

Response-Adaptive Randomization for Multi-arm Clinical Trials Using the Forward Looking Gittins Index Rule

Sofía S. Villar,^{1,2,*} James Wason,^{1,**} and Jack Bowden^{1,***}

¹MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 0SR, U.K.

²Department of Mathematics and Statistics, Lancaster University, Bailrigg, Lancaster LA1 4YW, U.K.

* *email*: sofia.villar@mrc-bsu.cam.ac.uk

** *email*: james.wason@mrc-bsu.cam.ac.uk

*** *email*: jack.bowden@mrc-bsu.cam.ac.uk

SUMMARY. The Gittins index provides a well established, computationally attractive, optimal solution to a class of resource allocation problems known collectively as the multi-arm bandit problem. Its development was originally motivated by the problem of optimal patient allocation in multi-arm clinical trials. However, it has never been used in practice, possibly for the following reasons: (1) it is fully sequential, i.e., the endpoint must be observable soon after treating a patient, reducing the medical settings to which it is applicable; (2) it is completely deterministic and thus removes randomization from the trial, which would naturally protect against various sources of bias. We propose a novel implementation of the Gittins index rule that overcomes these difficulties, trading off a small deviation from optimality for a fully randomized, adaptive group allocation procedure which offers substantial improvements in terms of patient benefit, especially relevant for small populations. We report the operating characteristics of our approach compared to existing methods of adaptive randomization using a recently published trial as motivation.

KEY WORDS: Bayesian adaptive designs; Clinical trials; Gittins index; Multi-armed bandit; Sequential allocation.

1. Introduction

Consider a clinical trial to test the effectiveness of several treatments on a group of patients. A rational observer, who is unfamiliar with established medical convention, might well suppose that patients would be allocated to treatments with the aim of optimizing their collective health. Indeed, using this idea as the initial motivation, Gittins and Jones (1979) first proposed an optimal, deterministic rule to perform such a task within a multi-arm clinical trial, which is termed the *Gittins index*. Since then, the Gittins index and its extensions have been extensively used to address resource allocation problems (collectively known as “multi-arm bandit” models) in a variety of scientific fields. Applications include: queuing theory and optimal scheduling (Aalto et al., 2009), computer-communication networks (Cechi and Jacko, 2013) and internet marketing (Hauser et al., 2009), among others.

A large body of literature has accumulated on the use of bandit models to address ethical issues in trial design (see e.g., Berry and Fristedt, 1985; Berry and Eick, 1995; Press, 2009). Ironically, bandits, and more specifically the Gittins index, have never been used in clinical practice (to the best of the authors’ knowledge) due to certain realities of medical research. These are discussed at length in a recent review by Villar et al. (2015), the main findings of which are now summarized.

The most immediate limitation is practical: in order to apply the Gittins index each patient’s outcome needs to be observed before the next patient is assigned. This rules out the

vast majority of medical conditions and diseases where the outcome of treatment is observed with non-negligible delay. Another major barrier, common to all optimal decision-analytical designs (Cheng and Berry, 2007), is that the deterministic nature of the Gittins index runs contrary to the predominant mode of clinical trial design and analysis over the last 60 years. Under this framework patients are randomized to different treatment arms with equal probability. This has several desirable consequences. First, it guarantees (asymptotically) that patient groups will be balanced with respect to all characteristics (known or unknown), thus reducing the potential for confounding bias. Second, it provides a perfect vehicle to facilitate a double-blind experiment, making it harder for the trial organizers to induce allocation bias (see e.g., Chow and Chang, 2008). Third, fixed and equal randomization maximizes the learning about all treatments, i.e., it maximizes power to detect a significant difference between any two treatments in the trial by assigning similar numbers of patients to both. Having a high power is usually the primary aim of a trial and relegates the benefit to patients (within the trial) into second place. However, there are signs that this established view, as summarized in the Belmont report (Beauchamp, 2008), is changing. It is now not as common in rare disease settings, for example in pediatric oncology (Wang and Arnold, 2002).

The reduction in statistical power caused by adaptive allocation is not unique to these index rules, rather it affects all adaptive allocation procedures to a greater or lesser extent

(Hu and Rosenberger, 2003). It is simply not possible to simultaneously maximize trial power while maximizing patient outcomes by skewing treatment allocation. That is why an ideal adaptive design would seek to assign more patients to the better treatment with the minimum sacrifice of inferential power.

In a multi-arm setting in which several treatments are assessed simultaneously against a shared control group, adaptive randomization does not face the power–patient benefit trade-off as separate two-arm trials do. The power of an adaptive multi-arm design can be increased by protecting the allocation to the control treatment (Trippa et al., 2012). By incorporating this feature, as shown in Villar et al. (2015), one can identify adaptive trial designs that score highly in both statistical power and patient benefit. Moreover, by testing many new promising treatments at the same time, multi-arm trials increase the probability of finding a successful new treatment and speed up the process of doing so.

This article addresses the remaining two limitations, i.e., its determinism and its fully sequential nature. We present an implementation of the Gittins index that introduces randomization, trading off a small amount of its optimality to gain applicability to real clinical trials. This is achieved by a modified algorithm that is probabilistic and is applied to blocks of patients rather than individuals, as suggested by Rosenberger and Lachin (1993).

One of the first works on randomized bandits was Yang and Zhu (2002). The authors consider a multi-arm problem with covariates and introduce a myopic randomized solution that is asymptotically consistent. Instead, we consider a problem without covariates and concentrate on introducing randomization to the Gittins index, by taking into account future sequences of allocations under Gittins' rule. Our simple, heuristic approach thereby aims to achieve a near-optimal mean total rewards criterion in a computationally feasible way.

The closest idea we know of in the literature is in Cheng and Berry (2007), where the introduction of randomization into an optimal design is considered. Treatment decisions are made using a decision–analytic approach whose goal is to maximize overall successful patient treatment over the patient population under the constraint that each arm must have at least probability r of being assigned to patients in the trial. Our algorithm does not impose any limitation on the allocation probabilities. This allows for arms to be dropped or promoted within a trial if there is enough evidence.

In Section 2, we introduce the Gittins index and demonstrate how it can be applied in a clinical trial setting with a binary outcome. We introduce our probabilistic implementation of the Gittins index, which we call the forward looking Gittins index (FLGI), in Section 3. In Section 4, we compare our approach to alternative procedures by performing simulations in the context of a recently published clinical trial: NeoSphere, (Gianni et al., 2012). We conclude with a discussion of our results and point to future research in Section 5.

