An information-theoretic approach for selecting arms in clinical trials

Pavel Mozgunov[1] and Thomas Jaki[1,2]

[1] *Department of Mathematics and Statistics, Lancaster University, Lancaster, UK.*

[2] *MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.*

Summary. The question of selecting the “best” amongst different choices is a common problem in statistics. In drug development, our motivating setting, the question becomes, for example: which treatment gives the best response rate. Motivated by recent developments in the theory of context-dependent information measures, we propose a flexible response-adaptive experimental design based on a novel criterion governing arm selections which can be used in adaptive experiments with simple (e.g. binary) and complex (e.g. co-primary, ordinal, nested) endpoints. It was found that for specific choices of the context dependent measure, the criterion leads to a reliable selection of the correct arm without any parametric or monotonicity assumptions, and provides noticeable gains in settings with costly observations. The asymptotic properties of the design are studied for different allocation rules, and the small sample size behaviour is evaluated in simulations in the context of Phase II clinical trials with different endpoints. We compare the proposed design to currently used alternatives and discuss its practical implementation.

Keywords: Dose Finding; Experimental Design; Information Gain; Multinomial Outcomes; Response-Adaptive Design, Shannon’s Differential Entropy

1. Introduction

Over the past decades, a variety of different methods for clinical trials aiming to select the “optimal” arm (e.g. dose, combination of treatments, treatment regimen, etc.) have been proposed in the literature (see e.g. O’Quigley et al., 2017, for a recent review of novel methods). Given m arms, the aims of Phase I and Phase II clinical trials are often to select the target arm (TA), the arm whose toxicity probability is closest to the maximal accepted target, $0 < \gamma_t < 1$, or (and) whose efficacy probability is closest to the target efficacy, $0 < \gamma_e \leq 1$, where higher values of γ_e corresponds to more effective arms. Despite the similar problem formulation for Phase I (evaluating toxicity) and Phase II (evaluating efficacy) trials, quite different approaches are generally used.

In Phase I dose-escalation trials, designs assuming a monotonic dose-toxicity relationship have been shown to have good operating characteristics in the context of single agent trials (Iasonos et al., 2016; Clertant and O’Quigley, 2017). There is, however, considerable uncertainty in the toxicity ordering for clinical trials investigating combinations of agents or when considering different treatment schedules (Wages et al., 2011). Methods based on a monotonicity assumption are of limited use for such trials. To overcome this issue and to relax the monotonicity assumption, some specialised approaches have been proposed, see e.g. Riviere et al. (2015) for a review of recent methods for combination trials, and Wages et al. (2014); Guo et al. (2016) for approaches to dose-schedule studies. The majority of novel Phase I methods relaxing the monotonicity assumption rely either on a complex parametric model or on explicit orders of toxicity. While such methods allow borrowing information between treatment arms, they might fail to find the TA in trials with a large number of potential orderings and a limited sample size. Furthermore, the majority of such designs consider a single binary endpoint only while more complex outcomes are becoming more frequent in dose finding trials, see Lee et al. (2017) for an example with multiple toxicity grades, and Thall and Cook (2004) for examples of trials with multinomial outcomes assuming a monotonic dose-toxicity relationship. Despite this, methods for studies with non-binary outcomes relaxing monotonicity assumption are sparse to date.

While relaxing the assumption of monotonicity between treatment arms in Phase I studies is relatively novel, designs that consider arms independently have been proposed for a long time in the Phase II setting (see e.g. Stallard and Todd, 2003; Koenig et al., 2008; Magirr et al., 2012). Williamson et al. (2016) have recently advocated designs maximising the expected number of responses in small populations trials. As a result, adaptive randomisation methods and optimal Multi-Arm Bandit (MAB) approaches are starting to be considered as appropriate candidates to fulfill this objective. Although MAB designs outperform other well-established methods in terms of the expected number of successes, they can suffer low statistical power for testing comparative hypotheses (Villar et al., 2015a). This problem corresponds to the "exploration vs exploitation" trade-off (Azriel et al., 2011). Solutions to tackle this balance and to achieve a high power while still assigning the majority of patients to superior treatments is an emerting topic in the MAB field. In particular randomised versions of the optimal MAB designs (Villar et al., 2015b), and approaches fixing the allocation of patients to the control arm (Villar et al., 2015a,b; Williamson et al., 2016; Villar et al., 2018) were proposed. These methods have primarily been developed for binary endpoints and selecting the TA corresponding to the highest response probability, and as a result, cannot be applied to a problem of selecting arm with the arbitrary target probability, γ_e, such as studies looking to select the ED80, the dose giving 80% of the maximum efficacy. At the same time, MAB approaches for non-binary endpoints, e.g. for multinomial (Glazebrook, 1978), normal (Jones, 1970, 1975), and exponential (Gittins et al., 2011) endpoints have been known for a long time, but only recently started to be explored in more details for application in clinical trials (Smith and Villar, 2018; Williamson and Villar, 2019).

While current guidelines generally recommend single endpoints for primary analyses of confirmatory clinical trials, it is recognized that certain settings require inference on multiple endpoints for comprehensive conclusions on treatment effects (Ristl et al., 2018). Consequently, Phase II clinical trials evaluating several endpoints, for example, toxicity and efficacy endpoints, co-primary efficacy endpoints or nested efficacy endpoints, start to attract attention in the literature (Song, 2015; Zhou et al., 2017). While formal testing for a difference in treatment responses remains the main focus of designs proposed for such trials, maximising the number of patients receiving the superior treatment is also of crucial importance – specifically in small population trials. Despite that, response-adaptive designs for settings with multiple endpoints have not been extensively studied yet.

This work is motivated by several Phase I and Phase II clinical trials which could benefit from an experimental design that does not require a parametric or monotonicity assumption between arms, and for which the authors contributed as statistical collaborators. One of them is the TAILoR (Pushpakom et al., 2019) trial which considered three active arms and placebo with the primary objective to find whether the response of at least one active arm is significantly different from the placebo group. The second objective was to find the optimal arm defined as the arm with the largest difference compared to placebo. The original study employed a two stage design in which half of the patients were equally randomised to four arms initially before a selection of all promising arms was undertaken. This design is expected to lead to a reliable answer to the first question, but will result in a low number of patients on the optimal arm. Therefore, response-adaptive designs such as MAB are of interest. The MAB approaches, however, can result in a failure to answer the primary goal of the trial. Therefore, a design able to balance these objectives is of interest.

The research problem described above can be considered as the general issue of correct selection of the TA whose response probability is closest to the percentile, $0 < \gamma \leq 1$. Importantly, an investigator aims to assign the majority of patients to the TA, but has limited information about the dependencies between arms.

In this work, we propose a general response-adaptive experimental design for studies with multinomial outcomes to solve a generic problem of selecting the TA under ethical constraints (e.g. maximize the number of patients at the superior arm) and when each observations is costly. Based on the theory of weighted (or, context-dependent) information measures (Belis and Guiasu, 1968; Kelbert et al., 2016), we propose to use the information gain (found as a difference of the Shannon differential entropy and the weighted Shannon differential entropy) as a criterion for the decision-making in clinical trials. The proposed approach allows incorporation of the context of the outcomes (e.g. avoid high toxicity or low efficacy) in the information measures themselves. This is achieved by assigning a greater "weight" to the information obtained about arms with desirable characteristics. Through specifying an arbitrary parametric weight function, the proposed approach can be applied to various experiments with (ethical) constraints tailored for the specific investigator's needs. In this work, two families of weight functions with a particular interest in arms whose response probabilities are in the neighbourhood of $\boldsymbol{\gamma}$ are considered in more details. We show that, subject to appropriate tuning, the design employing the derived criteria allocates each patient to the treatment estimated to be the best while taking into account the uncertainty about the estimates for each arm and can lead to better operating characteristics than alternative approaches. This leads to fulfilling of statistical goals of the experiment under the ethical constraints.

The idea of applying information-theoretic concepts, and specifically the Shannon entropy (Shannon, 1948), to govern treatment selection dates back to the work by Klotz (1978) who introduced the Maximum Entropy Constrained Balance Randomization design which seeks to maximize the Shannon entropy subject to the expected imbalance. This and related ideas of using the Shannon entropy, however, have received little attention in the literature until very recently when other designs for clinical trials using the information gain principle have been proposed (see e.g. Barrett, 2016; Kim and Gillen, 2016). These works, however, employ the standard definitions of information measures and include the ethical considerations through additional constraints on the derived information-theoretic criteria. The need for these constraints arises as standard measures of information do not depend on the value of the outcomes themselves but only the corresponding probabilities of these outcomes (Kelbert and Mozgunov, 2017). Therefore, they are called "context-free" (Kelbert et al., 2016). While the context-free nature gives the notion of information great flexibility which explains its successful application in various fields, it might be also considered as a drawback in many application areas such as clinical trials. It was found that the "context" of the experiment can be included into the information measures directly using a weight function (Belis and Guiasu, 1968; Kelbert and Mozgunov, 2015) that gives more value to the points of specific interest. Based on this, a Phase I/II dose-finding clinical trial design with trinary outcomes that utilizes an information gain criterion has been developed by Mozgunov and Jaki (2019). Furthermore, similar arguments and weight functions were used to derive a loss function for Phase I dose-escalation trials with binary responses (Mozgunov and Jaki, 2020).

The current work builds on these recent developments and expands the ideas in the following ways. Firstly, we consider a generic setting with multinomial outcomes and study a family of weight functions parametrised by the newly introduced penalisation parameter κ that generalises the criteria used by Mozgunov and Jaki (2019, 2020) and extends the potential applications beyond dose-finding clinical trials. Secondly, we propose an asymptotically unbiased and consistent estimator of the derived criterion and study the theoretical properties of the design based on this criterion. Finally, we propose a unified framework for using the weighted information gain to govern the treatment selection that can be used with an arbitrary parametric weight function specific to the ethical considerations of a given experiment.

The remainder of the paper is organized as follows: derivations of the criterion and assignment rules are given in Section 2. The procedure for finding a robust optimal value of the penalization parameter κ of the proposed design is given in Section 3. The proposed design is applied to the motivating setting of Phase II in Section 4 and to a trial with co-primary efficacy endpoints in Section 5, respectively. We conclude with a discussion in Section 6.

2. Methods

2.1. Selection Criteria

Consider a discrete random variable, corresponding to treatment arm j, taking one of d values and a corresponding random probability vector $\mathbf{Z}_j = \left[Z_j^{(1)}, Z_j^{(2)}, \ldots, Z_j^{(d)}\right] \in \mathbb{S}^d$ defined on a unit simplex

$$\mathbb{S}^d = \{\mathbf{Z}_j : Z_j^{(1)} > 0, Z_j^{(2)} > 0, \ \ldots, Z_j^{(d)} > 0; \ \sum_{i=1}^{d} Z_j^{(i)} = 1\}. \tag{1}$$

Assume that $\mathbf{Z}_j$ has a prior Dirichlet distribution $\text{Dir}(\mathbf{v}_j + \mathbf{J})$ where $\mathbf{v}_j = \left[v_j^{(1)}, \ldots, v_j^{(d)}\right]^{\mathrm{T}} \in \mathbb{R}_+^d$, $\sum_{i=1}^{d} v_j^{(i)} = \beta$ and $\mathbf{J}$ is a d-dimensional unit vector. After n_j realizations of a discrete random variable in which $x_j^{(i)}$ outcomes of i are observed, $i = 1, \ldots, d$, the random vector $\mathbf{Z}_{n_j}$ has a Dirichlet posterior distribution with density function

$$f_{n_j}(\mathbf{p}_j|\mathbf{x}_j) = \frac{1}{B(\mathbf{x}_j + \mathbf{v}_j + \mathbf{J})} \prod_{i=1}^{d} \left(p_j^{(i)}\right)^{x_j^{(i)}+v_j^{(i)}}, \; B(\mathbf{x}_j+\mathbf{v}_j+\mathbf{J}) = \frac{\prod_{i=1}^{d} \Gamma(x_j^{(i)} + v_j^{(i)} + 1)}{\Gamma\left(\sum_{i=1}^{d}(x_j^{(i)} + v_j^{(i)} + 1)\right)} \tag{2}$$

where $\mathbf{p}_j = \left[p_j^{(1)}, \dots, p_j^{(d)}\right]^{\mathrm{T}}$, $\mathbf{x}_j = \left[x_j^{(1)}, \dots, x_j^{(d)}\right]$, $\sum_{i=1}^{d} x_j^{(i)} = n$, $0 < p_j^{(i)} < 1$, $\sum_{i=1}^{d} p_j^{(i)} = 1$ and $B(\mathbf{x}_j + \mathbf{v}_j + \mathbf{J})$ is the Beta-function and $\Gamma(x)$ is the Gamma-function.

