Bayesian modelling strategies for borrowing of information in randomised basket trials

Luke O. Ouma1✉ | Michael J. Grayling1✉ | James M. S. Wason1✉ | Haiyan Zheng2✉

1Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
2MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Correspondence
Luke O. Ouma, Population Health Sciences Institute, Ridley Building 1, Queen Victoria Road, Newcastle upon Tyne NE1 7RU, UK.
Email: luke.ouma@newcastle.ac.uk

Funding information
Cancer Research UK, Grant/Award Number: RCCPDF\100008

Abstract
Basket trials are an innovative precision medicine clinical trial design evaluating a single targeted therapy across multiple diseases that share a common characteristic. To date, most basket trials have been conducted in early-phase oncology settings, for which several Bayesian methods permitting information sharing across subtrials have been proposed. With the increasing interest of implementing randomised basket trials, information borrowing could be exploited in two ways; considering the commensurability of either the treatment effects or the outcomes specific to each of the treatment groups between the subtrials. In this article, we extend a previous analysis model based on distributional discrepancy for borrowing over the subtrial treatment effects (‘treatment effect borrowing’, TEB) to borrowing over the subtrial groupwise responses (‘treatment response borrowing’, TRB). Simulation results demonstrate that both modelling strategies provide substantial gains over an approach with no borrowing. TRB outperforms TEB especially when subtrial sample sizes are small on all operational characteristics, while the latter has considerable gains in performance over TRB when subtrial sample sizes are large, or the treatment effects and groupwise mean responses are noticeably...
heterogeneous across subtrials. Further, we notice that TRB, and TEB can potentially lead to different conclusions in the analysis of real data.

**KEYWORDS**

biomarker-guided trial, master protocol, personalised medicine, precision medicine, randomised controlled basket trial

1 | INTRODUCTION

Improved understanding of molecular biology has facilitated rapid advances in identification of patient subgroups with better response to a targeted therapy. For instance, ado-trastuzumab emtansine in HER2 amplified ovarian and lung cancers (Li et al., 2018) and pembrolizumab in metastatic, microsatellite-instability–high cancer patients (Lemery et al., 2017) have demonstrated efficacy leading to FDA approval. Consequently, several innovative trial designs have emerged to efficiently evaluate targeted therapies under a single master protocol with the potential to expedite drug development (Woodcock & LaVange, 2017). The core idea in these designs is to (i) stratify patients with a given disease based on disease characteristics most predictive of treatment response and evaluate several linked targeted therapies (umbrella trials) (Siden et al., 2019; Woodcock & LaVange, 2017) or (ii) stratify patients with a shared potential therapeutic target in different underlying conditions based on their disease subgroups, evaluating a single therapy (basket trials) (Kaizer et al., 2019; Siden et al., 2019; Woodcock & LaVange, 2017). Trial designs (i) and (ii) can be randomised or non-randomised and can allow addition of arms and incorporation of adaptive design considerations.

This paper will focus on basket trials that evaluate a single treatment for multiple disease indications with a common characteristic (Siden et al., 2019; Tao et al., 2018; Woodcock & LaVange, 2017), which could be a common genomic biomarker, mechanism of drug activity or clinical symptom that the treatment targets. Figure 1 illustrates randomised and non-randomised basket trial designs, respectively. An example of a basket trial is that of TRK inhibitor Larotrectinib in 17 unique rare TRK fusion-positive cancers (Drilon et al., 2018). This study enrolled both paediatric and adult patients and reported a 75% overall response rate, demonstrating the utility of basket designs in streamlining drug development in rare diseases.

In general, statistical analysis of basket trials follows one of three approaches: (i) stand-alone analysis where subtrials are conducted and analysed separately as if they are individual studies (an example is the imatinib basket trial [Heinrich et al., 2008]); (ii) pooled analysis where all patients are considered as ‘homogenous’ and thus combined as one pooled data set, as in the basket trial of Larotrectinib (Drilon et al., 2018); or (iii) an analysis strategy that accounts for potential heterogeneity in treatment response across subtrials (Kaizer et al., 2019; Pohl et al., 2021). In practice, considerable heterogeneity in treatment response across subgroups within a basket trial can be observed. An example is the basket trial by Hyman et al. where vemurafenib displayed efficacy in BRAF-V600-mutant NSCLC and Erdheim-Chester disease patients but not in colorectal cancer patients (Hyman et al., 2015). Consequently, there is a need to account for potential heterogeneity in both design and analysis; one basket trial accounting for heterogeneity is the ongoing Chinese Bayesian basket trial of Danzhu Fuyuan Granule as Adjunctive Therapy for Chronic Stable Angina, Vascular Dementia, and Idiopathic Membranous Nephropathy (NCT04498962).
Several statistical analysis methods accounting for potential heterogeneity (type [iii] analyses above) in basket trials have been proposed and/or implemented, using either Bayesian approaches, frequentist approaches or a combination of both. Such methods enable borrowing of information across subtrials; consequently, the statistical power may be increased for detecting that a treatment benefits a specific subgroup. This could be particularly important when a basket trial involves rare subgroups. Pohl et al. present a comprehensive review of existing statistical analysis methods that incorporate borrowing of information in basket trials (Pohl et al., 2021). Notable Bayesian methods include Bayesian hierarchical modelling (Berry et al., 2013), EXNEX (Neuenschwander et al., 2016), Simon’s basket design (Simon et al., 2016), calibrated Bayesian hierarchical models (Chu & Yuan, 2018a), Bayesian latent subgroup design (Chu & Yuan, 2018b), Bayesian model averaging (Asano & Hirakawa, 2020; Psioda et al., 2019), Bayesian cluster hierarchical models (Chen & Lee, 2020), and the commensurate predictive prior (CPP) approach (Zheng & Wason, 2022). Notably, majority of these methods consider a single-arm, non-randomised trial setting that is often used in early-phase oncology trials.

Recently, there has been increased interest in conducting randomised precision medicine trials, whenever possible, to improve the chances of a successful phase III trial (Saad et al., 2017), even in rare diseases (Lasch et al., 2017; Prasad & Oseran, 2015). The potential and growing use of basket trials outside of oncology carries even greater prospects for randomisation as an important feature in the design of future basket trials. Currently, randomised basket trials have been implemented in oncology (IMPACT II trial, NCT02152254), HIV/AIDS (Moore et al., 2021), autoimmune and neurological disorders like primary biliary cholangitis (PBC) and Parkinson’s disease (Zheng & Wason, 2022), auto-inflammatory syndromes like familial Mediterranean fever and TRAP syndrome (De Benedetti et al., 2018), and have been proposed for immune-mediated inflammatory diseases (Grayling et al., 2021). Evidently, development of statistical methodology for randomised basket trials will be of great importance.
A key question to randomised basket trials relates to how borrowing of information is handled in the analysis. In general, approaches to borrowing information involve consideration of how similar the treatment response is across subtrials. In a two-arm setting, however, this could be achieved by considering the treatment effects across subtrials (referred to as ‘Treatment Effect Borrowing’, TEB) or the responses per treatment arm (later called groupwise mean responses in this paper, and referred to as ‘Treatment Response Borrowing’, TRB) across subtrials. This raises two important issues. Firstly, there is a need for a suitable method for TRB to be proposed for randomised basket trials. Secondly, the question of which borrowing approach is better is important; it is plausible that either method might be best in different situations. It is these points that this paper seeks to address.

This paper builds on work by Zheng and Wason (2022), which proposed a CPP approach that models the commensurability of treatment effects to leverage information from complementary subtrials (i.e., it facilitates TEB). We propose a new modified implementation of their methodology that models the commensurability of the groupwise mean responses to permit borrowing of information across subtrials (i.e., we extend their approach to allow for TRB). The groupwise mean responses refer to the mean response in either the control or experimental treatment arm within a subtrial. It is expected that this new TRB approach could outperform the original TEB modelling strategy if the response in one arm (e.g., the experimental treatment) is more similar across subtrials than the other (e.g., the control). We then evaluate the new and existing Bayesian modelling strategies for borrowing of information in randomised basket trials under different scenarios, through a comprehensive simulation study.

Our research is motivated by a phase II randomised basket trial (hereafter referred to as the OACS trial) which evaluates the efficacy of Obeticholic acid for treating cognitive dysfunction in three distinct patient subgroups: early-stage PBC, late-stage PBC and Parkinson’s disease. Each patient subgroup has a corresponding subtrial in which patients are randomised equally to a new targeted treatment or placebo. The primary outcome is the change in a composite score that measures the patients overall cognitive performance after 26 weeks of treatment, which is regarded as a continuous endpoint. Each subtrial is to be analysed separately; a secondary analysis will allow borrowing of information across subtrials. The exact model to be used, and whether TEB or TRB will be carried out, is currently not determined. This work is part of on-going research to ascertain which approach will be used.

The remainder of this article is organised as follows. In Section 2 we elaborate on our proposed TRB approach, provide an overview of the TEB methodology, and describe the simulation study including considered scenarios and trial characteristics. We also provide a description of case studies that will be used to further investigate the relative merits of the methods considered. Section 3 presents our results and Section 4 a discussion.

2 | METHODS

In this paper, we focus on the commensurate prior approach (Zheng & Wason, 2022) to implementing borrowing of information across patient subgroups in randomised basket trials based on two modelling strategies. To be more precise, we parameterise the level of commensurability in terms of (i) the subtrial-specific treatment effects (TEB) or (ii) the mean responses by treatment groups across the subtrials (TRB). We elaborate on an existing commensurate prior approach incorporating TEB in Section 2.1 (Zheng & Wason, 2022), and develop a new strategy for TRB in
Section 2.2. Both will be compared with stand-alone analyses (i.e., no borrowing of information; referred to as ‘No Borrowing’, NB) in a comprehensive simulation study.