2. Methods: the Gittins Index

Consider a clinical trial to test the effectiveness of K experimental treatments against a control treatment on a fixed sample of T patients. When a patient t ($t = 1, \dots, T$) is allocated to

treatment k ($k = 0, \dots, K$), where $k = 0$ denotes the control, then a binary outcome $Y_{k,t}$ is observed with $\Pr(Y_{k,t} = 1) = p_k$, the true unknown response probability for treatment k . Assume that $Y_{k,t} = 1$ if the treatment is successful and 0 otherwise. Patients enter the trial one-by-one and the outcome for patient t is observed before patient $t + 1$ appears.

Let each parameter p_k be assigned a $\text{Beta}(s_{k,0}, f_{k,0})$ prior density at the start of the trial, where $s_{k,0}$ and $f_{k,0}$ denote prior beliefs about the relative chances of success and failure of treatment k respectively. Given the conjugacy of the prior- and Bernoulli-distributed outcome, these priors are converted into Beta posteriors for p_k via Bayes theorem as patients enter the trial, which are assigned to a treatment arm and subsequently experience a success or a failure. Let $\mathbf{X}_{k,t} = (s_{k,0} + S_{k,t}, f_{k,0} + F_{k,t})$ be the two state-vector of available information on treatment k at time t , where the random vector $(S_{k,t}, F_{k,t})$ represents the number of successful and unsuccessful outcomes for arm k up to patient t . The posterior for p_k after having treated patient t and observing $s_{k,t}$ successes and $f_{k,t}$ failures is $f(p_k | \mathbf{x}_{k,t}) \sim \text{Beta}(s_{k,0} + s_{k,t}, f_{k,0} + f_{k,t})$ with its posterior mean being $\mathbb{E}[p_k | \mathbf{x}_{k,t}] = \frac{s_{k,t} + s_{k,0}}{s_{k,t} + f_{k,t} + s_{k,0} + f_{k,0}}$. Finally, let $a_{k,t}$ be the binary indicator variable denoting whether patient $t + 1$ is assigned to treatment k or not. The multi-armed bandit optimization problem is to find an allocation rule π such that:

$$V_D^*(\tilde{\mathbf{x}}_0) = \max_{\pi \in \Pi} \mathbb{E}^\pi \left[\left(\sum_{t=0}^{T-1} \sum_{k=0}^K d^t \mathbb{E}[p_k | \mathbf{x}_{k,t}] a_{k,t} \right) \middle| \tilde{\mathbf{x}}_0 \right]. \quad (1)$$

where $\tilde{\mathbf{x}}_0 = (\mathbf{x}_{k,0})_{k=0}^K$ is the initial joint state with all the prior parameters, $\mathbb{E}^\pi[\cdot]$ denotes expectation under allocation rule π , and d is a discount factor (i.e., $0 \leq d < 1$) introduced for reasons of tractability so that a trial of infinite size ($T = \infty$) can be assumed. Thus, $V_D^*(\tilde{\mathbf{x}}_0)$ is the optimal expected total-discounted value function conditional $\tilde{\mathbf{x}}_0$ over Π , the family of admissible allocation rules (i.e., those for which it holds that $\sum_{k=0}^K a_{k,t} = 1$ for all t). Put simply, (1) is the maximum average (discounted) number of patients responses attainable given the initial information on the available treatments before the start of the trial.

Since rewards are geometrically discounted, i.e., patient t 's success yields a reward of d^t , the choice of d plays a fundamental role. To see this, interpret d as the probability of the trial continuing after a given patient, so that $(1 - d)$ is the probability of a trial terminating after patient t . If d is small (say 0.5), then optimization is done for an small expected number of patients (say 15, since $(1 - 0.5)^{15} < 10^{-4}$). An index for the undiscounted finite horizon problem can be considered, though it introduces further complexities (Villar et al., 2015).

The solution to (1) can be found via dynamic programming, using a backward induction algorithm. This becomes computationally infeasible for relatively small values of T and K and is extremely difficult to implement in practice, since it specifies an arm to use in each possible trial accumulated history. An elegant and computationally tractable solution to (1) when considered for $T = \infty$ was given by Gittins and Jones (1979). They showed that the optimal rule obtained by backward induction is equivalent to the rule of allocating patient $t + 1$ to the arm with the highest Gittins index at time t . For

treatment k , this is denoted $\mathcal{G}(\mathbf{x}_{k,t})$ where:

$$\mathcal{G}(\mathbf{x}_{k,t}) = \sup_{\tau \geq 1} \frac{\mathbb{E} \left[\left(\sum_{i=0}^{\tau-1} \mathbb{E}[p_k | \mathbf{x}_{k,t+i}] d^i \right) | \mathbf{x}_{k,t} \right]}{\mathbb{E} \left[\left(\sum_{i=0}^{\tau-1} d^i \right) | \mathbf{x}_{k,t} \right]}. \quad (2)$$

Gittins and Jones (1979) proved that one can attach an index $\mathcal{G}(\mathbf{x}_{k,t})$ to each treatment k , which depends only on its current information state, such that the optimal action for patient i is to be allocated to the treatment whose current index is greatest. The Gittins index is calculated by solving the problem of allocating patients optimally between treatment k and a known treatment which yields a constant reward. Calculations of the indices (2) have been reported in tables as in Gittins et al. (2011). A detailed explanation of how the Gittins index rule is deployed and a table with values are given in Tables 4 and 5 of Web Appendix A.

3. The Forward-Looking Gittins Index Algorithm

Assume that instead of enrolling T patients one-by-one, patients are enrolled in groups of size b over J stages, so that $J \times b = T$. We wish to specify a rule based on the Gittins index that sequentially randomizes the next b patients among the $K+1$ treatments at stage j ($j=1, \dots, J$) given the data up to block $j-1$. This translates to determining $\pi_{k,j}$, where: $\pi_{k,j}$ = the probability of allocation to treatment k at stage j ($j=1, \dots, J$), which is common to all patients in block j , when using the Gittins index and given all data observed up to the stage $j-1$ (and therefore patient $(j-1) \times b$), denoted by $\tilde{\mathbf{x}}_{(j-1)b}$. Note that $\tilde{\mathbf{x}}_{(j-1)b}$ can be written as a $(K+1) \times 2$ matrix in which row k represents the parameters of treatment k 's current posterior distribution up to patient $(j-1)b$. This marginal probability is obtained via the following formula:

$$\pi_{k,j} = \frac{1}{b} \sum_{t=(j-1)b+1}^{jb} \left[\sum_{\tilde{\mathbf{x}}_{t-1} \in \Omega_{t-1}} \Pr(a_{k,t}^{\text{GI}} = 1 | \tilde{\mathbf{X}}_{t-1} = \tilde{\mathbf{x}}_{t-1}) \times \Pr(\tilde{\mathbf{X}}_{t-1} = \tilde{\mathbf{x}}_{t-1} | \tilde{\mathbf{X}}_{(j-1)b} = \tilde{\mathbf{x}}_{(j-1)b}) \right] \quad (3)$$