Let $\boldsymbol{\alpha_j} = \left[\alpha_j^{(1)}, \dots, \alpha_j^{(d)}\right]^{\mathrm{T}} \in \mathbb{S}^d$ be the vector in the neighbourhood of which f_{n_j} concentrates as $n_j \to \infty$. For example, in the clinical trials with binary responses considered in Section 4, the outcomes are response/no response, and α is the probability of response at a given arm. A classic question of interest in this setting is to estimate the probability vector, $\boldsymbol{\alpha_j}$. The information required to answer the estimation question can be measured by the Shannon differential entropy of f_{n_j} (Cover and Thomas, 2012)

$$h(f_{n_j}) = -\int_{\mathbb{S}^d} f_{n_j}(\mathbf{p}_j|\mathbf{x}_j)\log f_n(\mathbf{p}_j|\mathbf{x}_j)\mathrm{d}\mathbf{p}_j \tag{3}$$

with convention $0\log 0 = 0$. The classic formulation of the estimation question, however, does not take into account the fact that an investigator would like to find the target arm (TA) having pre-specified characteristics $\boldsymbol{\gamma} = \left[\gamma^{(1)}, \dots, \gamma^{(d)}\right] \in \mathbb{S}^d$. These target pre-specified characteristics could be, e.g. $\gamma = 1$ in a Phase II trial targeting the most efficient arm, or $\gamma = 0.25$ in a Phase I trial targeting the maximum tolerated dose corresponding to a 25% toxicity risk.

To take the context of the experiment and the nature of the outcomes $\mathbf{p}_j$ into account, one can consider an estimation experiment with "sensitive" area (i.e. the neighbourhood of $\boldsymbol{\gamma}$). The information required in such an experiment can be measured by the *weighted* Shannon differential entropy (Belis and Guiasu, 1968; Clim, 2008; Kelbert et al., 2016; Kelbert and Mozgunov, 2017) of f_{n_j} with a positive weight function $\phi_{n_j}(\mathbf{p}_j)$

$$h^{\phi_{n_j}}(f_{n_j}) = -\int_{\mathbb{S}^d} \phi_{n_j}(\mathbf{p}_j) f_{n_j}(\mathbf{p}_j|\mathbf{x}_j)\log f_{n_j}(\mathbf{p}_j|\mathbf{x}_j)\mathrm{d}\mathbf{p}_j. \tag{4}$$

The crucial difference between the information measures given in Equation (3) and Equation (4) is the weight function, $\phi_{n_j}(\mathbf{p}_j)$, which emphasizes the interest in the neighbourhood of $\boldsymbol{\gamma}$ rather than on the whole $\mathbb{S}^d$. It reflects that the information about the probability vector which lies in the neighbourhood of $\boldsymbol{\gamma}$ is more valuable in the experiment. Note that the context-free measure (3) can be interpreted as a weighted measure (4) with equal weights on every point in $\mathbb{S}^d$.

In actual studies, an investigator is typically interested in answering the question: Which arm has an associated probability vector closest to $\boldsymbol{\gamma}$. For this question, the information gain from considering the experiment with a sensitive area equals to

$$\Delta_{n_j} = h(f_{n_j}) - h^{\phi_{n_j}}(f_{n_j}). \tag{5}$$

Here, Δ_{n_j} is the average amount of additional statistical information required when considering the context-dependent estimation problem instead of the traditional one.

Following the information gain approach, the first term in the equation above is the information in a classic experiment using the context-free measure, while the second term is the information when the context is taken into account.

Our central proposal is to use this measure Δ_{n_j} to govern the arm selection in a sequential experiment. The weight function used to compute the information gain can be of different forms to reflect the question an investigator is interested in and define the "value" of the information in different areas of the simplex $\mathbb{S}^d$. The weight function should, therefore, be set in line with the objectives of the clinical trial. Drawing a parallel with the MAB approaches, Equation (5) can be also interpreted as defining an index for allocation of sampling observation to each arm. In this work, we will consider two families of weight functions that are suitable for two different clinical settings. First, we focus on a family of weight functions for the sensitive estimation question as above, and introduce a second family accounting for minimum and (or) maximum thresholds in Section 2.5.

We begin by considering a family of weight functions in the Dirichlet form

$$\phi_{n_j}(\mathbf{p}_j) = C(\mathbf{x}_j, \boldsymbol{\gamma}, n_j) \prod_{i=1}^{d} \left(p_j^{(i)}\right)^{\gamma^{(i)} n_j^{\kappa}} \tag{6}$$

parametrised by $\kappa \in (0,1)$ where $C(\mathbf{x}_j, \boldsymbol{\gamma}, n_j)$ is a constant satisfing the normalization condition $\int_{\mathbb{S}^d} \phi_{n_j}(\mathbf{p}_j) f_{n_j}(\mathbf{p}_j|\mathbf{x}_j)\mathrm{d}\mathbf{p}_j = 1$. The parameter κ is restricted to the unit interval to ensure asymptotically unbiased estimates of $\boldsymbol{\alpha_j}$: $\lim_{n_j\to\infty} \int_{\mathbb{S}^d} \mathbf{p}_j \phi_{n_j}(\mathbf{p}_j) f_{n_j}(\mathbf{p}_j)\mathrm{d}\mathbf{p}_j = \boldsymbol{\alpha_j}$. A weight function in the form of the probability density function, f_{n_j}, allows for an analytical expression of the information gain, Δ_{n_j}, and for tracing the dependence of the resulting information gain criteria on the weight function explicitly. Theorem 1 provides an analytical solutions for the asymptotic behaviour of the information gain under the weight function $\phi_{n_j}(\mathbf{p}_j)$ and provides insights on the information gain's relevance to the formulated problem of the TA selection under ethical constraints.

THEOREM 1. *Let $h(f_{n_j})$ and $h^{\phi_{n_j}}(f_{n_j})$ be the standard and weighted differential entropies of (2) with weight function (6) corresponding to arm j. Let $\lim_{n_j\to\infty} \frac{x_j^{(i)}(n_j)}{n_j} = \alpha_j^{(i)}$, $i = 1, 2, \dots, d$ and $\sum_{i=1}^{d} x_j^{(i)} = n_j$, then*

$$\Delta_{n_j} = O\left(\frac{1}{n_j^{1-2\kappa}}\right) \text{ as } n_j \to \infty \text{ if } \kappa < \frac{1}{2};$$

$$\Delta_{n_j} = -\frac{1}{2}\left(\sum_{i=1}^{d} \frac{\left(\gamma^{(i)}\right)^2}{\alpha_j^{(i)}} - 1\right) n_j^{2\kappa-1} + \omega(\boldsymbol{\alpha_j}, \boldsymbol{\gamma}, \kappa, n_j) + O\left(\frac{1}{n_j^{\eta(1-\kappa)-\kappa}}\right) \text{ as } n_j \to \infty \text{ if } \kappa \geq \frac{1}{2}$$

where

$$\omega(\boldsymbol{\alpha_j}, \boldsymbol{\gamma}, \kappa, n_j) = \sum_{u=3}^{\eta} \frac{(-1)^{u-1}}{u} n_j^{u\kappa-u+1} \left(\sum_{i=1}^{d} \frac{\left(\gamma^{(i)}\right)^u}{\left(\alpha_j^{(i)}\right)^{u-1}} - 1\right) \text{ and } \eta = \lfloor (1-\kappa)^{-1} \rfloor$$

All proofs are provided in the Supplementary Materials.

The information gain, Δ_{n_j}, tends to 0 for $\kappa < 1/2$ which implies that assigning a value of information with a rate less than $1/2$ is insufficient to emphasize the importance of the context of the study. However, the limit is non-zero for $\kappa \geq 1/2$. Following the conventional information gain approach, one would like to make a decision that maximises the statistical information in the experiment. The leading terms of the information gain Δ_{n_j} is always non-positive, and for any fixed n the asymptotics terms achieve the maximum value 0 at the point $\alpha_j^{(i)} = \gamma^{(i)}$, $i = 1, \ldots, d$ (all constants are cancelled out). This reflects the fact that, by adding one more research question into the information measure through the weight function, the uncertainty in the experiment (in terms of the differential information measure) is increased. There is no additional uncertainty when the answer to the both research questions coincide. Therefore, it follows that collecting more information about the arm which has characteristics $\boldsymbol{\alpha_j}$ close to the target $\boldsymbol{\gamma}$ (the ethical constraint of the experiment) implies maximisation of the information gain, Δ_{n_j}. Consequently, each patient tends to be assigned to the TA, and the criterion Δ_{n_j} is a patient's gain criterion (Whitehead and Williamson, 1998). It will be further demonstrated that, for certain values of the parameter κ, Δ_{n_j} also takes into account the statistical uncertainty of the arm and achieves the goal of the trial under ethical constraints. Therefore, we propose to use the information gain Δ_{n_j} for the arm selection in a sequential experiment.

To keep the tractable solution which can be easily interpreted in applications and could be argued to be easier to justify to use, we construct the arm selection criterion under the weight function $\phi_{n_j}(\cdot)$ nearly maximising the information gain using the leading term of the asymptotic expression for Δ_{n_j} derived in Theorem 1:

$$\delta^{(\kappa)}(\boldsymbol{\alpha_j}, \boldsymbol{\gamma}) := \frac{1}{2}\left(\sum_{i=1}^{d} \frac{\left(\gamma^{(i)}\right)^2}{\alpha_j^{(i)}} - 1\right) n_j^{2\kappa-1}. \tag{7}$$

Note that maximising the leading term of the information gain asymptotics is equivalent to minimising $\delta^{(\kappa)}(\boldsymbol{\alpha_j}, \boldsymbol{\gamma})$. The criterion (7) possesses some desirable properties: $\delta^{(\kappa)}(\cdot) \geq 0$ and $\delta^{(\kappa)}(\cdot) = 0$ iff $\boldsymbol{\alpha_j} = \boldsymbol{\gamma}$ for all κ and n_j. The boundary values $\alpha_j^{(i)} = 0$, $i = 1, \ldots, d$ correspond to infinite values of $\delta^{(\kappa)}(\boldsymbol{\alpha_j}, \boldsymbol{\gamma})$ which is advocated by Aitchison (1992) as one of the important properties for functions defined on a simplex, $\mathbb{S}^d$. Interestingly, as pointed out by one of the referees, the selection criterion (7) is similar to a testing approach based on a Wald-test but obtained using an independent argument.

While the term in brackets reflects how close the vector of the parameters $\boldsymbol{\alpha_j}$ is to the vector of the target characteristics, $\boldsymbol{\gamma}$, the balance in the "exploration vs exploitation" trade-off is controlled by the term $n_j^{2\kappa-1}$ reflecting the penalty on the number of observations on the same arm. A larger number of patients on an arm makes it less desirable to be chosen. Therefore, as the experiment progresses the design requires an increasing level of confidence that the selected arm is the TA. Increasing values of κ correspond to a greater penalty of the number of patients allocated to a specific arm and hence is expected to lead to a more spread allocation. This corresponds to a greater interest in the statistical power of the experiment. In contrast, $\kappa = 1/2$ corresponds to no penalty and is of particular interest in trials with small sample sizes. We will refer to κ as the *penalization parameter*. The penalization term on the number of observation in a given arm is of a growing interest in reinforcement learning, where it is considered as a way to address the exploration-exploitation trade-off similar to the considered problem (see e.g. an overview of the related literature by Browne et al., 2012).