### 2.1 Treatment effect borrowing

Consider a basket trial that enrols patients in \( K \) subgroups. Furthermore, \( n_k \) patients are randomised (equally) to receive a new treatment or a control per subtrial \( k = 1, \ldots, K \). The primary endpoint in this basket trial is a continuous outcome, denoted by \( y_{ik} \), is collected from each patient \( i = 1, \ldots, n_k \) in each subtrial \( k = 1, \ldots, K \). Let \( \theta_k \) be the subtrial-specific treatment effect and \( T_{ik} \) is a binary treatment assignment indicator; \( T_{ik} = 1 \) if patient \( i \) in subtrial \( k \) receives the experimental treatment and 0 otherwise (control). For simplicity, we consider here a regression model that includes just the intercept and treatment effect parameters

\[
y_{ik} = \beta_k + \theta_k T_{ik} + \epsilon_{ik}.
\]

(1)

Then, for each subtrial \( k = 1, \ldots, K \), we have

\[
E(y_{ik} | T_{ik} = 1) = \beta_k + \theta_k, \quad E(y_{ik} | T_{ik} = 0) = \beta_k,
\]

\( i = 1, \ldots, n_k \), where \( T_{ik} \) denotes treatment allocation for the \( i \)th patient in subtrial \( k \) to either control \( (T_{ik} = 0) \) or the experimental treatment \( (T_{ik} = 1) \) and \( \theta_k \) is the estimator of the experimental treatment benefit over control.

We use \( k^* \) to label a specific subtrial of our local estimation interest. We wish to assess the degree of commensurability between \( \theta_{k^*} \) and \( \theta_k \), for \( k \neq k^* \), to determine how much information we can leverage from complementary subtrial \( k \). We introduce a precision parameter, denoted by \( \nu_{kk^*} \), which captures the degree of consistency between the two location parameters \( \theta_{k^*} \) and \( \theta_k \).

Then, \( \theta_{k^*} \) would have a normal predictive prior centred on the complementary subtrial parameter \( \theta_k \) such that

\[
\theta_{k^*} | \theta_k, \nu_{kk^*} \sim N \left( \theta_k, \frac{1}{\nu_{kk^*}} \right), \quad k^* \neq k.
\]

(2)

We place a spike and slab prior (Hobbs et al., 2012; Mitchell & Beauchamp, 1988) \( g(\nu_{kk^*}) \) on each precision parameter \( \nu_{kk^*} \). The spike and slab prior distribution is a discrete mixture distribution defined as locally uniform between two limits \( 0 \leq \alpha_l \leq \alpha_u \) (the slab component), and with a probability mass concentrated at a point \( S > \alpha_u \) (the spike component). Explicitly, the prior assumes that \( \omega \)

\[
\mathbb{P}(\nu_{kk^*} < \alpha_l) = 0,
\]

\[
\mathbb{P}(\nu_{kk^*} < u) = \omega_{kk^*} \left\{ \frac{u - \alpha_l}{\alpha_u - \alpha_l} \right\}, \quad \alpha_l \leq u \leq \alpha_u,
\]

\[
\mathbb{P}(\nu_{kk^*} > \alpha_u) = \mathbb{P}(\nu_{kk^*} = S) = 1 - \omega_{kk^*},
\]

(3)

where \( \omega_{kk^*} \) denotes the prior probability that \( \alpha_l \leq \nu_{kk^*} \leq \alpha_u \). A small \( \omega_{kk^*} \) corresponds to a large value of \( \nu_{kk^*} \) at \( S \), thus allowing for strong borrowing of information. When the complementary subtrials are incommensurate, a greater probability mass is allocated to the slab prior through a large \( \omega_{kk^*} \), hence information from the complementary subtrial is discounted.
To quantify \( \omega_{kk^*} \), the prior belief about the incommensurability between subtrial \( k \) and \( k^* \), we proceed by computing the distributional divergence between the posterior distributions \( \pi_k(\theta_k | y_k) \) and \( \pi_{k^*}(\theta_{k^*} | y_{k^*}) \) and relate these distances to the spike prior probability \( (y_k \text{ and } y_{k^*} \text{ are outcome data from subtrial } k \text{ and } k^*, \text{ respectively, } k^* \neq k) \). Following Zheng and Wason (2022), we use the Hellinger distance \( (d_H) \) to quantify this distributional divergence for three key reasons: (i) \( d_H \) is a natural choice since \( d_H(\pi_{\theta_{k^*}}, \pi_{\theta_k}) \in [0, 1] \) that is, it falls strictly in the desired interval such that \( \omega_{kk^*} \in [0, 1] \); (ii) \( d_H \) is invariant to transformation; and (iii) symmetry, \( d_H(\pi_{\theta_{k^*}}, \pi_{\theta_k}) = d_H(\pi_{\theta_k}, \pi_{\theta_{k^*}}) \). We compute \( d_H \) as follows

\[
d_H(\pi_{\theta_{k^*}}, \pi_{\theta_k}) = \frac{1}{2} \int_{-\infty}^{\infty} \left( \sqrt{\frac{d\pi(\theta_{k^*} | y_{k^*})}{d\theta}} - \sqrt{\frac{d\pi(\theta_k | y_k)}{d\theta}} \right)^2 d\theta.
\]

(4)

The prior probability \( \omega_{kk^*} \) can be defined as a function of the computed Hellinger distance, for the desire of incorporating commensurate information only. For simplicity, we let \( \omega_{kk^*} = d_H(\theta_{k^*}, \theta_k) \).

As \( d_H(\pi_{\theta_{k^*}}, \pi_{\theta_k}) \to 1 \), more probability mass is allocated to the slab prior, hence a smaller precision parameter (larger variance), which denotes a high degree of incommensurability between subtrials \( k^* \) and \( k \). Consequently, we down-weight information from the corresponding complementary subtrial. On the contrary, as \( d_H(\pi_{\theta_{k^*}}, \pi_{\theta_k}) \to 0 \), more probability mass is allocated to the spike prior, leading to a large precision parameter \( \nu_{kk^*} \) (smaller variance). Consequently, information from the complementary subtrial is largely incorporated. The Hellinger distances are then normalised into a series of weights to determine how much information is leveraged from a complementary subtrial. When \( K \geq 3 \), the \( K - 1 \) CPPs are combined into a marginal predictive prior for \( \theta_{k^*} \).

The resulting marginal predictive prior for \( \theta_{k^*} \) (Equation 5) is updated to the posterior using data from subtrial \( k^* \) to inform decision-making.

\[
\pi^{MPP}(\theta_{k^*} | y_{k^*}, y_{(-k^*)}) \propto L(y_{k^*} | \theta_{k^*}) \times \pi^{MPP}(\theta_{k^*} | y_{(-k^*)}),
\]

(5)

where \( \pi(\cdot) \) and \( L(\cdot) \) denote the prior/posterior distribution and likelihood function respectively. Specifically, go and no-go decisions are made which refer to whether to recommend continuation or termination of investigation of the experimental therapy. Such decisions are made within all \( K \) subtrials, each testing the hypothesis that the experimental therapy provides benefit over control:

\[ H_{0k} : \theta_k \leq 0 \text{ vs. } H_{1k} : \theta_k > 0, \quad k = 1, \ldots, K. \]

Precisely, decisions are based on probabilistic inferences about \( \theta_k \); a go decision is made if the estimated posterior probability that \( \theta_k \) is greater than a prespecified threshold \( \theta_0 \) exceeds a desired probability level (e.g., 0.975), that is, \( P(\theta_k > \theta_0) > 0.975 \), otherwise a no-go decision is taken. The threshold \( \theta_0 \) represents a nominated magnitude of improvement that provides compelling evidence of experimental treatment benefit over control.

This TEB approach accommodates (i) complete discarding of information if \( \theta_k \) and \( \theta_{k^*} \) are largely incommensurate and (ii) reduces to the complete pooling approach if the parameters are commensurate. In the special case when \( K = 2 \), and there is no a priori assumption about which subtrial has a larger treatment effect, the degree of borrowing between \( \theta_1 \) and \( \theta_2 \) can be
anticipated to be symmetric (although this may differ) due to the symmetry property of the Hellinger distance, following (Equation 4) to define the spike and slab prior.

2.2 Treatment response borrowing

Let \( j = E, C \) index the experimental treatment and control groups, respectively. Denote the group-wise mean response by \( \varphi_{jk} \), for subtrial \( k = 1, \ldots, K \). The group-wise mean response refers to the mean for a treatment group \( j \) within a subtrial \( k \). Given that \( \varphi_{jk} \) is our subtrial-specific parameter of interest, we define subtrial \( k^* \) and \( k (k^* \neq k) \) as our contemporary and complementary subtrials, respectively. Let \( y_{jk}^{(i)} \) represent outcome measurement from treatment arm \( j \) in subtrial \( k \). We place an operational prior \( \pi_{0k^*} \left( \varphi_{jk^*} \right) \) on the parameter \( \varphi_{jk^*} \). Using Bayes’ Theorem, we have an operational posterior as follows

\[
\pi_{k^*} \left( \varphi_{jk^*} | y_{k^*}^{(i)} \right) \propto L \left( y_{k^*}^{(i)} | \varphi_{jk^*} \right) \pi_{0k^*} \left( \varphi_{jk^*} \right),
\]

which represents information carried by the contemporary subtrial \( k^* \).