Here Ω_{t-1} represents the set of all possible values for $\tilde{\mathbf{X}}_{t-1}$ given initial data $\tilde{\mathbf{x}}_{(j-1)b}$ for every future patient t in $(j-1)b+1, \dots, jb$ under the Gittins index rule (summarized by $a_{k,t}^{\text{GI}}$). Each term of the summation within the square brackets of equation (3) represents the joint probability of allocating a future patient t in block j to treatment k and the current information state at patient $t-1$, given the data at the beginning of block $j-1$. Notice that

- For the initial stage, i.e., for $j=1$, $\tilde{\mathbf{x}}_{(j-1)b}$ contains only the prior beliefs on the effectiveness of each treatment. When all treatments are assigned identical priors, it is clear that $\pi_{k,j} = \frac{1}{K+1}$ for all k . This is proved in Web Appendix B.

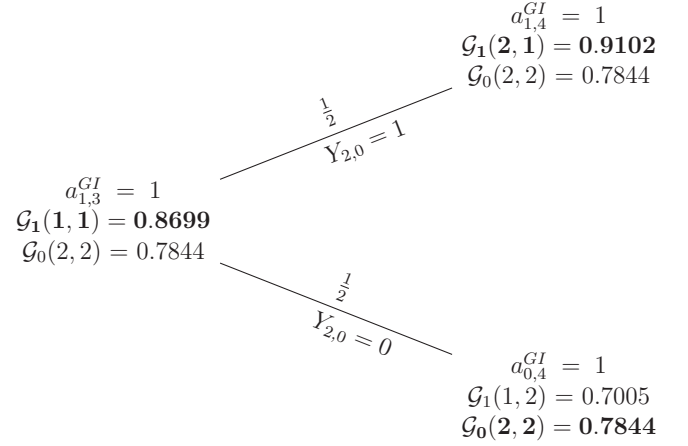


Figure 1. The FLGI rule and a probability tree of all trial histories using the Gittins rule when $K=2$, $b=2$, and $\tilde{\mathbf{x}}_2 = [(2,2); (1,1)]$. Bold text indicates the allocated treatment under the Gittins index rule $\{a_{k,t}^{\text{GI}}\}$. Gittins index values used correspond to $d=0.99$ (See Table 4 in Web Appendix A).

- For $j \geq 2$, the allocation decision for the first patient in block j (i.e., for patient $t = (j-1)b + 1$) is deterministic if a single treatment has a unique maximum Gittins index given $\tilde{\mathbf{x}}_{(j-1)b}$. However, if multiple treatments $k^* \in \{0, \dots, K\}$ are joint maxima, ties are broken at random among the k^* .
- For $j \geq 2$ and for subsequent patients in block j ($t = (j-1)b + 2, \dots, jb$), allocation probabilities are determined by averaging over the posterior predictive distribution of future data given $\tilde{\mathbf{x}}_{(j-1)b}$. For Bernoulli data and Beta-prior, this is the well known Beta-Binomial distribution. Again, ties between treatments are broken at random.

3.1. Worked Example

To clarify how these probabilities are computed and their computational cost, we present a simple example of a two-arm trial testing a control treatment ($k=0$) against an experimental treatment ($k=1$) with a block of size 2 (i.e., $b=2$ and $K=2$). Suppose further that in the first block of the trial, having started with Beta(1,1) priors, the resulting allocation was, two patients to control—one success and another one a failure—and no patient to the experimental treatment. Therefore, for the second block the priors for each treatment are $\tilde{\mathbf{x}}_2 = [(2,2); (1,1)]$.

Figure 1 illustrates, via a probability tree, how the FLGI probabilities for block 2 given the data in block 1 are computed. Given that the experimental treatment would have the maximum Gittins index, the allocation decision for the first patient of the second block (i.e., patient 3) is deterministic. It follows that $\Pr(a_{0,3}^{\text{GI}} = 1 | \tilde{\mathbf{x}}_2) = 0$ and $\Pr(a_{1,3}^{\text{GI}} = 1 | \tilde{\mathbf{x}}_2) = 1$.

When the second patient of the second block is to be allocated (i.e., patient 4), given that we have allocated the experimental treatment to the first patient, two possible outcomes can occur. If a success occurs, which happens with probability $1/2 = 0.5$, then the experimental treatment is allocated again. If a failure occurs, then the control treatment is allocated. Hence, $\Pr(a_{k,4}^{\text{GI}} = 1 | \tilde{\mathbf{x}}_2) = \sum_{\tilde{\mathbf{x}}_3 \in \Omega_3} \Pr(a_{k,4}^{\text{GI}} = 1 | \tilde{\mathbf{X}}_3 =$

Table 1

Comparison of the mean value of successes under the optimal rule for $K = 2, b = 2$ (obtained via dynamic programming), fixed randomized (FR), the FLGI, and the GI (Gittins index) in 10^3 replicas of a trial of size $T = 30$. The discount factor used for GI and FLGI is $d = 0.7$ (so as to get $(1 - d^T \approx 0)$ when $T = 30$.)

$p = [p_0 \ p_1]$	FR	Optimal ($b = 2$)	FLGI ($b = 2$)	GI ($b = 1$)	$p_{max} * T$
[0.1 0.1]	3.02	3.01	3.06	2.99	3
[0.2 0.9]	16.35	25.80	25.92	26.24	27
[0.1 0.3]	5.88	7.53	7.70	7.56	9
[0.35 0.65]	15.06	17.74	17.65	17.79	19.5
[0.4 0.5]	13.59	13.75	13.89	13.86	15
[0.7 0.8]	22.55	22.82	22.67	22.72	24

$\tilde{x}_3)Pr(\tilde{X}_3 = \tilde{x}_3|\tilde{x}_2)$ reduces to $1/2$ for $k = 0, 1$. Using equation (3), we can obtain $\pi_{0,2} = (0 + 1/2)/2 = 1/4, \pi_{1,2} = (1 + 1/2)/2 = 3/4$.

From this example, it is clear that the computational cost of computing the π_{kj} 's, which depends on the joint state for the $K + 1$ arms, i.e., \tilde{x}_t (instead of the one-arm state x_t), will grow exponentially as b and K increase. Hence, we use a Monte-Carlo algorithm for this purpose.