In the context of studies with binary outcomes considered in the example in Section 4, the selection criterion takes the form

$$\delta^{(\kappa)}(\alpha_j,\gamma) := \frac{2(\alpha_j-\gamma)^2}{\alpha_j(1-\alpha_j)} n_j^{2\kappa-1} \tag{8}$$

where α_j is the probability of an event for arm j and γ is the target probability. The squared distance term ensures that the arm with α_j closest to γ is selected. At the same time, the denominator is the variance of a probability of a binary event and, therefore, takes into account the uncertainty about the arm. Thus, the first term can be considered as a normalised distance between the target probability and the probability corresponding to a particular arm. Mozgunov et al. (2019) have shown that the first term of the criterion (8) shares the properties of the squared distance between $\alpha_j, \gamma \in (0,1)$ on the logit-transform scale proposed by Aitchison (1982, 1992) but, additionally, is convex and resembles a well-known squared distance formula.

2.2. Estimation

While the desirable characteristics of the TA, $\boldsymbol{\gamma}$, are known and fixed prior to the trial, the selection criterion (7) also depends on the true unknown parameters, $\boldsymbol{\alpha_j}$. Below, we propose an estimator of the selection criterion (7).

Consider a discrete set of m arms, $A_1,\ldots,A_m$, associated with $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ and $n_1,\ldots,n_m$ observations. Arm $A_{j^\star}$ is optimal if $\delta^{(\kappa)}(\boldsymbol{\alpha}_{j^\star},\boldsymbol{\gamma}) = \inf_{j=1,\ldots,m} \delta^{(\kappa)}(\boldsymbol{\alpha}_j,\boldsymbol{\gamma})$. To estimate $\delta^{(\kappa)}(\boldsymbol{\alpha}_j,\boldsymbol{\gamma})$, consider a random variable $\tilde{\delta}_{n_j}^{(\kappa)} \equiv \delta^{(\kappa)}(\mathbf{Z}_{n_j},\gamma)$ with $\mathbf{Z}_{n_j}$ having Dirichlet distribution (2). Theorem 2 shows that $\tilde{\delta}_{n_j}^{(\kappa)}$ is asymptotically unbiased, consistent and asymptotically normal.

Theorem 2. *Let $\bar{Z}$ be a standard Gaussian RV and $\widetilde{\mathbf{Z}}_{n_j} = \Sigma^{-1/2}\left(\boldsymbol{Z}_{n_j} - \boldsymbol{\alpha_j}\right)$ be a random variable with pdf $\widetilde{f}_{n_j}$ where pdf of $\mathbf{Z}_{n_j}$ is given in (2) with $\lim_{n_j\to\infty} \frac{x_j^{(i)}(n_j)}{n_j} = \alpha_j^{(i)}$ for $i = 1,2,\ldots,d$, $\sum_{i=1}^{d} x_j^{(i)} = n_j$ and Σ_j is a d-dimensional square matrix with elements $\Sigma_{j[uv]} = \frac{\alpha_j^{(u)}(1-\alpha_j^{(u)})}{n_j}$ if $u = v$ and $\Sigma_{j[uv]} = -\frac{\alpha_j^{(u)}\alpha_j^{(v)}}{n_j}$ if $u \neq v$. Let $\tilde{\delta}_{n_j}^{(\kappa)} = \delta^{(\kappa)}(\boldsymbol{Z}_{n_j},\boldsymbol{\gamma})$, $\nabla\delta^{(\kappa)}(\boldsymbol{z},\boldsymbol{\gamma}) = \left[\frac{\partial\delta^{(\kappa)}(\boldsymbol{z},\boldsymbol{\gamma})}{\partial z^{(1)}},\ldots,\frac{\partial\delta^{(\kappa)}(\boldsymbol{z},\boldsymbol{\gamma})}{\partial z^{(d)}}\right]^{\mathrm{T}}$, $\bar{\delta}_{n_j}^{(\kappa)} = \bar{\Sigma}_j^{-1/2}\left(\delta^{(\kappa)}(\boldsymbol{Z}_{n_j},\boldsymbol{\gamma}) - \delta^{(\kappa)}(\boldsymbol{\alpha_j},\boldsymbol{\gamma})\right)$ where $\bar{\Sigma}_j = \nabla_{\boldsymbol{\alpha_j}}^{\mathrm{T}}\Sigma_j\nabla_{\boldsymbol{\alpha_j}}$ and $\nabla_{\boldsymbol{\alpha_j}} \equiv \nabla\delta^{(\kappa)}(\boldsymbol{z},\boldsymbol{\gamma})$ evaluated at $\boldsymbol{z} = \boldsymbol{\alpha_j}$. Then, $\lim_{n_j\to\infty}\mathbb{E}\tilde{\delta}_{n_j}^{(\kappa)} = \delta^{(\kappa)}(\boldsymbol{\alpha_j},\boldsymbol{\gamma})$, $\lim_{n_j\to\infty}\mathbb{V}\tilde{\delta}_{n_j}^{(\kappa)} = 0$, and $\bar{\delta}_{n_j}^{(\kappa)}$ weakly convergences to $\bar{Z}$.*

A single summary statistics for $\delta^{(\kappa)}(\mathbf{Z}_{n_j},\boldsymbol{\gamma})$ is needed to select the most promising arm in the sequential experiment. We consider a “plug-in” estimator, $\hat{\delta}^{(\kappa)}(\hat{\mathbf{p}}_{n_j},\gamma) \equiv \hat{\delta}_{n_j}^{(\kappa)}$ with $\hat{\mathbf{p}}_{n_j} = [\hat{p}_{n_j}^{(1)},\ldots,\hat{p}_{n_j}^{(i)},\ldots,\hat{p}_{n_j}^{(d)}]$ and $\hat{p}_{n_j}^{(i)} = \frac{x_j^{(i)}+v_j^{(i)}}{n_j+\beta_j^{(i)}}$, $i = 1,\ldots,d$, the mode of the posterior Dirichlet distribution. The estimator for the arm A_j takes the form

$$\hat{\delta}^{(\kappa)}_{n_j} = \delta^{(\kappa)}(\hat{\mathbf{p}}_{n_j}, \boldsymbol{\gamma}) = \frac{1}{2}\left(\sum_{i=1}^{d} \frac{\left(\gamma^{(i)}\right)^2}{\hat{p}^{(i)}_{n_j}} - 1\right) n_j^{2\kappa-1}, \ j = 1, 2, \ldots, m. \tag{9}$$

The estimator (9) requires prior parameters $\mathbf{v}_1, \ldots, \mathbf{v}_m$ to start the experiment.

2.3. Assignment Rules

The estimator (9) is used to govern the selection among arms during the experiment and summarizes the arm's characteristics. It can be applied to different types of sequential experiments. We consider two assignment rules: A deterministic "select the best" rule, and a randomisation rule that randomizes patients to arms. These rules follow the setting of the motivating clinical trials. For example, the deterministic rule prioritizes the exploitation over exploration and can be used in Phase I trials evaluating toxicity where the randomization to all doses might not ethical (an example is provided in the Supplementary Materials) or in the Phase II setting if the goal of maximizing the number of successes is prioritised (considered in Section 4). The randomised rule could be favoured when an investigator is primarily interested in a high statistical power (Section 4).

2.3.1. Deterministic "Select the Best" rule

Let n be a total sample size and begin with the experiment with the arm that minimizes $\hat{\delta}^{(\kappa)}_{\beta_j}$, $j = 1, \ldots, m$. Given n_j observations, $\mathbf{x}_j$ outcomes for the arm A_j, $j = 1, \ldots, m$ and using the "plug-in" estimator, an arm $A_{j^\star}$ is selected if $\hat{\delta}^{(\kappa)}_{n_{j^\star}} = \inf_{j=1,\ldots,m} \hat{\delta}^{(\kappa)}_{n_j}$. The method proceeds until the total number of n is attained. The arm $A_{j^\star}$ satisfying

$$\hat{\delta}^{(1/2)}_{n_{j^\star}} = \inf_{j=1,\ldots,m} \hat{\delta}^{(1/2)}_{n_j}. \tag{10}$$

is adopted for the final recommendation, where $\sum_j n_j = n$. The value $\kappa = 0.5$ in (10) is used as there is no more scope for exploration at the end of the study, and the final recommendation should be not penalized by the sample size.

2.3.2. Randomised rule

Under this rule, the arm selected in the experiment is randomised with probabilities $\tilde{w}_j \equiv \frac{1/\tilde{\delta}^{(\kappa)}_{n_j}}{\sum_{i=1}^m 1/\tilde{\delta}^{(\kappa)}_{n_i}}$, $j = 1, \ldots, m$. When no observations have yet been collected, the procedure randomizes according to the criterion based on the prior distribution alone, $\hat{\delta}^{(\kappa)}_{\beta_j}$, $j = 1, \ldots, m$. Then, given n_j observations, $\mathbf{x}_j$ outcomes for arm A_j, $j = 1, \ldots, m$ and using the "plug-in" estimator (9), arm A_j is selected with probability $\hat{w}_j = 1$ if $\hat{\delta}^{(\kappa)}_{n_j} = 0$ and with probability

$$\hat{w}_j = \frac{1/\hat{\delta}^{(\kappa)}_{n_j}}{\sum_{i=1}^m 1/\hat{\delta}^{(\kappa)}_{n_i}} \text{ if } \hat{\delta}^{(\kappa)}_{n_i} > 0, \ i = 1, \ldots, m. \tag{11}$$

The method proceeds until n observations are attained. We adopt $A_{j^\star}$ as in Equation (10) for the final recommendation.

2.4. Design's Consistency

Although large sample size is never achieved in early phase clinical trials, the consistency condition of the design ensures that the approach provides a more reliable selection of the TA as sample size increases. The consistency condition for the proposed design under two assignment rules under the weight function $\phi_n(\cdot)$ is given in Theorem 3.

Theorem 3. *Let us consider the experimental design with a selection criteria based on $\tilde{\delta}_{n_j}^{(\kappa)}$, m arms and true probabilities vectors $\boldsymbol{\alpha}_j$, $j = 1, \ldots, m$. Then,* ***(a)*** *The design is consistent under the randomised rule for $\kappa \geq 0.5$;* ***(b)*** *the design is consistent under the deterministic rule for $\kappa > 0.5$.*

Note that under the deterministic rule, $\kappa = 0.5$ leads to a lack of consistency of the design. The effect on this will be considered in the setting of Phase I clinical trial with a small sample size (see Supplementary Materials) and in the setting of Phase II clinical trial with moderate sample sizes.

2.5. An Alternative Weight Function

The weight function, $\phi_{n_j}(\cdot)$, above can be a suitable choice when an investigator is interested in the TA with particular characteristics. At the same time, alternative research questions (e.g. composite) can be of interest to an investigator in the trial. For example, in some clinical trials, lower and/or upper bounds on the characteristics of interest can be imposed. The proposed information-theoretic approach can also be applied to such more complex questions. We provide an example below.

Consider a trial in which, one is still interested in the TA as close as possible to $\boldsymbol{\gamma}$ but only if these characteristics are "close enough" to the target. For example, in the setting of a Phase II clinical trial with binary responses (Section 4), the goal can be formulated as "to select the TA with the highest response probability that is above the minimum efficacy bound ψ". One of the possible weight functions reflecting these trial objectives can be formulated as follows.