To facilitate information borrowing, we make use of a CPP—a normal predictive prior centred at \( \varphi_{jk} \), with a precision parameter \( \nu_{kk^*} \). For treatment arm \( j \), we can represent the normal predictive distribution for \( \varphi_{jk^*} \) as

\[
\varphi_{jk^*} | \varphi_{jk}, \nu_{kk^*}^{(j)} \sim N \left( \varphi_{jk}, \nu_{kk^*} \right).
\]

As such, we have the CPP as

\[
\pi_{CPP} \left( \varphi_{jk^*}, \nu_{kk^*}^{(j)} | y_{k^*}^{(i)}, \varphi_{jk} \right) \propto L \left( y_{k^*}^{(i)} | \varphi_{jk} \right) \pi_{0k} \left( \varphi_{jk} \right) \times \nu_{kk^*}^{(j)} \varphi \left( \left( \varphi_{jk^*} - \varphi_{jk} \right) \nu_{kk^*}^{(j)} \right) \times g \left( \nu_{kk^*}^{(j)} \right),
\]

where \( g \left( \nu_{kk^*}^{(j)} \right) \) denotes the prior distribution of the unknown parameter \( \nu_{kk^*}^{(j)} \) and \( \varphi \) is the standard normal density function. The choice of \( g \left( \nu_{kk^*}^{(j)} \right) \) is important to allow for robust borrowing on treatment arm \( j \) from subtrial \( k \) to \( k^* \); specifically, we expect that (i) information can be completely discarded if \( \varphi_{jk^*} \) and \( \varphi_{jk} \) are incommensurate, \( \nu_{kk^*} \sim 0 \), and (ii) strong borrowing occurs when subtrial \( k^* \) and \( k \) are largely commensurate, \( \nu_{kk^*} \gg 0 \).

Our new modelling strategy allows for more flexible borrowing of information between subtrials, using their commensurability in either the experimental or control treatment, or both. Thus, for each precision parameter \( \nu_{kk^*}^{(j)} \), we place a spike and slab prior with respective prior probabilities of \( 1 - \omega_{kk^*}^{(j)} \) and \( \omega_{kk^*}^{(j)} \):

\[
\mathbb{P} \left( \nu_{kk^*}^{(j)} < \alpha_1 \right) = 0,
\]

\[
\mathbb{P} \left( \nu_{kk^*}^{(j)} < u \right) = \omega_{kk^*}^{(j)} \left\{ \frac{u - \alpha_1}{\alpha_u - \alpha_1} \right\}, \alpha_1 \leq u \leq \alpha_u,
\]

\[
\mathbb{P} \left( \nu_{kk^*}^{(j)} > \alpha_u \right) = \mathbb{P} \left( \nu_{kk^*}^{(j)} = S \right) = 1 - \omega_{kk^*}^{(j)}.
\]

Specifically, when \( \omega_{kk^*}^{(j)} \to 0 \), \( \nu_{kk^*}^{(j)} \) is as large as \( S \gg 0 \), thus allowing for strong borrowing of information.
To quantify $\omega_{kk^*}^{(j)}$, we compute the distributional divergence between the posterior distributions $\pi_{kk^*} (\varphi_{jk^*} | y_{k^*}^{(j)})$ and $\pi_k (\varphi_{jk} | y_k^{(j)})$ using the Hellinger distance discrepancy measure, $d_H$, (Equation (10)) as previously described in Section 2.1 above.

$$d_H \left( \pi_{\varphi_{jk}}, \pi_{\varphi_{jk^*}} \right) = \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} \left( \sqrt{\frac{d\pi \left( \varphi_{jk^*} | y_{k^*}^{(j)} \right)}{d\varphi}} - \sqrt{\frac{d\pi \left( \varphi_{jk} | y_k^{(j)} \right)}{d\varphi}} \right)^2 d\varphi. \quad (10)$$

Likewise, we let $\omega_{kk^*} = d_H \left( \varphi_{jk}, \varphi_{jk^*} \right)$ to incorporate information from the corresponding complementary subtrial as $d_H \left( \varphi_{jk}, \varphi_{jk^*} \right) \to 0$ (high commensurability) and discount information when $d_H \left( \varphi_{jk}, \varphi_{jk^*} \right) \to 1$ (high incommensurability).

In a randomised basket trial with $K \geq 3$ subtrials, we have $K - 1$ complementary subtrials from which we may borrow information when estimating $\varphi_{jk^*}$. From Equation (8), this translates to having $K - 1$ CPPs marginally on $\varphi_{jk^*}$, hence we need to obtain a collective prior for $\varphi_{jk^*}$. In theory, we can see the location parameter estimate $\varphi_{jk^*}$ as a weighted sum of $K - 1$ hypothetical random variables,

$$\varphi_{jk^*} = \sum_{k=1,k \neq k^*}^{K} p_{kk^*}^{(j)} \hat{\varphi}_{jk}, \forall k^* = 1, \ldots, K, \quad (11)$$

where $p_{kk^*}^{(j)}$ quantifies the relative importance of information from arm $j$ in subtrial $k$.

Given each of the $K - 1$ CPPs as in Equation (8) can be represented approximately as a $N \left( \hat{\varphi}_{jk}, \tau_{k}^{(j)} \right)$ for ease of notation, we can suppose that $\hat{\varphi}_{jk} \sim N \left( \hat{\varphi}_{jk}^{(j)}, \tau_{k}^{(j)} \right)$. Then, the collective marginal predictive prior for $\varphi_{jk^*}$ would be as follows,

$$\varphi_{jk^*} | y_{(-k^*)} \sim N \left( \sum_{k=1,k \neq k^*}^{K} p_{kk^*}^{(j)} \hat{\varphi}_{jk}, \sum_{k=1,k \neq k^*}^{K} p_{kk^*}^{(j)} \tau_{k}^{(j)} \right), \quad (12)$$

To obtain $p_{kk^*}^{(j)}$ (the weight representing the degree of commensurability between $\varphi_{jk^*}$ and $\varphi_{jk}$), we organise the pairwise discrepancy measurements into a symmetric $K \times K$ matrix

$$d_H \left( \pi_{\varphi_{jk}}, \pi_{\varphi_{jk^*}} \right) = \begin{pmatrix} 0 & d_{12}^{(j)} & \ldots & d_{1k}^{(j)} \\ d_{21}^{(j)} & 0 & \ldots & d_{2k}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{k1}^{(j)} & d_{k2}^{(j)} & \ldots & 0 \end{pmatrix}, \quad (12)$$

where each column lists the pairwise discrepancies between $\varphi_{jk^*}$ and $\varphi_{jk}$, for $k^* = 1, \ldots, K$. We then normalise the discrepancies columnwise. A decreasing function given by $p_{kk^*}^{(j)} = \frac{\exp(-d_{kk^*}^{(j)} / R_0)}{\sum_k \exp(-d_{kk^*}^{(j)} / R_0)}$, has been shown to have satisfactory properties to compute these weights (Zheng et al., 2020; Zheng & Wason, 2022). The largest weight is assigned to a subtrial with the smallest $d_{kk^*}^{(j)}$ per treatment arm $j$. Here $d_{kk^*}^{(j)}$ is the Hellinger distance when performing borrowing in treatment arm $j$ across the $k$ subgroups while $R_0$ controls the influence of the $d_{H}$ on the weights. The value $R_0$ of is pre-defined here to be 0.15 as in Zheng and Wason (2022). Simulation results...
elsewhere have shown that small values are preferred as the methodology identifies commensurate subtrials more sensitively relative to when \( R_0 \) is large where it loses this ability (Zheng & Wason, 2022). When \( R_0 \) is sufficiently large (\( R_0 \gg d_{kk^*} \)), the weights will be approximately the same irrespective of the magnitude of the pairwise Hellinger distance. However as \( R_0 \) tends to 0 (\( R_0 \to 0^+ \)), then the weights \( p_{kk^*} \to 1 \) as \( d_{kk^*} \to 0 \). Thereafter, we update the marginal predictive prior (Equation 12) using information from subtrial \( k \),

\[
\pi^{\text{MPP}}(\varphi_{jk^*} | y^{(j)}_{k^*}, y^{(j)}_{(-k^*)}) \propto L\left(y^{(j)}_{k^*} | \varphi_{jk^*}\right) \times \pi^{\text{MPP}}\left(\varphi_{jk^*} | y^{(j)}_{(-k^*)}\right),
\]

where \( y^{(j)}_{(-k^*)} \) denotes information from all subtrials except that from the contemporary subtrial \( k^* \) for arm \( j \).

Although we implement borrowing by treatment response based on \( \varphi_{jk^*} \), our interest lies on the subtrial-specific treatment effect defined as \( \theta_k = \varphi_{Ek} - \varphi_{Ck} \) for subtrial \( k \). The decision-making criteria (i.e., for go vs. no-go conclusions) is the same as in Section 2.1 above based on \( \theta_k \).

3 | SIMULATION STUDY

3.1 | Basic setting of the randomised basket trials

Consider a hypothetical phase II basket trial with \( K = 5 \) subtrials. A maximum of 336 patients are to be recruited overall, and within each subtrial, \( k = 1, \ldots, 5 \), unequal numbers of patients \( n_k \) are to be enrolled: \( n_1 = 70, n_2 = 66, n_3 = 64, n_4 = n_5 = 68 \). Our choice of sample sizes and number of subtrials is informed by recent basket trials. In particular, the CLUSTER randomised basket trial (De Benedetti et al., 2018) had 63–72 patients in two-thirds of its subtrials, while several basket trials (Hyman et al., 2015; Slosberg et al., 2018) have examined a fairly large number of patient subgroups.

We simulate data from a hypothetical randomised basket trial as follows; we suppose that \( y_{ik} \sim N\left(\mu_{ik}, \sigma^2\right) \) with \( \mu_{ik} = \beta_{0k} + \theta_k T_{ik} \). We assume \( \beta_{0k} = 5 \) and the inter-patient SD \( \sigma = 0.4 \). In the Bayesian analysis, we specify a random effects model for \( \beta_{0k} \); \( \beta_{0k} \sim N(0, 5) \) and \( \sigma^2 \sim \text{Half Normal}(1) \), where Half Normal(1) denotes a truncated \( N(0, 1) \) to cover the range (0, \( \infty \)).