The optimal allocation for any K, b , and T solves a dynamic programming problem. To assess the FLGI's performance, we consider a two-arm trial of 30 patients recruited in 15 separate blocks ($T = 30, J = 15, b = 2, K = 2$). The optimal rule in this case allocates one patient per treatment in the first block and both patients to the treatment with the highest posterior mean in the last block. For blocks between 2 and 14, optimal actions are less straightforward to summarize. The FLGI for $b = 2$ constitutes a heuristic that approximates the optimal policy as shown in Table 1 (See Web Appendix C for the problem formulation).

4. Simulation Study

We now evaluate the properties of the FLGI procedure by simulation, focusing on its: statistical power $(1 - \beta)$; type I error rate (α) ; expected number of patients assigned to the best treatment (p^*) ; and expected number of patient successes (ENS). We compare the FLGI to the following established group allocation procedures (see Table 2):

- (1) Fixed randomization (FR): patients are allocated between the K experimental arms and the control arm with a fixed and equal probability at every interim stage.
- (2) Thompson sampling (TS): assuming a uniform prior on $[0, 1]^{K+1}$ for (p_0, \dots, p_k) , following Thompson (1933), we allocate treatment k to patients in block j with a probability proportional to the posterior probability that p_k is the largest response rate given the observed data.
- (3) Trippa et al. procedure (TP): Trippa et al. (2012) proposed a Bayesian adaptive design similar to Thompson sampling for testing multiple experimental treatments

in a controlled trial setting for patients with recurrent Glioblastoma. Although the original cancer setting determined the use of a time-to-event outcome, the design approach has been transferred to the binary data setting (Wason and Trippa, 2014).

- (4) Optimal allocation ratios: following the implementation in Tymofyeyev et al. (2007), we consider two different rules: *Neyman allocation* (NA), which minimizes total sample sizes for a given power constraint, and *RSIHR allocation* (RSIHR), which minimizes expected failures given a power constraint and is named after the authors that first introduced it in a two-armed setting (Rosenberger et al., 2001).
- (5) Controlled FLGI (CFLGI): in addition to the rule defined by (3), we shall consider a controlled group allocation rule which, similarly to the Trippa procedure, protects the allocation to the control treatment, so it never goes below $1/(K + 1)$ during the trial.

Note that in choosing which rules to compare our procedure against, we wanted to include methods that are used in practice, although fixed randomization is still the most popular allocation rule in use. In all simulations, we used a uniform Beta(1,1) prior for each treatment arm and we computed the allocation probabilities for Thompson sampling, Trippa procedure, FLGI, and CFLGI using a simple Monte-Carlo approximation based on 10^2 replicas (Matlab code is available with this article at the Biometrics website).

4.1. *The NeoSphere Trial*

The NeoSphere trial, as reported by Gianni et al. (2012), evaluated the effect of a combination of drugs on pathological complete response in the breast for women with locally advanced inflammatory breast cancer. Out of 417 eligible patients, 107 were randomly assigned to group 0 (trastuzumab plus docetaxel), 107 to group 1 (pertuzumab and trastuzumab plus docetaxel), 107 to group 2 (pertuzumab and trastuzumab), and 96 to group 3 (pertuzumab plus docetaxel). The original trial allocated patients between the four arms with a fixed and equal randomization probability of $1/4$. The primary response outcome was available 24 weeks after randomization. At the end of the trial, the response rates observed in groups 0, 1, 2 and 3 were 29.0%, 45.8%, 16.8%, and 24.0%, respectively.

We show the impact of redesigning the study using the FLGI and the alternative patient allocation rules. Two response rate scenarios are initially considered : (i) the global null where $p_i = 0.290, i = 0, 1, 2, 3$; (ii) the observed values in NeoSphere, $p_0 = 0.290, p_1 = 0.458, p_2 = 0.168, \text{ and } p_3 = 0.240$. We first fix the block size to $b = 9$ patients and the number of interim analysis to $J = 46$. This leaves three remaining patients which are allocated using the probability allocation vector resulting after observing the outcomes of block 46. If we imagine that the trial's duration remains fixed at 2 years, a block size of nine is consistent with being able to observe the outcome after a 2 week delay. We subsequently review the method's performance for larger block sizes which are consistent with longer delays.

To assess the chance of a type I error being made under the global null, we report the family-wise error rate $\bar{\alpha}$. This is

Table 2
Allocation rules: definitions

Allocation rule	Probability vectors
Fixed randomization (FR)	$\pi_{k,j}^{FR} = \frac{1}{K+1}$
Thompson Sampling (TS)	$\pi_{k,j}^{TS} = \frac{Pr(\max_i p_i = p_k \bar{x}_{(j-1)b})^c}{\sum_{k=0}^K Pr(\max_i p_i = p_k \bar{x}_{(j-1)b})^c}$ with $c = \frac{j*b}{2T}$
Trippa et al. Procedure (TP)	$\pi_{k,j+1}^{TP} = \frac{\bar{\pi}_{k,j+1}}{\sum_{k=0}^K \bar{\pi}_{k,j+1}}$ where $\bar{\pi}_{k,j+1} = \frac{Pr(p_k > p_0 \bar{x}_{(j-1)b})^{\gamma_j}}{\sum_{k=1}^K Pr(p_k > p_0 \bar{x}_{(j-1)b})^{\gamma_j}}$ $k = 1, \dots, K$ $\frac{1}{K} \exp[\max[(n_{k,j})_{k=1}^K] - n_{0,j}]^{\eta_j}$ $k = 0$
and as in Trippa et al. (2012)	$\gamma_j = 10(j * b/T)^{0.75}$ and $\eta_j = 0.25(j * b/T)$
Optimal allocation ratios	$\rho_k^* = n_k^* / \sum_{k=0}^K (n_k^*)$
where n_k^* solve:	$\min_{n_0, \dots, n_K} \sum_{k=0}^K w_k n_k$ s.t $n_k \geq B \sum_{k=0}^K n_k \quad \forall k$ $\phi(n_0, \dots, n_K) \geq C$
Neyman allocation (NA) takes	$w_k = 1 \quad \forall k$
RSIHR allocation (RSIHRA) takes	$w_k = (1 - p_k) \quad \forall k$
For both NA and RSIHRA, we considered and defined $\pi_{k,j}$ using an allocation function controlled by a tuning parameter γ and a smoothing parameter σ .	$B = 0.1$ and $C = 0.80$ $\gamma = 2, \sigma = 1$ See equation (2) in Tymofyeyev et al. (2007)
Controlled FLGI (CFLGI)	$\pi_{k,j}^{CFLGI} = \frac{\bar{\pi}_{k,j}^u}{\sum_{k=0}^K \bar{\pi}_{k,j}^u}$, for $k = 0, \dots, K$ where $\bar{\pi}_{0,j}^u = 1/K$ and for $k = 1, \dots, K$ with $\bar{\pi}_{k,j}^u$ are the approximate FLGI probabilities in (3).
Modified controlled FLGI (MCFLGI)	$\pi_{k,j}^{MCFLGI} = \frac{\bar{\pi}_{k,j}^u}{\sum_{k=0}^K \bar{\pi}_{k,j}^u}$, for $k = 0, \dots, K$ where $\bar{\pi}_{0,j}^u = \max\{1/K, \min\{\bar{\pi}_{0,j}^u, (1 - 1/K)\}\}$, with $\bar{\pi}_{k,j}^u$ are the approximate FLGI probabilities in (3)

