive rates in basket trials. *JCO Precis. Oncol.* **2017**, *1*, 1–5. [[CrossRef](#)]
46. Dmitrienko, A.; Bretz, F.; Westfall, P.H.; Troendle, J.; Wiens, B.L.; Tamhane, A.C.; Hsu, J.C. Multiple testing methodology. In *Multiple Testing Problems in Pharmaceutical Statistics*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2009; pp. 53–116.
47. Freidlin, B.; Korn, E.L. Borrowing information across subgroups in phase II trials: Is it useful? *Clin. Cancer Res.* **2013**, *19*, 1326–1334. [[CrossRef](#)]
48. Zhou, T.; Ji, Y. On Bayesian sequential clinical trial designs. *New Engl. J. Stat. Data Sci.* **2023**. [[CrossRef](#)]
49. Saville, B.R.; Connor, J.T.; Ayers, G.D.; Alvarez, J. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. *Clin. Trials* **2014**, *11*, 485–493. [[CrossRef](#)]
50. Cunanan, K.M.; Iasonos, A.; Shen, R.; Begg, C.B.; Gönen, M. An efficient basket trial design. *Stat. Med.* **2017**, *36*, 1568–1579. [[CrossRef](#)]
51. Fisher, R.A. *Statistical Methods for Research Workers*, 6th ed.; Oliver and Boyd: Edinburgh, UK; London, UK, 1936.
52. Cochran, W.G. The comparison of percentages in matched samples. *Biometrika* **1950**, *37*, 256–266. [[CrossRef](#)]
53. Storer, B.E. Design and analysis of phase I clinical trials. *Biometrics* **1989**, *45*, 925–937. [[CrossRef](#)] [[PubMed](#)]
54. Liu, M.; Wang, S.J.; Ji, Y. The i3+3 design for phase I clinical trials. *J. Biopharm. Stat.* **2020**, *30*, 294–304. [[CrossRef](#)] [[PubMed](#)]
55. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
56. Venter, J.C.; Smith, H.O.; Adams, M.D. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351. [[CrossRef](#)] [[PubMed](#)]
57. Robert, C.; Ribas, A.; Wolchok, J.D.; Hodi, F.S.; Hamid, O.; Kefford, R.; Weber, J.S.; Joshua, A.M.; Hwu, W.J.; Gangadhar, T.C.; et al. Pembrolizumab versus ipilimumab in advanced melanoma. *N. Engl. J. Med.* **2014**, *372*, 2521–2532. [[CrossRef](#)] [[PubMed](#)]
58. Shah, M.; Rahman, A.; Theoret, M.R.; Pazdur, R. Optimizing Dosing in Oncology Drug Development Q&A. Friends of Cancer Research. 2022. Available online: <https://friendsofcancerresearch.org/blog/optimizing-dosing-in-oncology-drug-development/> (accessed on 22 December 2023).
59. Shah, M.; Rahman, A.; Theoret, M.R.; Pazdur, R. How to Get the Dose Right. The ASCO Post. 2022. Available online: <https://ascopost.com/issues/may-10-2022/how-to-get-the-dose-right/> (accessed on 22 December 2023).
60. Zirkelbach, J.F.; Shah, M.; Vallejo, J.; Cheng, J.; Ayyoub, A.; Liu, J.; Hudson, R.; Sridhara, R.; Ison, G.; Amiri-Kordestani, L.; et al. Improving dose-optimization processes used in oncology drug development to minimize toxicity and maximize benefit to patients. *J. Clin. Oncol.* **2022**, *40*, 3489–3500. [[CrossRef](#)]
61. Ji, Y.; Bi, D. Oncology dose optimization in early phase trials: A new dawn. *Biopharm. Rep.* **2023**, *30*, 16–26.
62. Park, J.J.; Siden, E.; Zoratti, M.J.; Dron, L.; Harari, O.; Singer, J.; Lester, R.T.; Thorlund, K.; Mills, E.J. Systematic review of basket trials, umbrella trials, and platform trials: A landscape analysis of master protocols. *Trials* **2019**, *20*, 1–10. [[CrossRef](#)] [[PubMed](#)]
63. Kaizer, A.M.; Koopmeiners, J.S.; Kane, M.J.; Roychoudhury, S.; Hong, D.S.; Hobbs, B.P. Basket designs: Statistical considerations for oncology trials. *JCO Precis. Oncol.* **2019**, *3*, 1–9. [[CrossRef](#)] [[PubMed](#)]
64. Pohl, M.; Krisam, J.; Kieser, M. Categories, components, and techniques in a modular construction of basket trials for application and further research. *Biom. J.* **2021**, *63*, 1159–1184. [[CrossRef](#)] [[PubMed](#)]
65. Lu, C.; Li, X.; Broglio, K.; Bycott, P.; Jiang, Q.; Li, X.; McGlothlin, A.; Tian, H.; Ye, J. Practical considerations and recommendations for master protocol framework: Basket, umbrella and platform trials. *Ther. Innov. Regul. Sci.* **2021**, *55*, 1145–1154. [[CrossRef](#)]
66. Yu, Z.; Wu, L.; Bunn, V.; Li, Q.; Lin, J. Evolution of phase II oncology trial design: From single arm to master protocol. *Ther. Innov. Regul. Sci.* **2023**, *57*, 823–838. [[CrossRef](#)] [[PubMed](#)]
67. Kaizer, A.M.; Koopmeiners, J.S.; Hobbs, B.P. Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics* **2018**, *19*, 169–184. [[CrossRef](#)] [[PubMed](#)]
68. Stallard, N.; Todd, S.; Parashar, D.; Kimani, P.K.; Renfro, L.A. On the need to adjust for multiplicity in confirmatory clinical trials with master protocols. *Ann. Oncol.* **2019**, *30*, 506–509. [[CrossRef](#)]
69. Collignon, O.; Gartner, C.; Haidich, A.B.; James Hemmings, R.; Hofner, B.; Pétavy, F.; Posch, M.; Rantell, K.; Roes, K.; Schiel, A. Current statistical considerations and regulatory perspectives on the planning of confirmatory basket, umbrella, and platform trials. *Clin. Pharmacol. Ther.* **2020**, *107*, 1059–1067. [[CrossRef](#)]
70. Chen, C.; Li, X.; Yuan, S.; Antonijevic, Z.; Kalamegham, R.; Beckman, R.A. Statistical design and considerations of a phase 3 basket trial for simultaneous investigation of multiple tumor types in one study. *Stat. Biopharm. Res.* **2016**, *8*, 248–257. [[CrossRef](#)]
71. Beckman, R.; Antonijevic, Z.; Kalamegham, R.; Chen, C. Adaptive design for a confirmatory basket trial in multiple tumor types based on a putative predictive biomarker. *Clin. Pharmacol. Ther.* **2016**, *100*, 617–625. [[CrossRef](#)]
72. He, L.; Ren, Y.; Chen, H.; Guinn, D.; Parashar, D.; Chen, C.; Yuan, S.S.; Korostyshevskiy, V.; Beckman, R.A. Efficiency of a randomized confirmatory basket trial design constrained to control the family wise error rate by indication. *Stat. Methods Med. Res.* **2022**, *31*, 1207–1223. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.