Besides TRB and TEB, we also consider stand-alone subtrial analyses, referred to as the NB approach. The NB approach implements a Bayesian analysis of the basket trial without sharing information across subtrials. Each of the subtrials is analysed in isolation. We set a normal prior \( N\left(0, 10^2\right) \) for each subtrial-specific treatment \( \theta_k \). If there is no information to be shared across subtrials, estimation of \( \theta_k \) using the treatment effect and the difference in treatment responses is equivalent. Hence, we do not need to consider two versions of NB.

3.2 | Simulation scenarios

We simulate randomised basket trial data under nine plausible scenarios listed in Table 1. We assume a known “true” treatment effect for each subtrial. Scenarios 1–4 feature the consistency
TABLE 1 Description of simulation scenarios with specification of the ‘true’ treatment effect $\theta_k$ and mean response in the control group $\varphi_{Ck}$

| Scenario | Assumed treatment effect per subtrial (mean response in the control arm) | Description |
|----------|--------------------------------------------------------------------------|-------------|
| 1        | 0 (5.0) 0 (5.0) 0 (5.0) 0 (5.0) 0 (5.0) | No treatment effect across all subtrials. |
| 2        | 0.35 (5.0) 0.35 (5.0) 0.35 (5.0) 0.35 (5.0) 0.35 (5.0) | Consistent $\theta_k$'s across all (small effect size) subtrials. |
| 3        | 0.55 (5.0) 0.55 (5.0) 0.55 (5.0) 0.55 (5.0) 1.20 (5.0) | $\theta_k$'s are consistent in all except subtrial 5, where the true effect is higher; mean response in control is the same. |
| 4        | 0.80 (5.0) 0.80 (5.0) 0.80 (5.0) 0.80 (5.0) 0.80 (10.0) | Consistent $\theta_k$ (large effect size) across subtrials, but subgroup 5 has larger mean response in both experimental and control arms. |
| 5        | 1.32 (5.0) 0.72 (5.0) 0.66 (5.0) 0.44 (5.0) 0.72 (5.0) | High heterogeneity in $\theta_k$'s across all subtrials ($\rho = 0.1$). |
| 6        | 0.73 (5.0) 1.07 (5.0) 0.73 (5.0) 0.46 (5.0) 1.16 (5.0) | Low heterogeneity in $\theta_k$'s across subtrials ($\rho = 0.8$). |
| 7        | 1.32 (3.5) 0.72 (5.0) 0.66 (4.2) 0.41 (6.5) 0.72 (5.6) | Same as scenario 5 except the mean response of control varies across subtrials. |
| 8        | 0 (2.5) 0 (4.0) 0 (5.5) 0 (7.0) 0 (8.5) | Same as scenario 1 except that the mean responses differ across subtrials. |
| 9        | 0 (5.0) 0 (5.0) 0.15 (5.0) 0.87 (5.0) 1.86 (5.0) | A mixture scenario of null, small, and large treatment effects. |

of responses on both the experimental and control arms (not only the treatment effect) across subtrials. For scenarios 5 and 6, the set of treatment effects, that is, the values set for $\theta_k$, are generated from distinct multivariate normal distributions, with a high pairwise correlation coefficient ($\rho = 0.8$) for low heterogeneity and a low pairwise correlation coefficient ($\rho = 0.1$) for high heterogeneity, respectively. Furthermore, scenarios 5 and 6 feature the consistency of responses on the control arm only. Scenario 7 represents cases of inconsistency of responses on both experimental and control arms (not only the treatment effect) across subtrials. Scenario 8 is a case with consistent (null) treatment effects while scenario 9 has consistent responses on the control. Scenarios 1 and 8 are global null scenarios while scenario 9 is a mixed null scenario (some subtrials have non-null treatment effect). The simulation parameters are chosen to represent plausible settings informed by varying specific parameters in recent basket trials with continuous endpoints.

We additionally evaluate the performance of the borrowing of information approaches for settings with smaller ($n_{S1} = 12, n_{S2} = 18, n_{S3} = 10, n_{S4} = 16, n_{S5} = 14$) and larger ($n_{L1} = n_{L2} = 100, n_{L3} = 106, n_{L4} = 110, n_{L5} = 104$) subtrial sample sizes.
3.3 Performance evaluation metrics

We investigate the performance of the modelling approaches, focusing on the bias, mean square error (MSE), empirical standard error (EmpSE), coverage and width of the 95% posterior credible interval for \( \theta_k \), and the Monte-Carlo standard error (MCSE). These are given by

\[
\text{Bias}(\theta_k) \approx \frac{1}{b_{\text{sim}}} \sum_{l=1}^{b_{\text{sim}}} \hat{\theta}_l^k - \theta_k,
\]

\[
\text{MSE}(\theta_k) \approx \frac{1}{b_{\text{sim}}} \sum_{l=1}^{b_{\text{sim}}} (\hat{\theta}_l^k - \theta_k)^2,
\]

\[
\text{EmpSE}(\theta_k) = \sqrt{\frac{1}{b_{\text{sim}} - 1} \sum_{l=1}^{b_{\text{sim}}} (\hat{\theta}_l^k - \bar{\theta}_k)^2},
\]

\[
\text{Coverage}(\theta_k) = \mathbb{P}(\hat{\theta}_{k,\text{low}} \leq \theta_k \leq \hat{\theta}_{k,\text{upp}}) \approx \frac{1}{b_{\text{sim}}} \sum_{l=1}^{b_{\text{sim}}} 1(\hat{\theta}_{k,\text{low}}^l \leq \theta_k \leq \hat{\theta}_{k,\text{upp}}^l),
\]

\[
\text{Width of the 95\% credible interval, } \epsilon = \frac{1}{b_{\text{sim}}} \sum_{l=1}^{b_{\text{sim}}} (Q_{\theta_k}^l(0.975) - Q_{\theta_k}^l(0.025)),
\]

\[
\text{MCSE}(\theta_k) \approx \sqrt{\frac{1}{b_{\text{sim}}(b_{\text{sim}} - 1)} \sum_{l=1}^{b_{\text{sim}}} (\hat{\theta}_l^k - \bar{\theta}_k)^2},
\]

where \( l = 1, \ldots, b_{\text{sim}} \) indexes the simulated basket trials (i.e., \( b_{\text{sim}} \) is the total number of replicates performed for each simulation scenario); \( \theta_k \) in the above is the assumed value in the particular simulation scenario; \( \hat{\theta}_l^k \) is the estimate from one of the analysis models in Section 2 from the \( l \)th simulation replicate; \( \bar{\theta}_k = \frac{1}{b_{\text{sim}}} \sum_{l=1}^{b_{\text{sim}}} \hat{\theta}_l^k \) is the mean of the \( \hat{\theta}_l^k \) across all replicates; and \( Q_{\theta}(a) \), with \( 0 < a \leq 100 \), denotes the \( a \)th quantile of the posterior distribution for \( \theta \); the lower and upper limits of the coverage definition are the 2.5th and 97.5th percentiles.

For each subtrial \( k \), we also obtain the estimated posterior probability that \( \theta_k \) exceeds the threshold \( \theta_0 \) to estimate type I error-rates and power, \( \mathbb{P}(\theta_k > \theta_0) > 0.975 \). For illustration, we specify \( \theta_0 = 0.3 \). Of interest is the subtrial-wise and familywise type I error-rates and the power for identifying effective subgroups. The subtrial-wise type I error-rate is the proportion of trials, with respect to a given subtrial, with an erroneous go decision made for a given null scenario (scenario 1). The familywise error-rate is computed as the proportion of trials in which an incorrect go decision is made for any null hypothesis that is true. A similar approach to type I error-rate and power computation has been used previously (Hobbs & Landin, 2018; Zheng & Wason, 2022).

We simulate 10,000 randomised basket trials for each of the simulation scenarios. For each replicate, we use two parallel chains each with 10,000 MCMC iterations. The first 3000 samples in each chain are discarded as burn-in. Convergence of the MCMC chains was checked using Gelman-Rubin diagnostic plots. All Bayesian analysis models are implemented in R (version 3.6.3) using the Rjags package. Both R and JAGS functions to reproduce the simulation study are available on GitHub (https://github.com/oondijo/RandBasketTrials).
3.4 | Results

Figure 2 compares the TRB, TEB and NB analysis strategies in terms of the bias and MSE of the posterior mean estimators for the subtrial-specific treatment effects. Under the global null scenarios, there is similar and noticeably small bias across the three approaches although TRB and TEB have smaller MSE especially in scenario 1. In scenario 8 (same as scenario 1 except that mean response across control, $\phi_{Ck}$, is different across subtrials), the bias and MSE for TEB and TRB are slightly different in some scenarios, occasionally higher for TRB. The MSEs of the point estimates differ considerably; the NB approach produces the highest MSE in scenarios 1–3 and comparable MSE to the TRB approach in scenarios 7–9.

When all or most subtrials are consistent in the mean responses on both the experimental and control arms (scenarios 1–4), the TRB and TEB approaches have comparable bias, but the TRB modelling approach has smaller. An exception to this is scenario 3, where all $\phi_{Ck}$’s and $\theta_1 - \theta_4$ are similar, but $\theta_5$ is different; we observe here that TRB has slightly higher bias although the MSE are similar. In the mixed null scenario (scenario 9) where we anticipate very limited borrowing or heavy discounting of information, TEB produces smaller bias (similar to the NB approach) than TRB as well as the smallest MSE. We observe similar results under varying degrees of the treatment effect across the subtrials (scenarios 5 and 6). Overall, the TRB approach consistently provides smaller bias and MSE when the $\theta_k$ are small and similar across subtrials where response in control and experimental is also similar (for instance scenarios 1, 2 and 4), while TEB provides more accurate estimates when at least one of the true treatment effects and groupwise mean responses ($\phi_{Ck}$) are different (for instance scenarios 3, 8 and 9). In Figure S3, we demonstrate that...
these differences hold for the considered circumstances of smaller and larger basket trial sample sizes.