the probability of rejecting at least one true null hypothesis. The Bonferroni method was used to account for multiple testing and ensure that $\bar{\alpha} \leq \alpha$, i.e., all hypothesis whose p-values p_k are less than $p_k \leq \frac{\alpha}{K}$ are rejected (with $\alpha = 0.05$). To assess power, we calculate the probability of rejecting the null for the truly best treatment under scenario (ii). That is the probability of rejecting $H_{01} : p_0 = p_1$ when $p_1 = 0.458$.

Hypothesis testing was performed using a normal cut-off value (when appropriate) and using an adjusted version of

Fisher’s exact test for comparing two binomial distributions. Fisher’s test is conservative, i.e., its actual rejection rate is far below the nominal significance level. To make the designs comparable, we chose its cut-off value so as to achieve the nominal type I error rate (Villar et al., 2015).

Under the global null, p^* is defined as the mean proportion of patients assigned to the control group, whereas under the alternative, p^* is defined as the mean proportion assigned to treatment 1. We also report the ENS value if all patients

Table 3

$5 * 10^3$ replicas of the NeoSphere trial under a FR design and alternative group designs with block size $b = 9$ and trial size $T = 417$. Top: Scenario (i). Middle: Scenario (ii) and Scenario (iii). Bottom: Scenario (iv).

$H_0 : p_0 = p_i = 0.29$ for $i = 1, \dots, 3$				
Scenario (i)	α		p^* (s.e.)	ENS (s.e.)
FR	0.043		0.250 (0.02)	120.88 (9.20)
TP	0.036		0.379 (0.05)	120.04 (9.24)
CFLGI	0.034		0.250 (0.02)	120.86 (9.22)
TS	0.054		0.250 (0.07)	120.86 (9.25)
RSIHR	0.024		0.248 (0.09)	120.80 (9.12)
NA	0.024		0.251 (0.09)	120.81 (9.30)
FLGI	0.046		0.251 (0.21)	120.96 (9.26)
GI	0.061		0.247 (0.23)	120.92 (9.49)
UB			1	120.93 (0.00)
$p_1 = [0.29 \ 0.458 \ 0.168 \ 0.24]$				
Scenario (ii)	$(1 - \beta)_{H_{02}}$		p^* (s.e.)	ENS (s.e.)
FR	0.642		0.250 (0.02)	120.62 (9.12)
TP	0.892		0.450 (0.04)	149.81 (10.9)
CFLGI	0.820		0.654 (0.06)	166.40 (11.9)
TS	0.782		0.585 (0.10)	155.93 (13.5)
RSIHR	0.480		0.479 (0.06)	139.51 (10.9)
NA	0.483		0.439 (0.05)	134.84 (10.0)
FLGI	0.178		0.847 (0.11)	179.64 (13.7)
GI	0.140		0.858 (0.12)	181.00 (14.1)
UB			1	190.99 (0.00)
$p_1 = [0.458 \ 0.29 \ 0.168 \ 0.24]$				
Scenario (iii)	$(1 - \beta)_{H_{01}}$	$(1 - \beta)_{H_{02}}$	p^* (s.e.)	ENS (s.e.)
FR	0.639	0.994	0.250 (0.02)	120.66 (9.17)
CFLGI	0.812	0.750	0.250 (0.02)	129.84 (11.2)
MCFLGI	0.878	0.728	0.382 (0.03)	140.69 (10.5)
RSIHR	0.493	0.960	0.480 (0.06)	139.47 (11.1)
NA	0.492	0.964	0.438 (0.06)	134.56 (10.3)
UB			1	190.99 (0.00)
$p_1 = [0.168 \ 0.458 \ 0.29 \ 0.24]$				
Scenario (iv)	$(1 - \beta)_{H_{01}}$		p^* (s.e.)	ENS (s.e.)
FR	0.994		0.250 (0.02)	120.36 (9.27)
CFLGI	0.987		0.634 (0.10)	152.85 (12.6)
RSIHR	0.961		0.480 (0.06)	139.65 (11.0)
NA	0.957		0.438 (0.06)	134.54 (10.3)
UB			1	190.99 (0.00)

received the best treatment (i.e., $0.458 * T$) as an upper bound on ENS of any design under scenario (ii).

4.2. Results

Table 3 displays the results from 5000 replications of the trial. As expected, under scenario (i) (top of Table 3) all the designs are equal in terms of ENS and p^* (ENS is $T * 0.29 = 120.93$). All rules allocate on average the same proportion of patients

to each treatment (close to 0.25). The exception to this is Trippa procedure for which control group allocation is always higher because it is matched to that of the best performing (by chance) treatment. Thompson sampling and the Gittins index-based designs have a variability in the allocation probabilities between 4–8 times larger than the other designs.

The most interesting differences among these designs occur under scenario (ii). The value of ENS attained under fixed

randomization in this case is close to the value reported in the study (which was of 121 complete remissions). We find that the Gittins index, FLGI, and CFLGI procedures increase the number of successes dramatically compared to fixed randomization (achieving 60, 58, and 45 more remissions on average, respectively). Trippa procedure also improves on fixed randomization (by approximately 29 remissions on average) and it achieves a higher power than the CFLGI, yet remissions attained are 10% lower.

Of particular interest is the *learn-versus-earn* trade-off. The first four designs (fixed randomization, Trippa procedure, CFLGI, and Thompson sampling) achieve more than 60% power to reject H_{01} , whereas the last four designs (RSIHRA, Neyman allocation, FLGI, and the Gittins index) fall below this power level. However, the designs that achieve the highest patient benefit (i.e., $p^* \geq 0.80$ and $ENS \geq 175$) are the ones that have the lowest power. The designs that have higher ENS and p^* also have a larger variability of allocation probabilities.