As before, let $\mathbf{Z}$ be a random probability vector, and denote the vectors of lower and upper bounds of the probabilities of interest by $\boldsymbol{\psi_L} = \left(\psi_L^{(1)}, \ldots, \psi_L^{(d)}\right)$ and $\boldsymbol{\psi_U} = \left(\psi_U^{(1)}, \ldots, \psi_U^{(d)}\right)$, $\psi_L^{(i)}, \psi_U^{(i)} \in (0,1), i = 1, \ldots, d$, respectively. Further let $\mathbb{S}_\psi^d$ be such that $\mathbb{S}_\psi^d = \{\mathbf{Z} : \ Z^{(1)} \in (\psi_L^{(1)}, \psi_U^{(1)}), Z^{(2)} \in (\psi_L^{(2)}, \psi_U^{(2)}), \ \ldots, Z^{(d)} \in (\psi_L^{(d)}, \psi_U^{(d)})\}$. Then, extending the weight function $\phi_{n_j}(\mathbf{p}_j)$, the weight function for the considered experiment can be written as

$$\phi_{n_j}^\star(\mathbf{p}_j) = \mathbb{I}\left(\mathbf{p}_j \in \mathbb{S}_\psi^d\right) \times \prod_{i=1}^{d} \left(p_j^{(i)}\right)^{\gamma^{(i)} n_j^\kappa} . \tag{12}$$

After scaling the weighted differential entropy, the following information gain criterion can be used to govern the arm selection in the experiment

$$\Delta^{\star}_{n_j} = h(f_{n_j}) - \frac{h^{\phi^{\star}_{n_j}}(f_{n_j})}{\int_{\mathbb{S}^d} \phi^{\star}_{n_j}(\mathbf{p}_j) f_{n_j}(\mathbf{p}_j) \mathrm{d}\mathbf{p}_j}. \tag{13}$$

The proposed information gain does not have an analytical solution. However, both the probability density function, f_{n_j} and the weight function have parametric forms and hence the criterion can be found using numerical integration. Consequently, instead of maximising the asymptotic expression of the information gain, one can maximise the exact information gain. The rest of the procedure stands: each cohort of patients is assigned to the arm using the obtained values of information gain $\Delta^{\star}_{n_j}$, $j = 1, \ldots, m$ that are updated once the outcomes are observed. Note that, the plug-in estimator in Equation (9) is not used as the integrals above are functions of the number of responses $x_j^{(i)}$ of outcomes i for arm j and the number of observations n_j.

To put the weight function with the boundary values into the specific context, we will consider a Phase II clinical trial with binary responses similar to the example considered in Section 4. Assume that the goal of the trial is to find the treatment arm with the highest response probability that is above the minimum efficacy bound $\psi_L = \psi = 0.70$. Then, the weight function takes the form $\phi^{\star}_{n_j}(p_j) = \mathbb{I}\left(p_j > \psi\right) p_j^{\gamma n_j^{2\kappa-1}} (1-p_j)^{(1-\gamma)n_j^{2\kappa-1}}$.

The inclusion of the boundary value into the weight function for $n_j = 100$, $\kappa = 0.5$, the minimum efficacy value $\psi = 0.70$, and different number of responses is demonstrated in Figure 1 with the target efficacy $\gamma = 1$.

The information gains for the different weight functions are nearly the same for the number of outcomes $x_j \geq 70$ corresponding to an estimated probability of efficacy of $\hat{p}_j \geq 0.70$, and both information gains are still maximised for the highest efficacy probability. However, when the estimated probability falls below the minimum efficacy threshold, the information gain $\Delta^{\star}_{n_j}$ decreases noticeably faster than Δ_{n_j}. As a result, $\Delta^{\star}_{n_j}$ allows for a better discrimination between efficacious $(p_j > \psi)$ and inefficacious arms. Importantly, the information gain $\Delta^{\star}_{n_j}$ can still distinguish treatment arms with an estimated efficacy probability of less than 0.70 because of the underlying uncertainty. Note that, although it was found that Δ_{n_j} tends to a non-positive value, its exact value the moderate sample sizes can be above zero as demonstrated in Figure 1. Nevertheless, a larger value of the criterion still corresponds to a more promising treatment and therefore can be used to discriminate between arms. We will consider how the weight function with the boundary values affects the performance of the design in more detail in Section 4.

3. A robust optimal penalization parameter κ

3.1. Procedure

The penalization parameter, κ controls the exploration-exploitation trade-off. Therefore, the choice of the optimal value of κ (e.g. in the sense of maximising the expected number of successes in a trial, ENS) is crucial. As follows from the proof of Theorem 3, the optimal value depends on the sample size, number of treatment arms and the true probabilities of the response of all treatment arms.

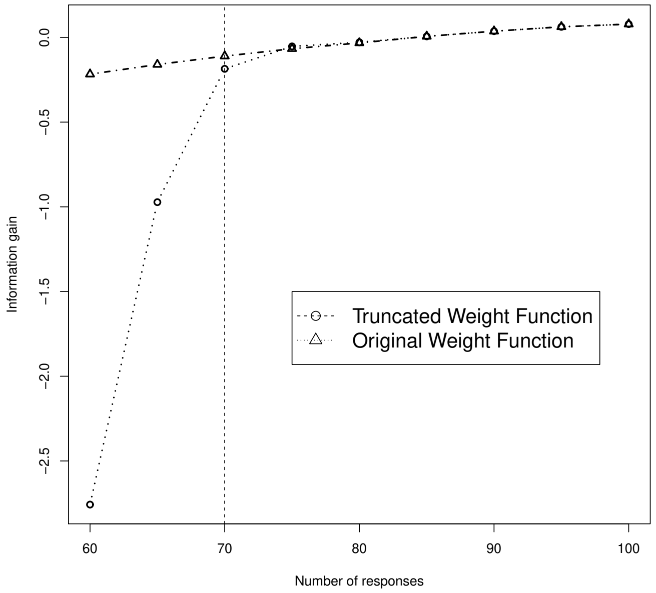

Fig. 1. Exact information gains using the weight function ϕ_{n_j} (dashed–dotted) and the weight function $\phi^{\star}_{n_j}$ (dotted) with the minimum efficacy value $\psi = 0.7$ for different values of the responses $x_j = 60, 65, \ldots, 100$.

As the true probabilities of response are unknown, we propose an approach to finding the robust optimal value of κ that does not require knowing the true probabilities of response and leads to near-optimal characteristics in the absence of prior information on the response probabilities. For a given optimality criterion, the approach builds on the algorithm by Clertant and O'Quigley (2017) and takes the form:

(a) Define a set of Z scenarios, $S_1, \ldots, S_Z$, where a scenario is the set of parameters $\boldsymbol{\alpha}_j$ defining the distribution of outcomes for the treatment arm j.
(b) Define the quantity of interest $q(\kappa)$ and the objective function $g(q(\kappa))$.
(c) Obtain $q(\kappa|S_z)$ for all κ on the prespecified grid and all $z = 1, \ldots, Z$.
(d) Find the optimal value of $\kappa^{opt} = \arg\min_{\kappa} \frac{1}{Z}\sum_{z=1}^{Z} g(q(\kappa|S_z))$.

Such a procedure results in a robust optimal design with the parameter κ^{opt} that optimizes the objective function $g(\cdot)$. In this work, we will consider two objective functions, $g(q(\kappa))$, with quantities of interest $q(\cdot)$ corresponding to different aims in the exploration-exploitation trade-off.

3.2. Objective Functions

To find the robust optimal design parameter, we use the context with binary responses. Specifically, we will consider two objectives function: (i) maximising the ENS, and (ii) achieving the pre-specified level of power under the Least Favourable Configuration (LFC).

Table 1. The objective function and corresponding quantities of interest for the maximising ENS and Achieving the pre-specified level of power criteria.

Approach	Quantity of Interest, $q_i(\kappa\|\cdot)$	Objective Function, $g_i(q_i)$
Maximising ENS	$q_1 = \sum_{j=1}^{m} p_j \times n_j(\kappa)$	$g_1 = \sum_{z=1}^{Z} \left(q_1(\kappa\|S_z) - q_1(\kappa^\star_{S_z}\|S_z)\right)^2$
Achieving Power	$q_2 = \mathbb{P}\,(\text{reject } H_0\|\text{LFC})$	$g_2 = \kappa \times$ $\mathbb{I}\left(\frac{1}{Z}\sum_{z=1}^{Z} \mathbb{I}\left(q_2(\kappa\|S_z) \geq 0.80 \times q_2^{FR}(S_z)\right) > \xi\right)$

Let $n_j(\kappa)$ be the total number of patients assigned to the treatment arm j using the design with parameter κ, p_j is the response probability for arm j, and $q_2^{FR}(S_z)$ be the power attained using fixed equal randomisation (FR). The objective functions, and the corresponding quantities of interests, q_1, q_2 are given in Table 1.

For the ENS criteria, $\kappa^\star_{S_z} = \arg\min_\kappa(Q(\kappa|S_z))$ is the scenario-specific optimal κ, and the objective function $g_1(\cdot)$ minimizes the expected losses in the ENS associated with the use of non scenario-specific optimal parameter. For the power criteria, the objective function is constructed to guarantee that with a probability of at least ξ, the design will achieve 80% of the power attained by the FR. Here, the power attained using FR under scenario S_z, $q_2^{FR}(S_z)$, normalises for different scenarios for the fixed sample size. We apply the procedure in Section 4.2 and evaluate its performance in Section 4.3 and Section 5.

3.3. Computation of the quantities of interest

For the values of the penalization parameter κ on the grid $\kappa = 0.50, 0.51, \ldots, 1$, finding the robust optimal values reduces to computing $q(\kappa)$ for a given value of κ. Similar to multi-arm bandit approaches (Villar et al., 2015a), the challenge here is that an analytical expression for the allocation of patients cannot be found. It can, however, be computed recursively.

To illustrate this recursive procedure, consider a two-arm trial with binary responses and sequential enrollment of patients. Let x_j and n_j be the number of responses and the number of patients assigned to arm j, respectively, α_j be the probability of response, $\hat{\delta}^{(\kappa)}(x_j, n_j)$, $j = 1, 2$ be the estimate of the criterion and $S^n = (x_1, n_1, x_2, n_2)$ be the state of the trial where $n_1 + n_2 = n$. Then, the probability of the state

$$\begin{aligned}
&\mathbb{P}\left(S^n\left(x_1, n_1, x_2, n_2\right)\right) = \\
&\alpha_1 \times \mathbb{I}\left(\hat{\delta}^{(\kappa)}(x_1 - 1, n_1 - 1) < \hat{\delta}^{(\kappa)}(x_2, n_2)\right) \times \mathbb{P}\left(S^{n-1}\left(x_1 - 1, n_1 - 1, x_2, n_2\right)\right) + \\
&(1 - \alpha_1) \times \mathbb{I}\left(\hat{\delta}^{(\kappa)}(x_1, n_1 - 1) < \hat{\delta}^{(\kappa)}(x_2, n_2)\right) \times \mathbb{P}\left(S^{n-1}\left(x_1, n_1 - 1, x_2, n_2\right)\right) + \\
&\alpha_2 \times \mathbb{I}\left(\hat{\delta}^{(\kappa)}(x_1, n_1) > \hat{\delta}^{(\kappa)}(x_2 - 1, n_2 - 1)\right) \times \mathbb{P}\left(S^{n-1}\left(x_1, n_1, x_2 - 1, n_2 - 1\right)\right) + \\
&(1 - \alpha_2) \times \mathbb{I}\left(\hat{\delta}^{(\kappa)}(x_1, n_1) > \hat{\delta}^{(\kappa)}(x_2, n_2 - 1)\right) \times \mathbb{P}\left(S^{n-1}\left(x_1, n_1, x_2, n_2 - 1\right)\right)
\end{aligned}$$

It is, however, known that this recursive procedure gets computationally demanding or even infeasible as the sample size and (or) number of arms increase. Therefore, following Villar et al. (2015b), we will use Monte Carlo simulations to approximate this distribution. It was found that the Monte Carlo simulations provide an accurate approximation of the distribution of allocations and noticeable gains in the computational time. A comparison of the exact computations and the Monte Carlo approximation for various values of n is provided in the Supplementary Materials.