Figure 3 visualises the median width of the posterior credible intervals for the $\theta_k$ yielded by the respective Bayesian analysis models. In all scenarios, we see that either TEB or TRB have the narrowest 95% credible intervals compared to the NB approach. The differences are particularly noticeable in scenarios 1–4 (apart from subtrial 5 of scenario 3 where differences are marginal) where treatment effects are consistent across all subtrials. This is because borrowing of information leads to greater efficiency gains and more precise posterior estimates of the treatment effect when subtrials are more commensurate. Although there are no observable differences in scenario 7 where both $\varphi_{CK}$ and $\theta_k$ are different across subtrials, it clear from Figure S6 that TRB works best under small sample sizes. However, in this scenario there are no differences across when the sample size is large, another indication that multiple metrics are necessary to distinguish TRB and TEB in that case.

For the global null scenario 1, we notice that TRB leads to more precise estimates, due to commensurability of mean responses of the control arm across all subtrials. However, when the mean response in control is different across subtrials (scenario 8), TEB gives the smallest posterior credible intervals across all subtrials. It is interesting to note that in circumstances of a small overall basket trial sample size, TRB gives the narrowest 95% credible intervals across nearly all...
scenarios; but for circumstances of a large sample size, TEB gives far narrower posterior credible intervals for the treatment effect in nearly all scenarios (Figure S6).

We also compared the two modelling strategies for borrowing of information based on their statistical power, type I error-rate, and familywise error-rate. Figure 4 compares the statistical power under small (S) or large (L) sample sizes. The first thing to note is the increase in statistical power to demonstrate a treatment benefit in a particular subtrial as we leverage information from other subtrials. Both TRB and TEB, achieve higher statistical power in almost all scenarios compared to the NB approach.

**FIGURE 4** A comparison of the Bayesian analysis models for randomised basket trials, in terms of the frequentist power, under small (S – \( n_{S1} = 12, n_{S2} = 18, n_{S3} = 10, n_{S4} = 16, n_{S5} = 14 \)) or large (L – \( n_{L1} = 100, n_{L2} = 106, n_{L3} = 110, n_{L4} = 104 \)) sample sizes. Frequentist power is defined as the proportion of trials with a correct go decision under \( H_{1k} : \theta_k > \theta_0 \). Horizontal dashed line represents 80% power. [Colour figure can be viewed at wileyonlinelibrary.com]
We observe very low statistical power in scenario 2 \( (\theta_k = 0.35 \text{ for all } k) \), which is expected because go decisions based on \( \mathbb{P}(\theta_k > 0.3) > 0.975 \) are made to be very unlikely. In scenarios 3 and 4, where treatment effects are commensurate across most subtrials (except subtrial 5), we observe that TRB leads to higher statistical power over TEB and no borrowing. As the treatment effect gets larger (scenario 4), there is an increase in power, apart from subtrial 5 which is different in scenarios 3 and 4. Notably, the differences between the borrowing approaches are more pronounced when the subtrial sample sizes are small. For instance, in scenarios 3 and 4, TRB gives 0.3%–11% and 2%–15% higher power than TEB on average in subtrials 1–4 when sample sizes are small. However, when the sample size is larger, TEB outperforms TRB marginally (<2%) in nearly all scenarios. For the mixed global null (scenario 9), TEB yields at least 13% higher power than TRB in subtrial 4. Scenarios 5 and 7 mimic situations where the true \( \theta_k \) are highly heterogeneous across all subtrials. Scenario 7 is further featured by the mean responses of the control group differing across the subtrials. We observe that TRB achieves the highest statistical power in scenario 5, but this is not always the case in scenario 7. In subtrial 5, TEB outperforms TRB by an increase of 14% in scenario 5, while in scenario 7 the gap is 10% in the opposite direction’ (Figure 4-S).

Principally, TRB achieves higher power compared to TEB when there is high commensurability across all subtrials, the mean response in control is similar across subtrials and the basket trial sample size is small. However, when either the treatment effects, and/or the mean response in the control/experimental arm are inconsistent across subtrials (scenarios 5–7 and 9), TRB still outperforms TEB in small sample sizes; in scenario 7 (subtrial 5) and scenario 9 (subtrial 4), TEB has higher power by at least 10% and 13%, respectively. When the sample size is larger, we notice TEB has marginal absolute gain in power in all scenarios over TRB. The observation that the power is high and similar across the three approaches in Figure 4 (bottom panel) is entirely expected. This is because when a treatment truly works, higher power will be observed as sample size gets larger for the same specified effect sizes. The low power in subtrial 4 in scenarios 5, 6 and 7 is because \( \theta_4 \) is assumed to be small (see Table 1) hence \( \Pr(\theta_k > \theta_0) > 0.975 \) is expected to be low.

Table 2 quantifies the coverage probability of the posterior distributions based on the respective Bayesian models, as we vary the sample size. Both TEB and TRB approaches guarantee high coverage that is well above the target coverage (95%) in most scenarios. However, we notice substantial differences in the coverage for cases of small subtrial sample sizes. For example, in the global null scenario where the mean response in control is different across subtrials (scenario 8), TEB gives higher coverage for \( \theta_k \) (which is also above the target) relative to TRB (approximately 4% in some cases). When the sample size is large, similar conclusions apply except that we observe smaller differences of 1%–2%. Further, TEB and TRB have similar coverage probability for \( \theta_k \) (differences of <1%) in the null scenario 1 where the mean response in the control arm was assumed to be similar across subtrials. When treatment effects are assumed to be heterogeneous across subtrials (with subtrials inconsistent in the mean responses of the experimental or control arm, or both), we observe mixed results where TRB can perform poorest of the three (subtrial 1 of scenarios 5 and 7).

In scenario 7 where the true \( \theta_k \) are highly heterogeneous across all subtrials and the true mean of the control arm varies across subgroups, all approaches produce coverage that is below the target (95%) in nearly all subtrials (Table 2). For instance, in subtrial 3, TEB gives coverage probability of nearly 3% higher than TRB when the overall sample size is small, but which is also closest to the target. Overall, TEB gives higher coverage probability in most scenarios when the sample size is large.
| Scenario | Borrowing approach | Small sample size | Large sample size |
|----------|--------------------|-------------------|-------------------|
|          | 1      | 2      | 3      | 4      | 5      | 1      | 2      | 3      | 4      | 5      |
| 1        | TRB    | 0.9706 | 0.9696 | 0.9708 | 0.9676 | 0.9667 | 0.9757 | 0.9798 | 0.9743 | 0.9764 | 0.9752 |
|          | TEB    | 0.9755 | 0.9776 | 0.9692 | 0.9731 | 0.9711 | 0.9716 | 0.9749 | 0.9721 | 0.974  | 0.978  |
|          | NB     | 0.9342 | 0.9422 | 0.9411 | 0.9438 | 0.9375 | 0.9505 | 0.9492 | 0.9471 | 0.9442 | 0.9526 |
| 2        | TRB    | 0.9710 | 0.9694 | 0.9712 | 0.9683 | 0.9669 | 0.9770 | 0.9791 | 0.974  | 0.9737 | 0.9753 |
|          | TEB    | 0.9587 | 0.9586 | 0.9561 | 0.9567 | 0.9524 | 0.9718 | 0.9733 | 0.9709 | 0.9714 | 0.9689 |
|          | NB     | 0.9341 | 0.9421 | 0.9410 | 0.9438 | 0.9374 | 0.9505 | 0.9492 | 0.9471 | 0.9442 | 0.9526 |
| 3        | TRB    | 0.9655 | 0.9646 | 0.9660 | 0.9626 | 0.9212 | 0.9658 | 0.9700 | 0.9644 | 0.9665 | 0.9515 |
|          | TEB    | 0.9575 | 0.9586 | 0.9558 | 0.9553 | 0.9384 | 0.9714 | 0.9726 | 0.9693 | 0.9695 | 0.9627 |
|          | NB     | 0.9340 | 0.9421 | 0.9411 | 0.9439 | 0.9373 | 0.9505 | 0.9493 | 0.9471 | 0.9442 | 0.9525 |
| 4        | TRB    | 0.9680 | 0.9697 | 0.9671 | 0.9634 | 0.9482 | 0.9695 | 0.9721 | 0.9690 | 0.9669 | 0.9574 |
|          | TEB    | 0.9596 | 0.9627 | 0.9579 | 0.9557 | 0.9511 | 0.9672 | 0.9699 | 0.9679 | 0.9671 | 0.9623 |
|          | NB     | 0.9340 | 0.9421 | 0.9410 | 0.9439 | 0.9373 | 0.9505 | 0.9493 | 0.9471 | 0.9442 | 0.9525 |
| 5        | TRB    | 0.9107 | 0.9620 | 0.9622 | 0.9303 | 0.9603 | 0.9430 | 0.9640 | 0.9603 | 0.9419 | 0.9621 |
|          | TEB    | 0.9359 | 0.9585 | 0.9546 | 0.9372 | 0.9519 | 0.9577 | 0.9645 | 0.9608 | 0.9504 | 0.9649 |
|          | NB     | 0.9341 | 0.9421 | 0.9410 | 0.9439 | 0.9374 | 0.9504 | 0.9493 | 0.9471 | 0.9442 | 0.9525 |
| 6        | TRB    | 0.9575 | 0.9483 | 0.9548 | 0.9286 | 0.9335 | 0.9542 | 0.9535 | 0.9557 | 0.9466 | 0.9502 |
|          | TEB    | 0.9548 | 0.9489 | 0.9509 | 0.9349 | 0.9430 | 0.9607 | 0.9615 | 0.9590 | 0.9546 | 0.9623 |
|          | NB     | 0.9340 | 0.9422 | 0.9410 | 0.9439 | 0.9373 | 0.9505 | 0.9493 | 0.9471 | 0.9442 | 0.9525 |
| 7        | TRB    | 0.9234 | 0.9556 | 0.9181 | 0.9405 | 0.9307 | 0.9375 | 0.9514 | 0.9394 | 0.9428 | 0.9461 |
|          | TEB    | 0.9089 | 0.9548 | 0.9458 | 0.9251 | 0.9438 | 0.9427 | 0.9532 | 0.9510 | 0.9352 | 0.9541 |
|          | NB     | 0.9340 | 0.9421 | 0.9413 | 0.9444 | 0.9377 | 0.9502 | 0.9493 | 0.9470 | 0.9443 | 0.9526 |
| 8        | TRB    | 0.9275 | 0.9475 | 0.9505 | 0.9486 | 0.9494 | 0.9418 | 0.9495 | 0.9540 | 0.9502 | 0.9570 |
|          | TEB    | 0.9656 | 0.9716 | 0.9639 | 0.9633 | 0.9654 | 0.9594 | 0.9679 | 0.9659 | 0.9657 | 0.9698 |
|          | NB     | 0.9365 | 0.9448 | 0.9365 | 0.9435 | 0.9385 | 0.9516 | 0.9493 | 0.9460 | 0.9435 | 0.9514 |
| 9        | TRB    | 0.9543 | 0.9533 | 0.9555 | 0.9385 | 0.9409 | 0.9559 | 0.9578 | 0.9584 | 0.9534 | 0.9567 |
|          | TEB    | 0.9617 | 0.9647 | 0.9553 | 0.9501 | 0.9473 | 0.9613 | 0.9647 | 0.9593 | 0.9594 | 0.9613 |
|          | NB     | 0.9342 | 0.9422 | 0.9411 | 0.9439 | 0.9376 | 0.9505 | 0.9492 | 0.9471 | 0.9442 | 0.9525 |