The optimal allocation ratio designs (Neyman allocation and RSIHRA) are constrained to attain at least a given probability to reject the global null for a test of homogeneity H_0 (i.e., to reject at least one H_{0k} with a two-sided test). In scenario (ii), they are required to attain the same level as fixed randomization to do that (i.e., $C = 0.80$ in Table 2). In fact, both designs produce levels substantially higher than that (0.99 and 0.97, respectively). To attain this, both optimal designs, after a minimal requirement of experimentation on all arms (i.e., $B = 0.1 \times b$), skew allocation toward the best (i.e., $k = 1$) and worst arm (i.e., $k = 2$) to reject H_{12} . Because of this, their probability of rejecting H_{02} is lower than that of fixed randomization, as the increased probability to reject the global null H_0 is gained by increasing the chance of rejecting H_{12} instead of H_{02} .

For multi-arm trials, power depends on the particular alternative hypothesis considered. In Neosphere, the worst and best treatments were different from the control arm and this explains the reduced marginal power of Neyman allocation and RSIHRA to reject H_{02} . It also explains the high marginal power attained by CFLGI, as its allocation quickly tends to assigning one in every four patients to control and the remaining three to the best experimental arm, which in Neosphere was $k = 2$. To illustrate this point, we consider two more cases where we compare the CFLGI to the optimal designs that target an ethical criterion and power simultaneously: Scenario (iii), where we let the control arm be the best treatment observed in Neosphere, and Scenario (iv) where control was assumed to be the worst treatment in Neosphere.

These results suggest that the power-ENS balance of the CFLGI is maintained except for the case in which the best arm is the control arm. In general, we expect the CFLGI to have a high marginal power to detect a difference between the best experimental treatment, if it exists, and control. In Web Appendix D, we discuss these extra scenarios and present a modification of the CFLGI rule that also performs well in the case where control is the best treatment.

4.3. Varying Block Size

We next evaluate the effect of changing the block size (or the number of interim analyses) on the ENS-power relation of the allocation methods in which power is not constrained.

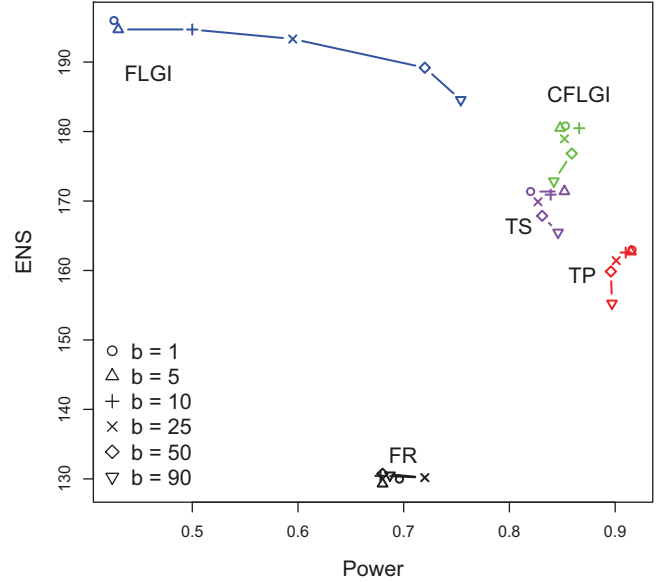


Figure 2. The expected number of successes (ENS) versus power for different designs: Thompson Sampling (TS), Trippa et al. (TP), Fixed Randomized (FR), FLGI, and CFLGI and block sizes b varying in $\{1, 5, 10, 25, 50, 90\}$ in 10^4 replicas.

By definition when $b = 1$ the FLGI is identical to the fully sequential Gittins index rule (i.e., it favors ENS at the expense of power). Also, from Proposition 1 (see Web Appendix B) we know that for $b = T$ (and $J = 1$) and all arms with identical initial priors, the FLGI group design is simply a fixed randomized design (i.e., it favors power at the expense of ENS). Hence, by varying b between 1 and T one can see how the FLGI approach trades off power for ENS within this range. This is illustrated in Figure 2. Note that the trial size is increased from 419 to 450 to allow several values of b to divide it exactly.

For the FLGI design, as the size of the block increases, its ENS decreases and its power increases. Conversely, for the other designs (CFLGI and Trippa procedure), the balance between power and ENS remains approximately constant when changing the block size. Block sizes with at least 80% power have ENS values below 180, while block sizes with ENS values above 180 have below 80% power. The designs that protect the allocation to control have a very similar performance in their power-ENS trade-off across different block sizes. For every block size, CFLGI achieves a higher ENS but a lower power than Trippa procedure.

4.4. The FLGI and the Patient Population Size

Cheng and Berry (2007) argue that the optimality of a particular design should be judged by taking into account the size of the trial (T), the total patient population (N , or the “patient horizon”) and, crucially, the ability of the trial’s allocation rule to facilitate identification of the truly best treatment at its conclusion while treating patients in the trial as effectively as possible. We now assess, via simulation, the designs considered from this conceptual viewpoint by considering the total expected number of successes (ENS_N) as the sum of the first T patients in the trial (ENS_T) and the remaining $N - T$ outside

of the trial (ENS_{N-T}):

$$ENS_N = ENS_T + ENS_{N-T}. \tag{4}$$

We explore two rules to choose the ‘best’ treatment (denoted by k^*) at the end of the trial. Rule 1 is as follows: if there is no treatment with a statistically significant effect compared to the control group, then the control group is chosen ($k^* = 0$). However, if one or more treatment is significantly effective, then the treatment with the largest posterior mean by the end of the trial is chosen. ENS_N in equation(4) can then be evaluated for each allocation rule by calculating ENS_T as described earlier in Section 4 and using

$$ENS_{N-T} = (N - T) \sum_{k=0}^K Pr(k^* = k) p_k, \tag{5}$$

where $Pr(k^* = k)$ is the probability treatment and k is selected at the end of the trial. Rule 2 simply chooses the treatment with the highest Gittins index by the end of the trial so that

$$ENS_{N-T} = (N - T) \sum_{k=0}^K Pr \left(\mathcal{G}(s_{k,T}, f_{k,T}) = \max_j \mathcal{G}(s_{j,T}, f_{j,T}) \right) p_k \tag{6}$$