4. Application to a Phase II Clinical Trial

4.1. Setting

Let us consider a Phase II clinical trial whose goals are (i) to find the most effective treatment and (ii) to treat as many patients as possible on the optimal treatment. Similar to the motivating trial, we consider $m = 4$ treatments. We assume that the primary endpoint is a binary measure of efficacy (e.g. response to treatment). While there is a number of competing approaches that could be applied in the considered setting, we limit the comparison to two alternative designs that are known to have good statistical properties in terms of either the number of treated patients or the statistical power. Specifically, we compare to the Gittins index (GI) approach (using the discount factor of 0.99 and non-informative priors, see Gittins and Jones, 1979; Villar et al., 2015a, for more detail), which is the near optimal design in terms of maximising the expected number of successes (ENS) and will serve as a benchmark for this characteristic. Additionally we also compare to fixed and equal randomization (FR) that is known to lead to high statistical power.

We consider two scenarios investigated by Villar et al. (2015a). Scenario 1 investigates $n = 423$ and the true efficacy probabilities are $(0.3, 0.3, 0.3, 0.5)$ while Scenario 2 considers $n = 80$ with true efficacy probabilities $(0.3, 0.4, 0.5, 0.6)$. Following Villar et al. (2015a), we consider the hypothesis $H_0:\ p_0 \geq p_i$ for $i = 1, 2, 3$ with the family-wise error rate calculated at $p_0 = \ldots = p_3 = 0.3$, where p_0 corresponds to the control treatment efficacy probability. The Dunnett test (Dunnett, 1984) is used for hypothesis testing in the FR setting. The hypothesis testing for GI and WE design is performed using an adjusted Fisher's exact test (Agresti, 1992). The adjustment chooses the cutoff values to achieve the same type-I error as the FR. The Bonferroni correction is used for GI and WE designs to correct for multiple testing and the family-wise error rate is set to be less or equal to 5%. Characteristics of interest are (i) the type-I error rate (α), (ii) statistical power ($1 - \eta$), (iii) the expected number of successes (ENS) and (iv) the average proportion of patients on the optimal treatment (p^*).

The proposed design requires a target value, γ. While in practice the target treatment effect can vary in different therapeutic areas, we consider the general setting in which no specific value is specified, and the arm with the highest success probability is of interest. We, therefore, use the highest possible value of a target probability, $\gamma = 0.999$. Investigating the dependence of the operating characteristics of the design on the target value γ in more detail, it was found that this choice might lead to a marginal decrease in the ENS compared to the setting when the true maximum treatment effect is known while fixing the target probability below the true maximum treatment effect can lead to a noticeable decrease in it - see the Supplementary Materials. The vector of the prior mode probabilities $p^{(0)} = [0.99, 0.99, 0.99, 0.99]^{\mathrm{T}}$ is chosen to reflect no prior knowledge about which arm has the highest success probability and the equipoise principle (Djulbegovic et al., 2000). We choose $\beta_0 = 5$ to observations on the control and $\beta_1 = \beta_2 = \beta_3 = 2$ to reflect no prior knowledge for competing arms. The higher value for β_0 compared to $\beta_1, \beta_2, \beta_3$ in the prior probabilities are intended to protect (to a certain extent) for higher number of patients on the control and to achieve a higher power. See Section 6 and the Supplementrary Materials for a more detailed discussion on the influence of prior assumptions on the operating characteristics. We fix $\kappa = 0.5$ for allocation the randomised rule and denote it by WE_{Ran}, and search for the optimal robust values of κ for each sample size under the deterministic "Select-the-Best" rule denoted by WE_{Det} as given below. The software in the form of R code to reproduce the findings of the work is available at `https://github.com/adaptive-designs/inf-theory`.

4.2. Choice of the robust optimal penalization parameter κ

The proposed design requires the specification of the penalization parameter κ. We apply the procedure in Section 3 for the two objective functions, (i) maximising the ENS, and (ii) achieving a particular level of statistical power. We will apply the procedure to the deterministic allocation rule.

Firstly, $Z = 5000$ random scenarios with $m = 4$ treatment arms are generated. For the ENS criterion, we assume an uniform distribution on the probability of responses at each treatment arm, $p_j \sim \mathcal{U}(0, 1)$, $j = 1, 2, 3, 4$. Note that, if there is some prior information on the plausible values of p_j it could be employed at this stage. For the power criterion, we power the trial under the LFC and generate the response probabilities as $p_1 = p_2 = p_3 \sim \mathcal{U}(0, 1)$, and $p_4 = \mathcal{U}(p_1, 1)$. We specify the values of κ on the grid $\kappa = 0.50, 0.51, \ldots, 1$, and conduct the procedure for sample sizes considered in the examples ($n = 80$ and $n = 423$) as well as an intermediate value $n = 165$ (used in the example in Section 5). We use 5000 Monte Carlo simulations to approximate the distribution of patients for each κ under each scenario. For the power objective function, we require that 80% of the power of the FR is achieved with probability at least 90%, $\xi = 0.90$. Note that this requirement is imposed to be satisfied with high probability over the 5000 random scenarios. This means that in any given scenario the achieved power can be both above and below the 80% of the power achieved by the FR design. Note also that the procedure can be computationally expensive. For $n = 423$, for example, the full calibration procedure took around 70 hours (Intel Core i7-8650U CPU @ 1.90GHz × 8) after being parallelized between 5 cores. Note that one can reduce this time by reducing number of Monte Carlo simulations and/or scenarios at the cost of lower precision in the optimal value of κ. The objective function $g_1(\cdot)$ and the quantity of interest q_2 for various values of κ are given in Figure 2.

For the maximum ENS criterion, the optimal values of κ increase as the sample size increases. As expected, when optimising the number of patients on the superior arms, a low value of the penalization parameter should be used if the sample size is small as more spread allocation will result in a decreased ENS. At the same time, for larger sample sizes, a low value of κ can result in allocating many patients to suboptimal arms, and the consequences of this get more severe as the sample size increases. Therefore, the value of $\kappa = 0.51$ and $\kappa = 0.56$ will be used for sample sizes $n = 80$ and $n = 423$, respectively to achieve the near-maximum ENS.

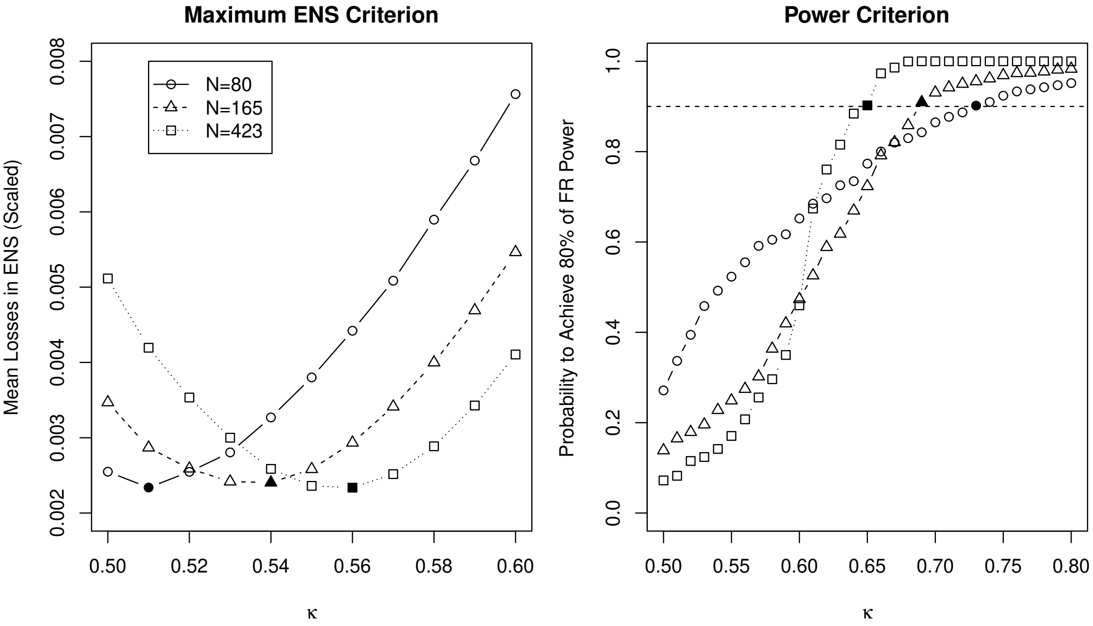

Fig. 2. The values of the objective function g_1 divided by the sample size (Left Panel), and the expected values of q_2 (Right Panel) for various values of the penalization parameter κ, and for various sample sizes $n = 80$ (solid line, circle), $n = 165$ (dashed line, triangle) and $n = 423$ (dotted line, square). Blacked filled shapes correspond to the robust optimal values of κ.

Regarding the power criterion greater values of κ correspond to greater power and, as a result, to a greater probability that the desirable power will be achieved. For various sample sizes, the optimal values are found to be in the interval (0.65,0.73). The minimum values of κ for which the probability to attain the target power is 90% are the robust optimal values and are used in the examples below when balancing the ENS and the statistical power.

4.3. Results

The trade-off between the expected number of successes (ENS) and the statistical power for different values of the penalty parameter κ under the deterministic rule in both scenarios is illustrated in Figure 3.

In both scenarios, greater values of κ correspond to greater power and lower ENS as the increase in penalty tends to more diverse allocations. The exception is $\kappa \in (0.5, 0.55)$ in Scenario 1 where the inconsistency for $\kappa = 0.5$ leads to locking-in on the suboptimal treatment. Subsequently, we use the robust optimal κ found above for the ENS (dashed line) and power (dotted line) criteria.

The operating characteristics of the considered designs in Scenario 1 are given in Table 2. Under the null hypothesis, the performance of all methods is similar and the type-I error is controlled. Under the alternative hypothesis, the $\mathrm{WE}_{\mathrm{Det}}$ design with calibrated optimal $\kappa = 0.56$ performs comparably to the GI in terms of the ENS with