Note: Small and large sample sizes:
- $n_{S1} = 12$, $n_{S2} = 18$, $n_{S3} = 10$, $n_{S4} = 16$, $n_{S5} = 14$, $n_{L1} = 100$, $n_{L2} = 106$, $n_{L3} = 110$, $n_{L4} = 104$.
- Abbreviations: NB, No Borrowing; TEB, Treatment Effect Borrowing; TRB, treatment response borrowing.
In Figures S1 and S5, we present graphical results of a comparison of the empirical standard error (EmpSE) from the Bayesian modelling approaches. We also present additional results on the subtrial-wise and familywise error-rate control for each of the three analysis approaches in Table S2. We notice that TRB has the smallest EmpSE when the sample size is small but as this gets larger, TEB has the smallest EmpSE. Furthermore, the error rates (FWER and type I error) from the borrowing strategies are smaller than those of the no borrowing strategy.

4 | APPLICATION TO CASE STUDY DATA SETS

4.1 | Data description

We apply the TRB, TEB and NB approaches to Bayesian analyses of two case study data sets. Currently, there are no published randomised basket trials with a continuous primary outcome, based on our assessment of recent systematic reviews, a search of clinicaltrials.gov, and an ongoing systematic review in our group. We instead identified trials with continuous primary outcomes whose design matches that of a randomised basket trial to utilise as case studies. The first case study data is drawn from a trial of 60 participants randomised between different structured exercise and diet interventions. Participants were first randomly assigned to one of three different exercise regimes, and within each exercise regime subgroup, they were further randomised to two dietary interventions. Although not labelled as a basket trial, this could be considered as a randomised basket trial with three subtrials defined by the exercise regime assigned. Each subtrial then comprises the 20 patients further randomised to either diet 1 or diet 2. Both diet 1 and diet 2 are similar across the subtrials. The post-intervention clinical outcome of interest was the pulse rate, a continuous measurement was taken at either three or four timepoints from baseline for each participant. We analyse the pulse rate data from the end of the trial. This data is publicly available here (Repeated measures analysis with R. UCLA: Statistical Consulting Group, n.d.).

The second case study data set is drawn from a multi-centre, randomised, double-blind, placebo-controlled, proof-of-concept study of iscalimab in patients with primary Sjogren’s syndrome (Fisher et al., 2020). Iscalimab is a novel anti-CD40 monoclonal antibody that targets the CD40–CD154-mediated T cell–B cell interactions which have been shown to be responsible for aberrant lymphocyte activation in inflamed tissue (Fisher et al., 2020). The trial was conducted in 10 sites across five different countries, United Kingdom, Germany, Switzerland, Hungary, and United States, recruiting patients aged 18–75 years. This trial had two cohorts. In cohort 1, 12 patients were randomised to receive subcutaneous iscalimab (3 mg/kg) or placebo for 12 weeks, after which patients on placebo would receive iscalimab up to week 32 (end of the trial). Cohort 2 (with 32 patients) followed a similar design as cohort 1 except that a higher dose of iscalimab (10 mg/kg) was administered. As such, this could be viewed as a randomised basket trial with two subtrials. The primary outcome was the ESSDAI score, which has a maximum possible score of 123 points, with higher scores indicating more severe symptoms. The minimum clinically important improvement in ESSDAI has been published as three points. Our interest is the comparison of ESSDAI scores between iscalimab and placebo patients at week 12. For this trial, the process to obtain individual level patient data would have been (lasting several months), and we resorted to obtaining all relevant information on the distribution of ESSDAI scores for each treatment arm in the two subtrials (mean ESSDAI score for baseline and at each timepoint, range of scores, SDs) from the chief investigator and the published paper. Using this information, we simulate data for 24 patients in subtrial 1 and 64 patients in subtrial 2. Note
that one key assumption in our analysis is that the difference in dosage of Iscalimab between the cohorts does not influence the outcomes. We view that as appropriate as our goal is to illustrate how different the results of the analyses can be, and what the differences may look like in a real trial.

### 4.2 Analysis of case studies

Table 3 presents the results of the three analyses (TRB, TEB, NB) of the two case study datasets. Our results demonstrate that TRB and TEB can give substantially different point estimates for the $\theta_k$. This is a consequence of the differential information borrowing from complementary subtrials.

For case study 1, we notice that TEB has the smallest estimates in subtrials 1 and 2 and the largest in subtrial 3. Considering the posterior probabilities $\mathbb{P}(\theta_k > 3|\text{data})$, we see that for subtrial 2, $\mathbb{P}(\theta_2 > 3|\text{data}) = 0.615, 0.852$ and 0.40 for no borrowing, TRB and TEB respectively (see Table 4). Thus, if a typical threshold probability to inform decision-making such as $\zeta = 0.9$ was specified (i.e., decision-making was based on $\mathbb{P}(\theta_k > 3|\text{data}) > \zeta$), we would arrive at identical no-go conclusions for all subtrials since all the posterior probabilities $\mathbb{P}(\theta_k > 3|\text{data})$ are less than $\zeta = 0.9$. However, if a smaller value such as $\zeta = 0.8$ is chosen for some trials, for example as in the case of a small proof-of-concept study, we can see the TRB and TEB analyses would then lead to different conclusions in subtrial 2 of case study 1; there is an 85.2% chance of recommending the intervention by the TRB analysis, while the NB and TRB analyses, this probability is 61.5% and 40%, respectively.

As in case study 1, we notice that the treatment effects for case study 2 are different between the NB approach and the two modelling strategies TEB and TRB (Table 3). Here TRB and TEB produce comparable results, both for the treatment effects and the 95% credible intervals, an observation more likely when the groupwise mean responses are comparable across subtrials. However, we notice that the posterior probabilities to inform go/no-go decisions were slightly different (0.365 vs. 0.468) (Table 4). While no more than three subtrials are analysed here and our simulation study had five subtrials, there is no expectation for the number of subtrials to have an impact. In principle, the distributional discrepancy approach here allows for borrowing depending only on the commensurability of information across subtrials.

| Subtrial | Case study 1: All $n_k = 20$ | Modelling strategy | 1 | 2 | 3 |
| --- | --- | --- | --- | --- | --- |
| TRB | 2.50 ($-2.11, 7.07$) | 1.73 ($-2.81, 5.97$) | 1.31 ($-0.12, 5.46$) | 4.0 ($-2.83, 10.75$) | 6.27 ($-0.29, 12.30$) | 2.23 ($-3.63, 7.85$) | 23.59 (16.39, 30.01) | 27.45 (19.96, 34.98) | 33.27 (23.91, 48.71) |
| TEB | 0.21 ($-2.26, 2.72$) | $-0.52$ ($-2.60, 1.55$) | $-0.46$ ($-2.61, 1.74$) | $-3.10$ ($-4.65, -1.46$) | $-2.73$ ($-4.18, -1.19$) | $-2.59$ ($-4.21, -0.82$) | — | — | — |
| Case study 2: $n_1 = 24, n_2 = 64$ | NB | TRB | TEB |

Notes: Data are the posterior mean treatment effects, $\theta_k$ (95% credible interval). NB refers to the analysis strategy without borrowing; TEB is where borrowing is performed based on the commensurability of the treatment effects $\theta_k$; TRB is where borrowing is performed based on the commensurability of the mean response across the respective treatment arms $\phi_k$. |
TABLE 4 Posterior probability that $\theta_k$ exceeds a pre-specified threshold, $P(\theta_k > \theta | \text{data})$

| Case study                        | Modelling strategy | 1     | 2     | 3     |
|----------------------------------|--------------------|-------|-------|-------|
| Case study 1: All $n_k = 20$     | NB                 | 0.419 | 0.615 | 0.999 |
|                                  | TRB                | 0.275 | 0.852 | 1.000 |
|                                  | TEB                | 0.239 | 0.398 | 1.000 |
| Case study 2: $n_1 = 24$, $n_2 = 64$ | NB                 | 0.016 | 0.247 | —     |
|                                  | TRB                | 0.039 | 0.365 | —     |
|                                  | TEB                | 0.028 | 0.468 | —     |

Note: In case study 1, $\theta = 5$ and $\theta = 3$ for case study 2.
Abbreviations: NB, No Borrowing; TEB, Treatment Effect Borrowing; TRB, treatment response borrowing.