The probabilities of selecting a treatment as the best treatment under both rules are computed via simulation after 10^3 replicas of the trial of size T . Let $K=3$ and the true response rate vector be the observed values in NeoSphere, i.e., $(p_0, p_1, p_2, p_3) = (0.29, 0.458, 0.168, 0.24)$. We consider two possible trial sizes: a small trial ($T = 80$) to represent a rare disease and a large trial (the original size of $T=417$) to represent a more common condition. Furthermore, we vary the population size as $N = (1 + l)T$ for $l \in \{1, 2, 3, \dots, 10, 20, 40, 60, 80, 100\}$. This allows, in the rare disease case, for the total population to range from 160 (equal numbers inside and outside of the trial) to 8080 (a hundred times as many outside as inside). We assess the relative gap between ENS_N using its upper bound (if all patients had been assigned to the best treatment from the start) and the value calculated from equation (4) (for a given trial allocation and treatment selection rule) using

$$\text{Gap} = \frac{Np_1 - ENS_N}{Np_1}. \tag{7}$$

Figure 3 (top) shows the relative gap (7) as a function of N under rule 1 for the small trial (top-left) and large trial (top-right). These results indicate that all rules, except for the Gittins index and FLGI, exhibit a type of asymptotic convergence similar to the one reported in Cheng and Berry (2007). That is, as the population that benefits from the trial grows (i.e., as $N - T \rightarrow \infty$), their suboptimality gap tends to zero. However, for FLGI and the Gittins index approach the opposite is true. For smaller population sizes the gap is smallest, and it grows as a function of N . Moreover, the gap

of the Gittins index and the FLGI rules is always above the other designs.

The poor performance of the uncontrolled Gittins index-based rules is caused by the use of a classical significance test. Such tests are appropriate when large enough numbers of patients are assigned to all arms and are thus perfectly suited to rules such as fixed randomization. If a truly best treatment exists, the Gittins index-based rules will assign highly unequal numbers across trial arms. Although this imbalance is a positive indicator of a superior treatment effect, it (paradoxically) leads to a low power. For this reason, Figure 3 (bottom) shows how the relative gap of the Gittins index, FLGI, and CFLGI decreases toward 0 under selection rule 2 for both trial sizes. The results for the other allocation methods using selection rule 1 are shown for comparison. Under rule 2, the convergence of the gap toward 0 as T and $N - T$ tend to infinity holds and moreover, the advantage of these rules over the other designs holds for all N .

These results illustrate that, for rare diseases (where there is little power to detect a statistically significant treatment effect) bandit-based designs are highly attractive. Furthermore, the value of such designs lies within the fact that they can be effectively used for treatment selection (i.e., using rule 2). Note that CFLGI performs well under either selection rule.

5. Discussion

Despite its optimality, and its original clinical motivation, the Gittins index rule has never been used in practice to conduct an actual clinical trial. In this article, we focus on three of the practical barriers to its use, namely: (1) insufficient statistical power (when computed using traditional hypothesis testing procedures); (2) the need for instantly observed treatment outcomes; and; (3) a lack of randomization to provide a basis for inference. To address these barriers, we propose the forward looking Gittins index algorithm, a randomized group sequential allocation procedure based on the index solution to the *classic* infinite horizon multi-arm bandit problem.

Simulations confirm that the FLGI procedure, combined with a protected control group allocation, enables adaptively block-randomized clinical trials that are as follows: statistically conservative when no treatment with additional benefit over control exists; statistically powerful when one does; and highly ethical in terms of patient benefit. These characteristics were found to persist when the block size was stochastic, instead of being fixed (results not shown). As with all adaptive allocation procedures, the feasibility of our approach is diminished when there are significant delays in observing patient responses.

The FLGI rule, although independently developed, resembles the policy improvement algorithm (see e.g., Howard, 1960) used to approximate computational expensive optimal value functions in Markov Decision problems. Semi-randomized rules based on the Gittins index have been proposed by Glazebrook (1980). These rules overcome the power limitation by continuing to allocate suboptimal arm with a positive, though decreasing, probability. However, they are not expressed in terms of allocation probabilities, and thus do not constitute a fully randomized procedure in the sense that we have outlined here. Gittins indices and analogous optimal-

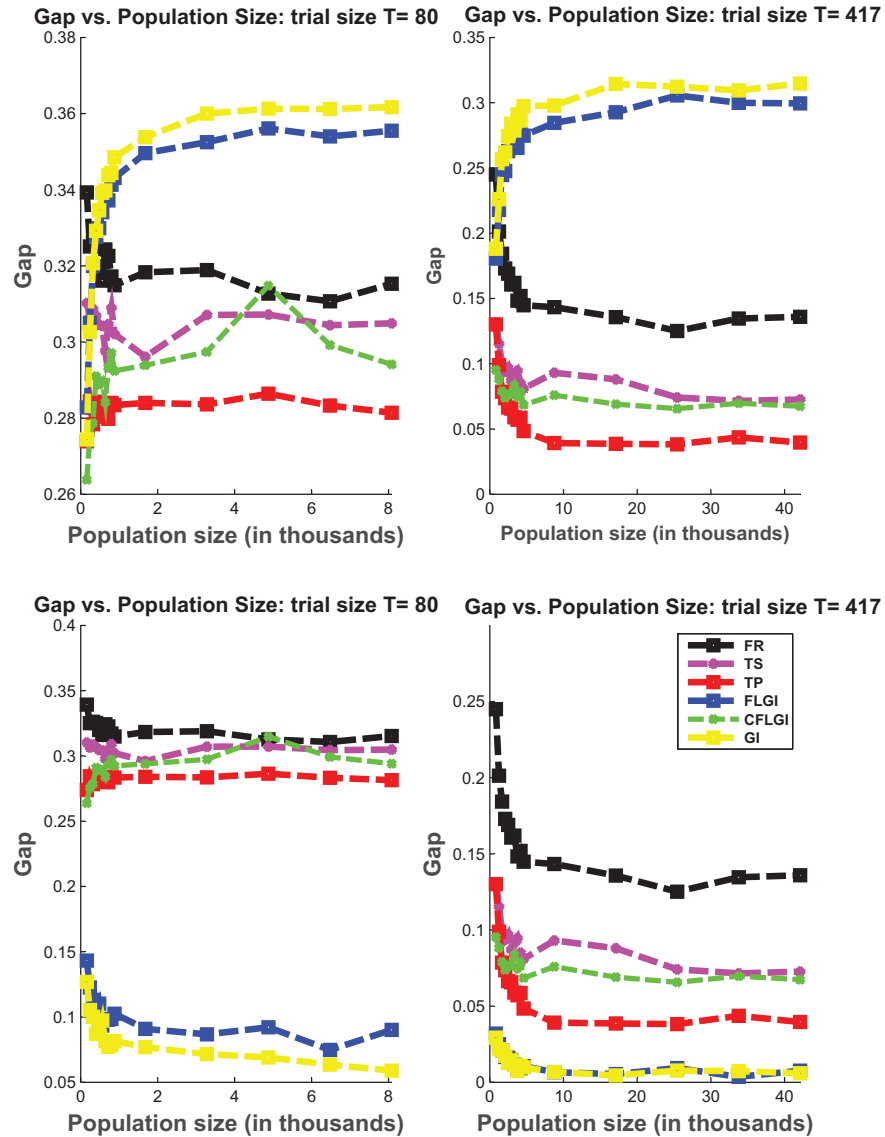


Figure 3. The suboptimality gap for $T = 80$ and $T = 417$ as the population size varies in $N = \{(2, 3, 4, \dots, 11)T, (21, 41, 61, 81, 101)T\}$ in 10^3 replicas. Top: simulations using power computations based on statistical significance. Bottom: simulations using the GI rule to make a final decision for FLGI, CFLGI, and GI.