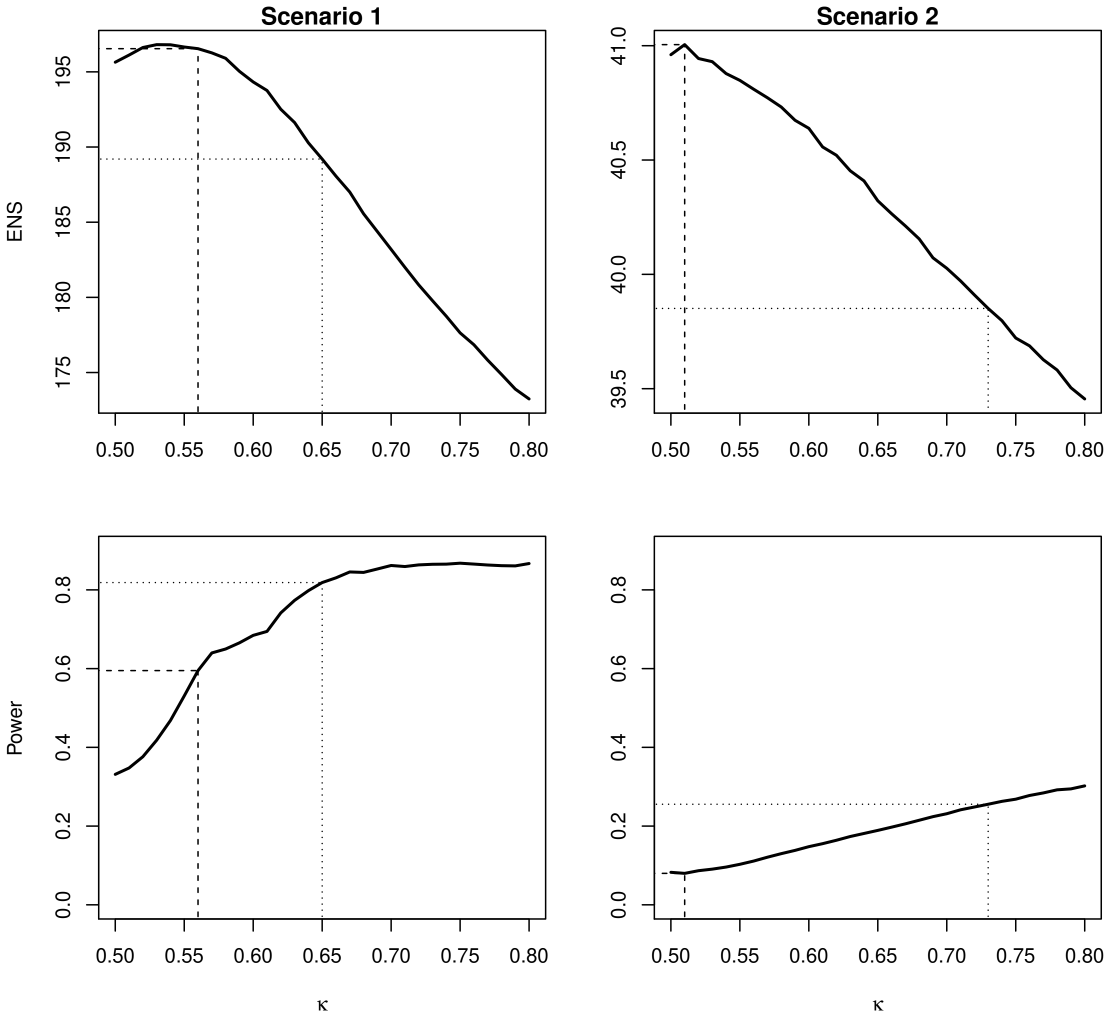

Fig. 3. ENS and power (fixed cutoff value) for the WE design under the deterministic rule for different κ. Dashed and dotted lines correspond to the values of the penalisation parameter κ chosen using the ENS and power criteria.

the GI design resulting in around 2 more respones on average, but increases power by nearly 18% points due greater number of patients on the control achieved through the penalisation of the number of observations at each arm using κ and the chosen prior β_0. Nevertheless, the statistical power is relatively low and can be increased by using higher values of the penalty parameter ($\kappa = 0.65$). It leads to an increase in the power from 0.61 to 0.85 at the cost of the slight ($\approx 4\%$) decrease in the ENS. In fact, $\mathrm{WE}_{\mathrm{Det}}$ then has comparable power to the FR, while treating almost 40 more patients on the superior treatment. Another way to increase the statistical power is to use $\mathrm{WE}_{\mathrm{Ran}}$ for which both the associated power and the ENS is higher than for the FR.

The operating characteristics of the designs in Scenario 2 with fewer patients and different probabilities of response under the alternative is given in Table 3. Under the

Table 2. Operating characteristics of the WE design under the randomised rule ($\mathrm{WE}_{\mathrm{Ran}}$), under the deterministic rule ($\mathrm{WE}_{\mathrm{Det}}$) for different κ (in brackets), GI design and FR in Scenario 1 with $n = 423$ under the null and alternative hypothesises. Results are based on 10^4 replicated trials.

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_0 = p_1 = p_2 = 0.3, p_3 = 0.5$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
GI	0.05	0.25 (0.18)	126.68 (9.4)	0.43	0.83 (0.10)	198.25 (13.7)
FR	0.05	0.25 (0.02)	126.91 (9.4)	0.82	0.25 (0.02)	147.91 (9.6)
$\mathrm{WE}_{\mathrm{Ran}}(0.50)$	0.05	0.24 (0.05)	127.02 (9.5)	0.89	0.39 (0.06)	160.02 (11.1)
$\mathrm{WE}_{\mathrm{Det}}(0.56)$	0.05	0.21 (0.19)	126.89 (9.5)	0.61	0.82 (0.16)	196.63 (17.4)
$\mathrm{WE}_{\mathrm{Det}}(0.65)$	0.05	0.23 (0.13)	126.89 (9.5)	0.85	0.74 (0.11)	189.20 (13.9)

Table 3. Operating characteristics of the WE design under the randomised rule ($\mathrm{WE}_{\mathrm{Ran}}$), under the under deterministic rule ($\mathrm{WE}_{\mathrm{Det}}$) for different κ (in brackets), GI design and FR in Scenario 2 with $n = 80$ under the null and alternative hypothesises. Results are based on 10^4 replicated trials.

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_i = 0.3 + 0.1i, i = 0, 1, 2, 3$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
GI	0.00	0.25 (0.13)	23.97 (4.1)	0.01	0.49 (0.21)	41.60 (5.4)
FR	0.05	0.25 (0.04)	24.02 (4.1)	0.50	0.25 (0.04)	35.98 (4.3)
$\mathrm{WE}_{\mathrm{Ran}}(0.50)$	0.05	0.23 (0.07)	24.01 (4.1)	0.59	0.33 (0.10)	37.55 (4.8)
$\mathrm{WE}_{\mathrm{Det}}(0.51)$	0.05	0.19 (0.16)	23.99 (4.1)	0.36	0.50 (0.28)	41.03 (6.1)
$\mathrm{WE}_{\mathrm{Det}}(0.73)$	0.05	0.22 (0.11)	23.99 (4.1)	0.58	0.44 (0.18)	39.82 (5.2)

null hypothesis, all designs perform similarly in terms ENS and all control the type-I error at the 5% level. Under the alternative hypothesis, the GI and $\mathrm{WE}_{\mathrm{Det}}$ with $\kappa = 0.51$, again, yield the highest (and similar) ENS among all alternatives, but also low statistical power. Note that, for the difference of 35% in power for both approaches, the GI design corresponds to the highly conservative type I error, nearly 0%, against 5% for $\mathrm{WE}_{\mathrm{Det}}(0.51)$. The $\mathrm{WE}_{\mathrm{Ran}}$ or increased κ for $\mathrm{WE}_{\mathrm{Det}}$ result in a considerable power increase. Both designs have a greater (or similar) power and result in more ENS than the FR.

Overall, the WE designs for the robust optimal values of κ perform comparably or with minor differences to the optimal GI design in terms of ENS, but with greater statistical power for both large and small sample sizes. Importantly, the proposed WE design uses an optimal robust value that was previous extensively calibrated and was found to yield beneficial operating characteristics subject to tuning of the penalisation parameter. The ENS and power trade-off can be tuned via the built-in parameter κ. Specifically, for greater values of κ of the randomised rule, the WDE designs can result in similar statistical power to the FR, but with the considerably greater ENS.

4.4. Application of the minimum efficacy bound weight function

The information-theoretic design studied above targets the most effective arm. This design does, however, not take into account that the response probabilities can be to low to be useful. Consequently, the selection of arms should be severely penalised if the response rate is below a minimum clinically interesting value ψ. To account for

Table 4. Operating characteristics of the WE design using the exact information gain under the randomised rule ($\mathrm{WE}_{\mathrm{Ran}}$), under the deterministic rule ($\mathrm{WE}_{\mathrm{Det}}$) for $\kappa = 0.50$ in Scenario 1 with $n = 423$ under the null and alternative hypothesises. Results are based on 2500 replicated trials.

Method	$H_0 : p_0 = p_1 = p_2 = p_3 = 0.3$			$H_1 : p_0 = p_1 = p_2 = 0.3, p_3 = 0.5$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
$\mathrm{WE}_{\mathrm{Det}}(\psi = 0.0)$	0.05	0.22 (0.17)	127.0 (9.6)	0.68	0.78 (0.12)	193.26 (14.7)
$\mathrm{WE}_{\mathrm{Det}}(\psi = 0.3)$	0.05	0.21 (0.20)	127.0 (9.6)	0.70	0.78 (0.15)	193.17 (16.1)
$\mathrm{WE}_{\mathrm{Det}}(\psi = 0.4)$	0.05	0.24 (0.13)	126.8 (9.6)	0.70	0.83 (0.12)	196.84 (13.7)
$\mathrm{WE}_{\mathrm{Det}}(\psi = 0.5)$	0.05	0.24(0.09)	126.9 (9.6)	0.69	0.83 (0.11)	197.25 (15.6)
$\mathrm{WE}_{\mathrm{Det}}(\psi = 0.6)$	0.05	0.24 (0.06)	127.0 (9.6)	0.86	0.67 (0.11)	183.81 (17.0)
$\mathrm{WE}_{\mathrm{Ran}}(\psi = 0.0)$	0.05	0.24 (0.06)	126.6 (9.8)	0.90	0.40 (0.06)	160.76 (10.6)
$\mathrm{WE}_{\mathrm{Ran}}(\psi = 0.3)$	0.05	0.24 (0.10)	126.6 (9.8)	0.91	0.44 (0.07)	163.52 (10.0)
$\mathrm{WE}_{\mathrm{Ran}}(\psi = 0.4)$	0.05	0.23 (0.12)	126.6 (9.8)	0.93	0.60 (0.11)	177.28 (10.9)
$\mathrm{WE}_{\mathrm{Ran}}(\psi = 0.5)$	0.05	0.24 (0.08)	126.6 (9.8)	0.92	0.74 (0.09)	189.07 (12.9)
$\mathrm{WE}_{\mathrm{Ran}}(\psi = 0.6)$	0.05	0.25 (0.06)	126.6 (9.8)	0.89	0.66 (0.10)	182.40 (16.4)

this minimum efficacy value ψ, the weight function $\phi_n^\star$ and the exact information gain criterion given in Equation (13) can be used. In Table 4, we apply this information gain criterion under Scenario 1 with sample size $n = 423$.

While in an actual clinical trial, the minimum efficacy bound, ψ, will be determined by expert's knowledge, we consider different bounds $\psi = 0.0, 0.3, 0.4, 0.5, 0.6$ to investigate how its value affects the operating characteristics. To track the influence of ψ, we study the designs using a fixed value of $\kappa = 0.50$.

For both $\mathrm{WE}_{\mathrm{Ran}}$ and $\mathrm{WE}_{\mathrm{Det}}$, the minimum efficacy bounds $\psi \leq 0.30$ results in similar operating characteristics as the information gain with the weight function ϕ_n without a minimum efficacy value as all treatment arms have greater efficacy probabilities compared to the bound. For $\psi = 0.4$ and $\psi = 0.5$, the first three arms are considered as inefficacious. Comparing to $\psi = 0.30$, the design results in a slightly higher power and in 5% more patients allocated to the superior arm. As ψ increases above the response probability of the superior arm, all treatment arms are considered as inefficacious resulting in more spread allocations (as the information gain is inflated for all arms) and lower ENS.

Overall, the design using the minimum efficacy bound weight function allows to improve both power and ENS if the threshold is correctly specified. It can, however, also lead to a decrease in the ENS if all of the arms are considered as ineffícious.

5. Application to a Phase II Clinical Trial with Co-Primary Efficacy Endpoints

5.1. Setting

In the previous example, a single binary endpoint was used. However, trials with co-primary efficacy endpoints are of growing interest in medical research (Zhou et al., 2017). As the proposed criterion can be applied to a trial with an arbitrary number of discrete outcomes, we investigate the performance of the novel response adaptive design in a setting of a Phase II trial in metastatic breast cancer considered by Song (2015) in this section.