5 DISCUSSION

The use of basket trials has grown substantially in the last decade, especially in early-phase trial settings targeting rare diseases (Hazim & Prasad, 2018; Park et al., 2019). Randomised basket trials provide higher-quality evidence of treatment benefit over the control relative to non-randomised basket trials and will increasingly characterise the future landscape of phase II and phase III basket trials given the enormous weight for randomised evidence in drug approvals (Prasad & Oseran, 2015; Saad et al., 2017). Some recent ongoing randomised basket studies include NCT04498962 and NCT04988087.

In this article, we sought to understand the best modelling strategy to borrow information in the Bayesian analysis of a randomised basket trial with a continuous endpoint, motivated in part by our desire to identify a suitable analysis approach for the OACS randomised basket trial. We have proposed a TRB approach focusing on the borrowing by treatment arms (groupwise responses) and compared its performance to TEB as well as the NB approach. Specifically, the TRB approach assesses how similar the responses to a specific treatment are across subtrials, in contrast to the TEB that assesses how similar the experimental treatment benefit over the control (treatment effects) is across subtrials, for leveraging information from commensurate subtrials.

Our work is relevant to debates on the arm-based (Hong et al., 2016) against contrast-based (Begg & Pilote, 1991) frameworks for performing meta-analysis of randomised controlled trials with at least two treatment arms (Dias & Ades, 2016). The newly proposed TRB offers added flexibility to implementing borrowing of information within basket trials. Firstly, data from complementary subtrials that are commensurate with the contemporary subtrial by either the experimental or control arm can be leveraged. Secondly, incommensurate data can be down weighted to a different extent by treatment arms. Our results are congruent with studies comparing arm-based and contrast-based models; estimates from arm-based models were slightly larger than those obtained from contrast-based models, but with smaller SEs (Karahalios et al., 2022).

Both TRB and TEB are based on the commensurate predictive prior approach to quantifying the degree of commensurability; synthesis weights are specified as a function of the pairwise Hellinger distance and assigned to the corresponding complementary subtrial. When fitting the Bayesian models with a TRB or TEB approach, we placed a spike and slab prior on the precision parameter that gauges the degree of borrowing, but alternative priors could be considered. Zheng et al. used a two-component gamma mixture prior instead for analytic tractability in
the exact Bayesian inference of treatment effects when leveraging pre-experimental data from multiple sources (Zheng et al., 2020). Using a two-component gamma mixture prior brings about the convenience of analytic tractability and would particularly facilitate the derivation of a sample size formula for the trial design. The two choices of prior on the commensurate parameter would mean different interpretation/implication of a specified value for the prior probability of incommensurability. For instance, setting $\omega_{kk}^*$ to a value as small as 0.3 may suggest a priori strong borrowing in the models coupled with a spike and slab prior, while the degree of borrowing might be attenuated with a two-component gamma mixture prior. Indeed, we expect differences associated with these two prior specifications, and it is our plan to perform another comprehensive simulation study to formally compare the marginal changes in the degree of borrowing induced by these two prior specifications.

Our simulation results show that permitting borrowing of information in the analysis of randomised basket trials offers substantial gains in efficiency over independent analysis of subtrials (NB) as shown in previous simulation studies (Jin et al., 2020; Zheng & Wason, 2022). The proposed approaches to TRB and TEB differ most widely in performance when subtrial sample sizes are small, with differences depending on the similarity of treatment effects or treatment responses within each subtrial. In circumstances of small subtrial sizes, TRB typically outperforms TEB, especially when either the treatment effects or the mean response in control (or both) are consistent across subtrials. On the contrary, TEB performs better when there is considerable heterogeneity in treatment effects and/or the mean response in either the control or experimental arm across subtrials. Our findings indicate that TEB is a more robust approach that provides more precise estimates with better coverage probability and marginally higher power over TRB in nearly all scenarios when the trial sample size is sufficiently large. We also applied both TRB and TEB to the analysis of two case study data sets, showing that a trial may arrive at a different conclusion based on the chosen analysis approach. An important note is that our choice of case study 1 is different from the common basket design (Figure 1). This helps illustrate that basket trials are equally applicable to non-oncology settings and non-pharmaceutical interventions and are not restricted to biomarker-guided pharmaceutical interventions as is common in oncology.

While we have considered a single Bayesian analysis methodology in Section 2, the concept underlying TRB can be implemented using any proposed Bayesian hierarchical method for basket trials. For illustrative purposes, we have focused on a setting where patients are equally randomised to a single experimental treatment or control within each subtrial, with a common continuous endpoint. A possible extension to our proposed methodology is thus to consider binary or time to event outcomes like progression-free survival. In a binary setting, we can assume response in each subgroup follows a binomial distribution, and we can model the log-odds of response. Borrowing in this case would be based on the similarity of the difference in log-odds per subgroup for TEB, while for TRB we account for the similarity of the log-odds separately for control and treatment arms across subgroups. The log-odds can be assumed to follow a normal distribution, with normal hyperpriors and the Bayesian hierarchical modelling proceeds as in Section 2.1 and 2.2.

There are also several avenues to enhance the practicality of our proposed methodology (particularly the TRB strategy) in precision medicine trials. One future direction is to explore its performance in a basket trial setting where patients are randomised to at least two experimental treatment groups plus control in all or some of the subtrials. Another type of randomised basket design could consider incorporating response-adaptive randomization (Lin et al., 2021; Ventz et al., 2017); more specifically, the randomization ratio could be adapted at an interim analysis.
to favour a treatment with the highest efficacy, as estimated by our proposed TRB. Ventz and colleagues recently developed a class of Bayesian response adaptive basket designs and demonstrated the advantages of adapting randomization with borrowing over not borrowing (Ventz et al., 2017). Furthermore, it is also possible to incorporate early stopping for futility or efficacy. For example, we may decide to drop subgroup $k$ if the posterior probability that the treatment effect $\theta_k$ exceeds a prespecified threshold $\theta_0$ is low, that is, $\Pr(\theta_k > \theta_0) \leq \delta$ for TRB. Similarly, if the posterior probability is high, the subgroup can be stopped for efficacy. An important note here is that the decision-making criteria using $\theta_k$ is the same TRB or TEB as this decision is made post-computation of the treatment effects.

The use of TRB is in no way restricted to basket trial designs. Our proposed methodology could potentially be used in umbrella trials with a common control in each subtrial. The Lung-MAP umbrella trial (Ferrarotto et al., 2015) is one such example where this may be plausible. Here, information borrowing across subtrials would be implemented for the control arm only. Relevant to this context, Zang et al. proposed a Bayesian hierarchical model to borrow information across subgroups receiving standard-of-care only in a Bayesian adaptive marker-stratified design (Zang et al., 2019). By using the TRB proposed in this paper, we could expect a more precise estimate of the mean response in each control arm to be achieved, which would aid the estimation of the treatment effect specific to each subgroup.

We caution that despite the efficiencies of borrowing across subgroups in basket trials, this should always be motivated by a strong biological and clinical rationale. For instance, when the control (e.g., the standard of care treatment) is different in certain subgroups, the strength of scientific and clinical evidence underpinning such borrowing is weak. This is because different drugs may likely have potentially different mechanisms of action and may elicit different responses across subgroups. As such, borrowing is more justifiable only across similar controls and TRB would be suitable if the experimental arm is the same, but the control arm is not. We also note that Bayesian simulations undertaken here demand high computational requirements. The computational time for a single scenario with 10,000 simulation replicates, two MCMC chains each with 10,000 iterations for TEB/TRB was 3.4 to five hours on average when parallelising on 44 cores in a high-performance computing cluster node (2.2 GHz Intel Xeon E5-2699 v4 processor). For the NB approach, this takes approximately 10 min.

Finally, we conclude that borrowing of information in randomised basket trials offers considerable efficiency gains but must be implemented carefully depending on the potential heterogeneity in effects across subgroups and the subtrial sample sizes. Consequently, investigators and statisticians must thoroughly investigate any borrowing approach based on plausible trial scenarios.

ACKNOWLEDGEMENTS
Dr Zheng’s contribution to this article was supported by Cancer Research UK (RCCPDF\100008).

CONFLICT OF INTEREST
None to be declared.

DATA AVAILABILITY STATEMENT
Case study data alongside relevant analysis code are available at https://github.com/oondijo/RandBasketTrials.
REFERENCES

Asano, J. & Hirakawa, A. (2020) A Bayesian basket trial design accounting for uncertainties of homogeneity and heterogeneity of treatment effect among subpopulations. *Pharmaceutical Statistics*, 19, 975–1000. Available from: https://doi.org/10.1002/pst.2049

Begg, C.B. & Pilote, L. (1991) A model for incorporating historical controls into a meta-analysis. *Biometrics*, 47, 899–906.