ity results have been derived for a variety of endpoints. Therefore, a natural extension of this work can consider other endpoints, e.g., exponentially distributed populations (see Gittins et al., 2011). Also, extending the FLGI rule for the case in which covariates are available, as in Yang and Zhu (2002), is very relevant.

Finally, the FLGI rule is especially useful for trials in which the main goal is the selection of a superior treatment for continuing further study (for example in Phase II or rare disease settings). This is because the FLGI (just as the Gittins Index) skews the allocation toward one of the treatments, selecting the best treatment within the trial when it exists, although it is often then unable to declare it effective at the desired significance level. This highlights the need to develop hypothesis testing procedures which complement bandit designs. High power may be less relevant than the goal of having

a large probability of selecting a superior within the trial and therefore treating patients better. If power is a concern, the controlled FLGI rule can be applied to gain both power and ethical advantages.

Adaptive designs are often criticized in terms of their frequentist operating characteristics. We have found that the FLGI approach controls the type one error conservatively and that the suboptimal arms (those discarded within the trial by the FLGI rule) also have their success rates slightly underestimated (Villar et al., 2015). Further research is needed to develop testing and estimation procedures that maintain the correct frequentist operating characteristics.

The motivation of our proposed algorithm was in the setting of clinical trials, but its use is not limited to medical applications. The heuristic applies to sequential allocation problems more generally. In particular, the FLGI can be used as a

heuristic to solve bandit problems with more than two actions (See Web Appendix C and E) or with random block sizes.

6. Supplementary Web Appendix

Web Appendices A–D and additional Tables referenced in Sections 2, 3, 3.1, 4.2, 4.3, and 5 are available with this article at the Biometrics website on Wiley Online Library. Example code and instructions for running it are also available as a web supplement (same website).

ACKNOWLEDGEMENTS

We thank the editor, an associate editor, and two referees for their insightful comments and suggestions. The first author thanks Professor Donald Berry for some helpful discussions on this article and his constructive suggestions. This work was funded by the UK Medical Research Council (Grant numbers G0800860, MR/J004979/1, and MR/L012286/1), Cambridge Biomedical Research Centre, and the Biometrika Trust.

REFERENCES

- Aalto, S., Ayesta, U. and Richter, R. (2009). On the Gittins index in the M/G/1 queue. *Queueing Systems* **63**, 437–458.
- Beauchamp, T. (2008). The Belmont report. Oxford University Press, USA. *The Oxford Textbook of Clinical Research Ethics* 149–55.
- Berry, D. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. London: Chapman & Hall.
- Berry, D. and Eick, S. (1995). Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine* **14**, 231–246.
- Cecchi, F. and Jacko, P. (2013). Scheduling of users with markovian time-varying transmission rates. In *Proceedings of the ACM SIGMETRICS/International Conference on Measurement and Modeling of Computer Systems*, 129–140. New York, NY: ACM.
- Cheng, Y. and Berry, D. (2007). Optimal adaptive randomized designs for clinical trials. *Biometrika* **94**, 673–689.
- Chow, S. C. and Chang, M. (2008) Adaptive design methods in clinical trials a review. *Orphanet Journal of Rare Diseases* **3**, 169–190.
- Gianni, L., Pienkowski, T., Im, Y., Roman, L., Tseng, L., Liu, M. et al. (2012). Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early her2-positive breast cancer (neosphere): a randomised multicentre, open-label, phase 2 trial. *The Lancet Oncology* **13**, 25–32.
- Gittins, J. and Jones, D. (1979). A dynamic allocation index for the discounted multiarmed bandit problem *Biometrika* **66**, 561–565.
- Gittins, J., Glazebrook, K., and Weber, R. (2011). *Multi-armed bandit allocation indices*. New York, NY: Wiley.
- Glazebrook, K. (1980). On randomized dynamic allocation indices for the sequential design of experiments. *Journal of the Royal Statistical Society, Series B (Methodological)* **42**, 342–346.
- Hauser, J., Guilherme Liberali, G., and Braun, M. (2009). Website morphing. *Marketing Science* **28**, 202–223.
- Howard, R. (1960). *Dynamic Programming and Markov Processes*. New York: J Wiley.
- Hu, F. and Rosenberger, W. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* **98**, 671–678.
- Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York: John Wiley & Sons.
- Press, W. (2009). Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research. *Proceedings of the National Academy of Sciences*, **106**, 22387–22392.
- Rosenberger, W. and Lachin, J. (1993). The use of response-adaptive designs in clinical trials. *Controlled Clinical Trials* **14**, 471–484.
- Rosenberger, W., Stallard, N., Ivanova, A., Harper, C., and Ricks, M. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57**, 0909–913.
- Thompson, W. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25**, 285–294.
- Trippa, L., Lee, E., Wen, P., Batchelor, T., Cloughesy, T., Parmigiani, G. et al. (2012) Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *Journal of Clinical Oncology* **30**, 3258–3263.
- Tymofeyev, Y., Rosenberger, W. F., and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association* **102**, 224–234.
- Villar, S., Bowden, J., and Wason, J. (2015). Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges. *To appear in Statistical Science*.
- Wang, L. and Arnold, K. (2002). Press release cancer specialists in disagreement about purpose of clinical trials. *Journal of the National Cancer Institute* **94**, 1819–1819.
- Wason, J. and Trippa, L. (2014). A comparison of bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Trials* **14**, 40. ISSN 1745-6215. URL <http://www.trialsjournal.com/content/14/S1/P40>.
- Yang, Y. and Zhu, D. (2002). Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *The Annals of Statistics* **30**, 100–121.

Received September 2014. Revised March 2015.

Accepted April 2015.