In this Phase II trial, the two key efficacy variables of interest were (i) the tumour objective response rate (ORR) and (ii) the absence of the deterioration in Global Health Status of European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (GHS) in the first two cycles of treatment. As outlined by Song (2015) both endpoints are "relatively rapidly observable" which makes the application of an response-adaptive design suitable. Given two co-primary binary efficacy endpoints, the response observed in each patient has four categories: (a) ORR and GHS, (b) ORR and no GHS, (c) no ORR and GHS, (d) no ORR and no GHS. Then, denoting the probabilities of these events for arm i by $\alpha_j^{(1)}, \alpha_j^{(2)}, \alpha_j^{(3)}, 1-\alpha_j^{(1)}-\alpha_j^{(2)}-\alpha_j^{(3)}$ and the probabilities of the target treatment effect by $\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}, 1-\gamma^{(1)}-\gamma^{(2)}-\gamma^{(3)}$ respectively, the proposed criterion takes the form

$$\delta^{(\kappa)}(\boldsymbol{\alpha}_j, \boldsymbol{\gamma}) := \frac{1}{2}\left(\frac{\left(\gamma^{(1)}\right)^2}{\alpha_j^{(1)}} + \frac{\left(\gamma^{(2)}\right)^2}{\alpha_j^{(2)}} + \frac{\left(\gamma^{(3)}\right)^2}{\alpha_j^{(3)}} + \frac{\left(1-\gamma^{(1)}-\gamma^{(2)}-\gamma^{(3)}\right)^2}{1-\alpha_j^{(1)}-\alpha_j^{(2)}-\alpha_j^{(3)}} - 1\right) n_j^{2\kappa-1}.$$

Extending the setting considered by Song (2015), who considered a single-arm trial, we investigate the behaviour of designs in a more general framework with two treatment arms (indexed by 1 and 2) and a control arm (a standard of care, indexed by 0). Following the sample size considered in the single-arm trial, 55 patients, the sample size in the three-arms trial is fixed to be $n = 55 \times 3 = 165$. Although four outcomes can be observed in the trial, Phase II trials are conventionally formulated in terms of the marginal probabilities of each binary event rather than in terms of the probabilities of joint events. Let $p_{orr,j}$ be the probability of ORR and $p_{ghs,j}$ be the probability of GHS corresponding to the treatment arm j. Motivated by the trial investigated by Song (2015), we consider the following hypothesis: $\mathcal{H}_0 : p_{k,0} = p_{k,j}$ for $j = 1, 2$ and $k = \{orr, ghs\}$. As in the single endpoint example, the hypothesis testing is performed using Fishers adjusted exact test, where the adjustment chooses the cutoff value to achieve a 5% type-I error (Villar et al., 2015a). Again, the Bonferroni correction is used to ensure that the family wise error rate is less or equal than 5%. Characteristics of interest are (i) the type-I error rate (α), (ii) statistical power ($1-\eta$), (iii) the average proportion of patients on the optimal treatment (p^*) and (iv) the expected number of ORR (ENS) (the expected number of GHS is suppressed for the sake of space).

5.2. Design Specification and Comparators

To adapt the novel criterion to the formulated context, we employ a re-parametrisation of the probabilities of joint events under the assumption of independence. Then, the target of the trial can be defined in terms of the target probability of ORR, γ_{orr}, and the target probability of GHS, γ_{ghs}. We define the probabilities of events for arm j as $\alpha_j^{(1)} = p_{orr,j} p_{ghs,j}$, $\alpha_j^{(2)} = p_{orr,j}\left(1-p_{ghs,j}\right)$, $\alpha_j^{(3)} = \left(1-p_{orr,j}\right) p_{ghs,j}$, and the corresponding targets as $\gamma^{(1)} = \gamma_{orr}\gamma_{ghs}$, $\gamma^{(2)} = \gamma_{orr}\left(1-\gamma_{ghs}\right)$, $\gamma^{(3)} = \left(1-\gamma_{orr}\right)\gamma_{ghs}$.

Following the single agent example, we specify the parameters for the proposed response-adaptive design as follows. As the upper bound for the ORR and GHS is not defined, the target values $\gamma_{orr} = \gamma_{ghs} = 1$ are taken to ensure that the arm corresponding to the highest probability is chosen. Given the re-parametrisation, both probabilities are considered as Beta random variables. The vectors of the prior mode probabilities $p_{orr}^{(0)} = [0.99, 0.99, 0.99]^{\mathrm{T}}$ and $p_{ghs}^{(0)} = [0.99, 0.99, 0.99]^{\mathrm{T}}$ are chosen to reflect equipoise. Again, we choose the following parameters of the Beta distribution for both probabilities: $\beta_0 = 5$ to ensure enough observations on the control and $\beta_1 = \beta_2 = 2$ to reflect no prior knowledge for competing arms. As before, we fix $\kappa = 0.5$ for the deterministic allocation rule and use the robust optimal values of κ for $n = 165$ obtained in Section 4.

We compare the performance to two alternative approaches: the first is a MAB approach that prioritises the exploitation objective and, therefore, is expected to result in high ENS; and the second one that is known to result in high power. The MAB approach seeking to obtain high ENS described below is referred to as "Max Prob". As for the proposed information-theoretic approach, under the independence assumption, we consider each efficacy endpoints as Beta random variables and assign each subsequent patient to the arm that corresponds to the maximum probability of having the highest p_{orr} and the highest p_{ghs} together (Wathen and Thall, 2017). Formally, the next patient is allocated to the treatment arm $j^\star$ such that $j^\star = \arg\max_j \left[\mathbb{P}\left(p_{orr,j} = \max_i(p_{orr,i})\right) \times \mathbb{P}\left(p_{ghs,j} = \max_i(p_{ghs,i})\right)\right]$. The `R`-package `bandit` is used to compute these probabilities (Lotze and Loecher, 2014). The design uses the same prior distribution as the proposed design. Fixed and equal randomisation (FR) is use as comparator expected to achieve high power.

Although, the designs' constructions employ the assumption of independence between efficacy endpoints, it is unlikely to be true in an actual trial. Therefore, we generate correlated efficacy endpoints in the simulation study using the approach by Tate (1955). We generate a bivariate standard Normal vector (x_{orr}, x_{ghs}) with mean $\mu = (0, 0)$ and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{14}$$

where ρ is the correlation coefficient. By applying the CDF of the standard Normal random variable marginally $(u_{orr}, u_{ghs}) = (\Phi(x_{orr}, \Phi(x_{ghs}))$, one can obtain two correlated random variables having Uniform distributions that subsequently can then be transformed to binary ones. As a strong correlation is anticipated between the co-primary efficacy endpoints, the correlation coefficient of $\rho = 0.75$ is chosen. The results for other values of the correlation, ρ (including the case of no correlation) are given in Supplementary Materials.

5.3. Results

The operating characteristics for two possible cases (based on the original study) are given in Table 5. The results show that the performance of the novel response-adaptive design are qualitatively similar to the previous example. Under the null hypothesis (Scenario 1), the performance of all methods is similar and the type-I error is controlled. Under the alternative hypotheses (Scenario 2), the $\mathrm{WE}_{\mathrm{Det}}$ design with the calibrated value of $\kappa = 0.54$ results in a similar proportion of patients assigned to the superior arm as the Max Prob (0.70 against 0.71) and nearly the same average number of ORR observed in the trial. However, as in the previous example, the $\mathrm{WE}_{\mathrm{Det}}$ design results in

Table 5. Operating characteristics of the WE design under the randomised rule ($\mathrm{WE}_{\mathrm{Ran}}$), under the deterministic rule ($\mathrm{WE}_{\mathrm{Det}}$) for different κ (in brackets), $\mathrm{Max\ Prob}$ design and FR in the trial with two co-primary efficacy endpoints and $n = 165$ under the null and alternative hypothesises. Results are based on 10^4 replicated trials.

	Scenario 1			Scenario 2		
	$p_{orr,1} = p_{orr,2} = p_{orr,3} = 0.10$			$p_{orr,1} = p_{orr,2} = 0.10, p_{orr,3} = 0.25$		
Method	$p_{ghs,1} = p_{ghs,2} = p_{ghs,3} = 0.45$			$p_{ghs,1} = p_{ghs,2} = 0.45, p_{ghs,3} = 0.60$		
	α	$p^*(s.e)$	ENS(s.e.)	$(1-\eta)$	$p^*(s.e.)$	ENS (s.e.)
Max Prob	0.05	0.25 (0.16)	16.45 (3.9)	0.33	0.71 (0.18)	33.95 (7.2)
FR	0.05	0.33 (0.04)	16.50 (3.8)	0.62	0.33 (0.04)	24.78 (4.5)
$\mathrm{WE}_{\mathrm{Ran}}(0.50)$	0.05	0.29 (0.07)	16.50 (3.8)	0.67	0.46 (0.08)	27.87 (5.4)
$\mathrm{WE}_{\mathrm{Det}}(0.54)$	0.05	0.24 (0.16)	16.45 (3.9)	0.49	0.70 (0.17)	33.92 (7.2)
$\mathrm{WE}_{\mathrm{Det}}(0.69)$	0.05	0.27 (0.12)	16.45 (3.9)	0.63	0.64 (0.12)	32.35 (6.5)

higher statistical power (0.49 against 0.33). To increase power (at the cost of lower ORR rate) a higher value of κ can be used or randomisation between arms can be employed. The value $\kappa = 0.69$ using $\mathrm{WE}_{\mathrm{Det}}$ leads to an increase in power to 0.63 for the cost of a slight decrease in the ENS by approximately 1 treated patients. Moreover, $\mathrm{WE}_{\mathrm{Det}}(0.69)$ implies nearly the same power as FR but results in nearly 8 more patients treated on the better treatment on average. Alternatively, using the randomised assignment rule, the proposed design results in even higher statistical power (0.67) but in 5 fewer ENS compared to $\mathrm{WE}_{\mathrm{Det}}(0.69)$ which is still higher than for FR.

Overall, the example with co-primary efficacy endpoints supports the previously found results. For the tuned robust optimal values of the penalisation parameter, the proposed design can perform comparably in terms of ENS to the MAB design that prioritises exploitation but can outperform it in terms of the power. Moreover, the WE design can result in similar power as the FR but noticeably greater ENS.

6. Discussion

In this work, we proposed a general criterion for the selection of arms in experiments with multinomial outcomes that is based on the weighted information measure and is of particular use in the setting with ethical and strict sample size constraints. We considered two families of weight functions and demonstrated how the proposed criterion can be used for an arbitrary weight function reflecting various objectives of experiments that are of interested to investigators. For the considered weight functions, the information gain criteria preserve the flexibility and allow to tailor the design parameters in light of the exploration-exploitation trade-off. The design parameters should be carefully tuned prior to the design application to ensure desirable statistical properties of the design with high probability and competitive advantages over the design considered in this work. Such a tuning procedure to find the optimal robust design for a generic objective function is proposed.

The prior distribution used in the illustrative examples was chosen to protect allocation of patients to control in a non-ruled based manner – by design itself. However, alternative specifications of prior distribution can be considered. In general, a prior distribution that does not secure more patients on the control (either through β_0 or

$p^{(0)}$) will require higher values of the penalisation parameter κ to reach the same level of power compared to the prior ensuring more patients on the control. In fact, the design under each of these prior would require the search of the robust optimal penalisation parameter κ as described in Section 3. We refer the reader to Supplementrary Materials for an evaluation of different prior distributions on the properties of the design for various values of κ. It was found that given the investigator's preferences in the power-ENS balance, other prior distributions can provide gains similar to the ones found for the considered prior assumptions.