Berry, S.M., Broglio, K.R., Groschen, S. & Berry, D.A. (2013) Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. *Clinical Trials*, 10, 720–734. Available from: https://doi.org/10.1177/1740774513497539

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, 2717–2732. Available from: https://doi.org/10.1177/0962280220910186

Chu, Y. & Yuan, Y. (2018a) A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clinical Trials*, 15, 149–158. Available from: https://doi.org/10.1177/1740774517755122

Chu, Y. & Yuan, Y. (2018b) BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67, 723–740. Available from: https://doi.org/10.1111/rssc.12255

de Benedetti, F.,Gattorno, M.,Anton, J., Ben-Chetrit, E.,Frenkel, J., Hoffman, H.M. et al. (2018) Canakinumab for the treatment of autoinflammatory recurrent fever syndromes. *The New England Journal of Medicine*, 378, 1908–1919. https://doi.org/10.1056/nejmoa1706314

Dias, S. & Ades, A.E. (2016) Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods*, 7, 23. https://doi.org/10.1002/1526-9000(201601)7:1<23::AID-RSSD2.0.CO;2-C

Drilon, A., Laetsch, T.W., Kummar, S., DuBois, S.G., Lassen, U.N., Demetri, G.D. et al. (2018) Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *The New England Journal of Medicine*, 378, 731–739. Available from: https://doi.org/10.1056/NEJMoa1714448

Ferrarotto, R., Redman, M.W., Gandara, D.R. & Herbst, R.S. (2015) Lung-MAP framework, overview, and design principles. *Chinese Clinical Oncology*, 4, 2–7. Available from: https://doi.org/10.3978/j.issn.2304-3865.2015.09.02

Fisher, B.A., Szanto, A., Ng, W.F., Bombardieri, M., Posch, M.G., Papas, A.S. et al. (2020) Assessment of the anti-CD40 antibody icsalimab in patients with primary Sjögren’s syndrome: a multicentre, randomised, double-blind, placebo-controlled, proof-of-concept study. *The Lancet Rheumatology*, 2, e142–e152. Available from: https://doi.org/10.1016/S2665-9913(19)30135-3

Grayling, M.J., Bigirimurame, T., Cherlin, S., Ouma, L., Zheng, H. & Wason, J.M.S. (2021) Innovative trial approaches in immune-mediated inflammatory diseases: current use and future potential. *BMC Rheumatology*, 5, 21. Available from: https://doi.org/10.1186/s41927-021-00192-5

Hazim, A. & Prasad, V. (2018) A pooled analysis of published, basket trials in cancer medicine. *European Journal of Cancer*, 101, 244–250. Available from: https://doi.org/10.1016/j.ejca.2018.06.035

Heinrich, M.C., Joensuu, H., Demetri, G.D., Corless, C.L., Apperley, J., Fletcher, J.A. et al. (2008) Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitivetyrosine kinases. *Clinical Cancer Research*, 14, 2717–2725. Available from: https://doi.org/10.1158/1078-0432.CCR-07-4575

Hobbs, B.P. & Landin, R. (2018) Bayesian basket trial design with exchangeability monitoring. *Statistics in Medicine*, 37, 3557–3572. Available from: https://doi.org/10.1002/sim.7893

Hobbs, B.P., Sargent, D.J. & Carlin, B.P. (2012) Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, 7, 639–674. Available from: https://doi.org/10.1214/12-BA722
Hong, H., Chu, H., Zhang, J. & Carlin, B.P. (2016) A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. Research Synthesis Methods, 7, 6–22. Available from: https://doi.org/10.1002/JRSM.1153

Hyman, D.M., Puzanov, I., Subbiah, V., Faris, J.E., Chau, I., Blay, J.Y. et al. (2015) Vemurafenib in multiple non-melanoma cancers with BRAF V600 mutations. The New England Journal of Medicine, 373, 726–736. Available from: https://doi.org/10.1056/nejmoa1502309

Jin, J., Riviere, M.K., Luo, X. & Dong, Y. (2020) Bayesian methods for the analysis of early-phase oncology basket trials with information borrowing across cancer types. Statistics in Medicine, 39, 3459–3475. Available from: https://doi.org/10.1002/SIM.8675

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. Available from: https://doi.org/10.1200/po.19.00194

Karahalios, A., McKenzie, J.E. & White, I.R. (2022) Contrast-based and arm-based models for network meta-analysis network meta-analysis (NMA) BT. In: Evangelou, E. & Veroniki, A.A. (Eds.) Meta-research: methods and protocols. New York: Available from:: Springer, pp. 203–221. https://doi.org/10.1007/978-1-0716-1566-9_13

Lasch, F., Weber, K., Chao, M.M. & Koch, A. (2017) A plea to provide best evidence in trials under sample-size restrictions: the example of pioglitazone to resolve leukoplakia and erythroplakia in Fanconi anemia patients. Orphanet Journal of Rare Diseases, 12, 1–6. Available from: https://doi.org/10.1186/S13023-017-0655-8

Lemery, S., Keegan, P. & Pazdur, R. (2017) First FDA approval agnostic of cancer site — when a biomarker defines the indication. New England Journal of Medicine (NEJM/MMS), 377, 1409–1412. Available from: https://doi.org/10.1056/nejmp1709968

Li, B.T., Makker, V., Buonocore, D.J., Offin, M.D., Olah, Z.T., Panora, E. et al. (2018) A multi-histology basket trial of ado-trastuzumab emtansine in patients with HER2 amplified cancers. Journal of Clinical Oncology, 36, 2502. Available from: https://doi.org/10.1200/jco.2018.36.15_suppl.2502

Lin, R., Thall, P.F. & Yuan, Y. (2021) A phase I–II basket trial design to optimize dose-schedule regimes based on delayed outcomes. Bayesian Analysis, 16, 179–202. Available from: https://doi.org/10.1214/20-BA1205

Mitchell, T.J. & Beauchamp, J.J. (1988) Bayesian variable selection in linear regression. Journal of the American Statistical Association, 83, 1023–1032. Available from: https://doi.org/10.1080/01621459.1988.10478694

Moore, C.L., Turkova, A., Mujuru, H., Kekitiinwa, A., Lugemwa, A., Kityo, C.M. et al. (2021) ODYSSEY clinical trial design: a randomised global study to evaluate the efficacy and safety of dolutegravir-based antiretroviral therapy in HIV-positive children, with nested pharmacokinetic sub-studies to evaluate pragmatic WHO-weight-band based dolutegravir dosing. BMC Infectious Diseases, 21, 5. Available from: https://doi.org/10.1186/s12879-020-05672-6

Neuenschwander, B., Wandel, S., Roychoudhury, S. & Bailey, S. (2016) Robust exchangeability designs for early phase clinical trials with multiple strata. Pharmaceutical Statistics, 15, 123–134. Available from: https://doi.org/10.1002/pst.1730

Park, J.J.H., Siden, E., Zoratti, M.J., Dron, L., Harari, O., Singer, J. et al. (2019) Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. Trials, 20, 572. Available from: https://doi.org/10.1186/s13063-019-3664-1

Pohl, M., Krisam, J. & Kieser, M. (2021) Categories, components, and techniques in a modular construction of basket trials for application and further research. Biometrical Journal, 63, 1159–1184. Available from: https://doi.org/10.1002/bimj.202000314

Prasad, V. & Oseran, A. (2015) Do we need randomised trials for rare cancers? European Journal of Cancer, 51, 1355–1357. Available from: https://doi.org/10.1016/J.EJCA.2015.04.015

Psioda, M.A., Xu, J., Jiang, Q., Ke, C., Yang, Z. & Ibrahim, J.G. (2019) Bayesian adaptive basket trial design using model averaging. Biostatistics Available from; 22, 19–34. https://doi.org/10.1093/biostatistics/kxz014

Repeatead measures analysis with R. UCLA: Statistical Consulting Group (n.d.). Available at: https://stats.oarc.ucla.edu/r/seminars/repeated-measures-analysis-with-r/ [Accessed 1st November 2021].

Saad, E.D., Paolletti, X., Burzykowski, T. & Buyse, M. (2017) Precision medicine needs randomized clinical trials. Nature Reviews Clinical Oncology, 14, 317–323. Available from: https://doi.org/10.1038/nrclinonc.2017.8
Siden, E.G., Park, J.J., Zoratti, M.I., Dron, L., Harari, O., Thorlund, K. et al. (2019) Reporting of master protocols towards a standardized approach: a systematic review. *Contemporary Clinical Trials Communications*, 15, 100406. Available from: https://doi.org/10.1016/j.conctc.2019.100406

Simon, R., Geyer, S., Subramanian, J. & Roychowdhury, S. (2016) The Bayesian basket design for genomic variant-driven phase II trials. *Seminars in Oncology*, 43, 13–18. Available from: https://doi.org/10.1053/j.seminoncol.2016.01.002

Slosberg, E.D., Kang, B.P., Peguero, J., Taylor, M., Bauer, T.M., Berry, D.A. et al. (2018) Signature program: a platform of basket trials. *Oncotarget*, 9, 21383–21395. Available from: https://doi.org/10.18632/oncotarget.25109

Tao, J.J., Schram, A.M. & Hyman, D.M. (2018) Basket studies: redefining clinical trials in the era of genome-driven oncology. *Annual Review of Medicine*, 69, 319–331. Available from: https://doi.org/10.1146/annurev-med-062016-050343

Ventz, S., Barry, W.T., Parmigiani, G. & Trippa, L. (2017) Bayesian response-adaptive designs for basket trials. *Biometrics*, 73, 905–915. Available from: https://doi.org/10.1111/biom.12668

Woodcock, J. & LaVange, L.M. (2017) Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine (NEJM/MMS)*, 377, 62–70. Available from: https://doi.org/10.1056/nejmra1510062

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, 2883–2896. Available from: https://doi.org/10.1002/SIM.8159

Zheng, H., Jaki, T. & Wason, J. M. S. (2020) Bayesian sample size determination using commensurate priors to leverage pre-experimental data. Available at: http://arxiv.org/abs/2011.01106 [Accessed 8th October 2021].

Zheng, H. & Wason, J.M.S. (2022) Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. *Biostatistics*, 23, 120–135. Available from: https://doi.org/10.1093/biostatistics/kxaa019

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

---

**How to cite this article:** Ouma, L.O., Grayling, M.J., Wason, J.M.S. & Zheng, H. (2022) Bayesian modelling strategies for borrowing of information in randomised basket trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(5), 2014–2037. Available from: https://doi.org/10.1111/rssc.12602