Throughout the paper, we have intensionally focused on two competing methods only under each example to provide a benchmark for comparison and to focus on the proposed method. In the provided examples above, it was found when compared to some MAB approaches that favour exploitation, the proposed design for the found robust optimal values of κ and considered prior distribution can yield better power while resulting only in minor reduction in ENS. At the same time, there are other modified MAB approaches that were proposed in the literature that could be applied to the considered problems and can result in a better power-ENS balance than original counterparts. Specifically, Villar et al. (2015b) proposed a randomised version of the GI design to tackle the exploitation-exloration trade-off. Furthermore, Villar et al. (2015a) proposed the GI modification that imposes a rule-based mechanism on controlling number of patients on the control treatment and found that it leads to a noticeable improvement in power while only minor losses in the ENS. A similar controlling procedure could be imposed on the proposed design. Similarly, there is a GI index defined for multinomial outcomes (Glazebrook, 1978) that could be an alterantive approach for the problem with co-primary outcome studied in Section 5. A comprehensive comparison of these procedured in a large number of potential simulation scenarios is of interest and is subject of future research.

Throughout the work, the examples concerned Phase II clinical trials evaluating efficacy. However, the design was also found to provide benefits in the setting of Phase I clinical trials seeking to select the MTD (i.e. the target probability γ is the toxicity probability at the MTD), particularly when the assumption of monotonicity is questionable. We refer the reader to the Supplementary Materials for the corresponding results. Therefore, while Phase I and Phase II trials state two different questions, the general formulation of the proposed design (to target the TA with specific characteristics) enable its application in both setting and in a wide range of trials.

In the presented evaluations a fixed target value of $\gamma = 0.999$ was considered. At the same time, there are many clinical settings in which the maximum clinically feasible efficacy probability can be specified prior to the trial. We study the effect of the target value in many different scenarios in the Supplementary Materials. Setting the target below the true response probability results in targeting an inferior arm and worsen the performance both in terms of ENS and statistical power. Therefore, it is preferred to be more conservative and ensure that the target probability is high enough. Studying various target values, we found that, under the "Select-the-Best" allocation rule, the influence of the target value on the operating characteristics is small. Under the "Randomization" allocation rule, however, specification of the target value close to the true maximum value yields noticeably more patients on the superior arm.

An important assumption employed by the proposed response-adaptive information-theoretic design (as well as for the majority of alternative response-adaptive procedures Villar et al., 2015a) is that the patients' responses are observed shortly after the treatment, or at least before a next patient is to be enrolled in the study. This, however, might not hold in many clinical trials. Consequently, the question of the delayed responses incorporation is of great practical interest.

In this work, multinomial outcomes were considered only. Generalising the proposed approach to experiments with continuous outcomes is subject of future research together with its non-parametric extension.

While clinical trials have been the main motivation for this work, the design can be applied to a wide range of problems of similar nature. For example, applications where the MAB approach has found the applications: online advertising, portfolio design, queuing and communication networks, etc. (see Gittins et al., 2011, and references there in). In these settings, however, the sample size is not one of the main constraints in constrast to the clinical trial setting considered in this work. Nevertheless, the general principles proposed can be applied in these problems and their merits in the setting with easy-to-collect observation is to be studied. On top of that, the proposed design can be used in more general problems of selecting an arm corresponding to target value γ rather than the selection of the highest success probability only. It is important to emphasize that the derived selection criterion can be also applied in conjunction with parametric models which also expands its possible applications. In fact, the parameters can be estimated by any desirable method and then 'plugged-in' in the criterion which preserves its properties.

Acknowledgement

The authors acknowledge the insightful and constructive comments made by the editor, associate editor and three reviewers. These comments have greatly helped us to refine the original submission. This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567, and is independent research supported by the National Institute for Health Research (NIHR Advanced Fellowship, Dr Pavel Mozgunov, NIHR300576) and by Prof Jaki's Senior Research Fellowship (NIHR-SRF-2015-08-001) supported by the National Institute for Health Research. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. TJ is also supported by UK Medical Research Council (grant number: MC UU 0002/14).

References

Agresti, A. (1992) A survey of exact inference for contingency tables. *Stat. Sc.*, 131–153.

Aitchison, J. (1982) The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 139–177.

— (1992) On criteria for measures of compositional difference. *Mathematical Geology*, **24**, 365–379.

Azriel, D., Mandel, M. and Rinott, Y. (2011) The treatment versus experimentation dilemma in dose finding studies. *Journal of Statistical Planning and Inference*, **141**, 2759–2768.

Barrett, J. E. (2016) Information-adaptive clinical trials: a selective recruitment design. *JRSS: Series C*, **65**, 797–808.

Belis, M. and Guiasu, S. (1968) A quantitative-qualitative measure of information in cybernetic systems (corresp.). *IEEE Transactions on Information Theory*, **14**, 593–594.

Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S. and Colton, S. (2012) A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, **4**, 1–43.

Clertant, M. and O'Quigley, J. (2017) Semiparametric dose finding methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 1487–1508.

Clim, A. (2008) Weighted entropy with application. *Anal. Bucuresti, Mat., Anul*, **57**, 223–231.

Cover, T. M. and Thomas, J. A. (2012) *Elements of information theory.* John Wiley.

Djulbegovic, B., Lacevic, M., Cantor, A., Fields, K. K., Bennett, C. L., Adams, J. R., Kuderer, N. M. and Lyman, G. H. (2000) The uncertainty principle and industry-sponsored research. *The Lancet*, **356**, 635–638.

Dunnett, C. W. (1984) Selection of the best treatment in comparison to a control with an application to a medical trial. *Design of experiments: Ranking and selection*, 47–66.

Gittins, J., Glazebrook, K. and Weber, R. (2011) *Multi-armed bandit allocation indices.* John Wiley & Sons.

Gittins, J. C. and Jones, D. M. (1979) A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 561–565.

Glazebrook, K. (1978) On the optimal allocation of two or more treatments in a controlled clinical trial. *Biometrika*, **65**, 335–340.

Guo, B., Li, Y. and Yuan, Y. (2016) A dose–schedule finding design for phase I–II clinical trials. *Journal of the Royal Statistical Society: Series C*, **65**, 259–272.

Iasonos, A., Wages, N. A., Conaway, M. R., Cheung, K., Yuan, Y. and O'Quigley, J. (2016) Dimension of model parameter space and operating characteristics in adaptive dose-finding studies. *Statistics in Medicine*, **35**, 3760–3775.

Jones, D. (1975) *Search Procedures for Industrial Chemical Research.* Ph.D. thesis, University of Cambridge.

Jones, D. M. (1970) *Sequential Method for Industrial Chemical Research.* Ph.D. thesis, University of Wales.

Kelbert, M. and Mozgunov, P. (2015) Asymptotic behaviour of the weighted Renyi, Tsallis and Fisher entropies in a Bayesian problem. *Eurasian Mathematical Journal*, **6**, 6–17.

— (2017) Generalization of Cramér-Rao and Bhattacharyya inequalities for the weighted covariance matrix. *Mathematical Communications*, **22**, 25–40.

Kelbert, M., Suhov, Y., Izabella, S. and Yasaei, S. S. (2016) Basic inequalities for weighted entropies. *Aequationes Mathematicae*, 1–32.

Kim, S. B. and Gillen, D. L. (2016) A Bayesian adaptive dose-finding algorithm for balancing individual-and population-level ethics in Phase I clinical trials. *Sequential Analysis*, **35**, 423–439.

Klotz, J. (1978) Maximum entropy constrained balance randomization for clinical trials. *Biometrics*, 283–287.

Koenig, F., Brannath, W., Bretz, F. and Posch, M. (2008) Adaptive Dunnett tests for treatment selection. *Statistics in Medicine*, **27**, 1612–1625.

Lee, S. M., Ursino, M., Cheung, Y. K. and Zohar, S. (2017) Dose-finding designs for cumulative toxicities using multiple constraints. *Biostatistics*.

Lotze, T. and Loecher, M. (2014) *bandit: Functions for simple A/B split test and multi-armed bandit analysis*. URL: `https://CRAN.R-project.org/package=bandit`. R package version 0.5.0.

Magirr, D., Jaki, T. and Whitehead, J. (2012) A generalized Dunnett test for multi-arm clinical studies with treatment selection. *Biometrika*, **99**, 494.

Mozgunov, P. and Jaki, T. (2019) An information theoretic phase I–II design for molecularly targeted agents that does not require an assumption of monotonicity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **68**, 347–367.

— (2020) Improving safety of the continual reassessment method via a modified allocation rule. *Statistics in Medicine*, **Epub**.

Mozgunov, P., Jaki, T. and Gasparini, M. (2019) Loss functions in restricted parameter spaces and their Bayesian applications. *Journal of Applied Statistics*, 1–24.

O'Quigley, J., Iasonos, A. and Bornkamp, B. (2017) Handbook of Methods for Designing, Monitoring, and Analyzing Dose-Finding Trials. In *Handbook of Methods for Designing, Monitoring, and Analyzing Dose-Finding Trials*. CRC Press, Taylor and Francis Group.

Pushpakom, S., Kolamunnage-Dona, R., Taylor, C., Foster, T., Spowart, C., García-Fiñana, M., Kemp, G. J., Jaki, T., Khoo, S., Williamson, P. et al. (2019) TAILoR (TelmisArtan and InsuLin Resistance in Human Immunodeficiency Virus [HIV]): An Adaptive-design, Dose-ranging Phase IIb Randomized Trial of Telmisartan for the Reduction of Insulin Resistance in HIV-positive Individuals on Combination Antiretroviral Therapy. *Clinical Infectious Diseases*.

Ristl, R., Urach, S., Rosenkranz, G. and Posch, M. (2018) Methods for the analysis of multiple endpoints in small populations: A review. *Journal of Biopharmaceutical Statistics*, 1–29.

Riviere, M.-K., Dubois, F. and Zohar, S. (2015) Competing designs for drug combination in phase I dose-finding clinical trials. *Statistics in Medicine*, **34**, 1–12.

Shannon, C. E. (1948) A mathematical theory of communication. *Bell system technical journal*, **27**, 379–423.

Smith, A. L. and Villar, S. S. (2018) Bayesian adaptive bandit-based designs using the gittins index for multi-armed trials with normally distributed endpoints. *Journal of applied statistics*, **45**, 1052–1076.

Song, J. X. (2015) A two-stage design with two co-primary endpoints. *Contemporary Clinical Trials Communications*, **1**, 2–4.

Stallard, N. and Todd, S. (2003) Sequential designs for phase III clinical trials incorporating treatment selection. *Statistics in Medicine*, **22**, 689–703.

Tate, R. F. (1955) The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, **42**, 205–216.

Thall, P. F. and Cook, J. D. (2004) Dose-finding based on efficacy–toxicity trade-offs. *Biometrics*, **60**, 684–693.

Villar, S. S., Bowden, J. and Wason, J. (2015a) Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science*, **30**, 199.

— (2018) Response-adaptive designs for binary responses: How to offer patient benefit while being robust to time trends? *Pharmaceutical Statistics*, **17**, 182–197.

Villar, S. S., Wason, J. and Bowden, J. (2015b) Response-adaptive randomization for multi-arm clinical trials using the forward looking Gittins index rule. *Biometrics*, **71**, 969–978.

Wages, N. A., Conaway, M. R. and O'Quigley, J. (2011) Continual reassessment method for partial ordering. *Biometrics*, **67**, 1555–1563.

Wages, N. A., O'Quigley, J. and Conaway, M. R. (2014) Phase I design for completely or partially ordered treatment schedules. *Statistics in Medicine*, **33**, 569–579.

Wathen, J. K. and Thall, P. F. (2017) A simulation study of outcome adaptive randomization in multi-arm clinical trials. *Clinical Trials*, **14**, 432–440.

Whitehead, J. and Williamson, D. (1998) Bayesian decision procedures based on logistic regression models for dose-finding studies. *J. of Biopharm. Stat.*, **8**, 445–467.

Williamson, S. F., Jacko, P., Villar, S. S. and Jaki, T. (2016) A Bayesian adaptive design for clinical trials in rare diseases. *Computational Statistics & Data Analysis.*

Williamson, S. F. and Villar, S. S. (2019) A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. *Biometrics.*

Zhou, H., Lee, J. J. and Yuan, Y. (2017) BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. *Statistics in Medicine*, **36**, 3302–3314.