Applied Stochastic
Models in Business
and Industry

Special Issue Paper

Received 9 September 2016, Revised 20 February 2017, Accepted 20 February 2017 Published online 22 March 2017 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2249

Combining Bayesian experimental designs and frequentist data analyses: motivations and examples

Steffen Ventz[a,b], Giovanni Parmigiani[b,c*†] and Lorenzo Trippa[b,c]

Recent developments in experimental designs for clinical trials are stimulated by advances in personalized medicine. Clinical trials today seek to answer several research questions for multiple patient subgroups. Bayesian designs, which enable the use of sound utilities and prior information, can be tailored to these settings. On the other hand, frequentist concepts of data analysis remain pivotal. For example, type I/II error rates are the accepted standards for reporting trial results and are required by regulatory agencies. Bayesian designs are often perceived as incompatible with these established concepts, which hinder widespread clinical applications. We discuss a pragmatic framework for combining Bayesian experimental designs with frequentists analyses. The approach seeks to facilitate a more widespread application of Bayesian experimental designs in clinical trials. We discuss several applications of this framework in different clinical settings, including bridging trials and multi-arm trials in infectious diseases and glioblastoma. We also outline computational algorithms for implementing the proposed approach. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: Bayesian experimental design; clinical trials; decision theory

1. Introduction

The Bayesian design of a clinical trial is characterized by the collection and subsequent formalization of available information through a prior distribution. Previous clinical trials, data from epidemiological studies, or disease models are standard examples of relevant information used to specify the prior. In summary, the design of the trial starts from a prior distribution π over a set of unknown parameters

$$\theta \sim \pi.$$

Throughout our discussion π will be a genuine representation of the investigators beliefs and uncertainties on key parameters θ. Typically in medicine, θ includes response probabilities, survival curves, or toxicity rates of different treatments. These parameters will be estimated and compared using the data generated by the clinical trial.

The information embedded in the prior π can be used in several contexts and for different purposes. Examples are (i) the choice of the sample size for a two-arm or a multi-arm study [1], (ii) the definition of a two stage design, with stage-specific samples sizes selected using the prior π [2], and (iii) Bayesian adaptive randomization, with reinforcement of the randomization probabilities during the trial towards the most promising arms [3–5].

Some of these designs, for example, two-arm studies, can be optimized by a direct application of the decision theoretic paradigm. The design is selected by the prior π and the utility function u, which is representative of the investigators' preferences. In general, the solution of the decision problem coincides with the design d that maximizes the expected value

$$\mathbb{E}_{\pi,d}(u)$$

of the utility generated by the experiment [6]. Here the utility $u = u(Y, d, \theta)$ is a random quantity, which is a function of the unknown parameter θ and the data Y collected during the study with design d. In clinical trials, the utility function

[a] *Department of Computer Science and Statistics, University of Rhode Island, Kingston, RI, U.S.A.*
[b] *Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, U.S.A.*
[c] *Department of Biostatistics, Harvard School of Public Health, Boston, MA, U.S.A.*
**Correspondence to: Giovanni Parmigiani, Dana Farber Cancer Institute, Boston, MA, U.S.A.*
[†] *E-mail: gp@jimmy.harvard.edu*

typically represents the preferences of multiple stakeholders, including medical investigators, pharmaceutical companies, and patients. We discuss some examples in the next paragraphs.

The sample size for a two-arm study d can be selected by specifying a utility function that captures the trade-off between marginal costs associated with the enrollment of each patient and the likelihood to correctly identify and recommend the best available treatment. In this example,

$$u(X,\theta,d) = \texttt{data support recommendation of the best treatment} \\ - \texttt{constant} \times \texttt{sample size.}$$

In other contexts, the selected design d is not the solution of a maximization problem. The use of a prior distribution π is combined with less explicit utility criteria. Examples include the use of adaptive randomization probabilities in multi-arm trials, with randomization probabilities proportional to the posterior probabilities of positive treatment effects [3, 7]. In this case, the utility criteria are not explicitly stated, but the intention is explicitly to increase the accrual towards the most promising arms. This type of studies uses the prior π and the data generated during the trial for interpretable decisions, such as variations of the randomization probabilities, or to drop arms during the trial. We refer to Berry and Fristedt [8] for discussions of the decision theoretic framework to define adaptive randomization probabilities, which illustrate computational complexities and justify the use of alternative heuristic algorithms.

For trials designed without the explicit use of a utility functions u and a prior distributions π, sample sizes and interim decision rules are often selected using substitutes of (π, u), such as tables, which report operating characteristics under a list of simulation scenarios. These evaluations typically involve several candidate designs. The list of scenarios, similarly to π, is representative of prior beliefs and predictions of the investigator. Symmetrically, the choice of the operating characteristics to be compared across potential designs mirrors the investigators preferences. We often have a one-to-one correspondence between the key components of the decision theoretic framework (π, u) and those of a simulation study, scenarios, and operating characteristics [9].

We list a few closely related advantages for using prior distributions and utility functions: first, a pragmatic aspect. The selection of a design based on examining tables and summaries across simulation scenarios, candidate designs, and competing operating characteristics can be quite challenging and time consuming. Second, the use of the decision theoretic approach forces investigators to think through and explicitly state goals and assumptions via a prior π and a utility function. In routine tasks, for example, selection of futility stopping boundaries, it is easier to interpret and subsequently agree or disagree on the choice of π and u, than having a debate over large tables of operating characteristics. Additionally, a clinical trial design selected based on decision theoretic arguments can always – and in most cases should – be scrutinized through interpretable summaries of the resulting operating characteristics. Still, skepticism can be an appropriate reaction towards attempts to declare exhaustive the evaluation of a design through simulations and tables of operating characteristics. These tables can be necessary but not sufficient for a solid evaluation the trial design. Third, in complex trials, it is difficult to replace prediction and posterior probabilities with alternative data summaries with comparable level of interpretability. In particular, prediction and posterior probabilities are useful and interpretable to specify interim analyses during the study. For instance, in studies with biomarker-treatment interactions, posterior probabilities can be used to modify arm-specific eligibility criteria [10, 11].

Limitations of the Bayesian framework that prevent a more widespread use in clinical trials, including computational demand, prior elicitation, and the acceptance of a single utility function from several stakeholders, have been discussed in the literature [12, 13]. The goal of the sections that follow is complementary to these discussions of the pros and cons of the Bayesian framework. We discuss a possible strategy to facilitate the use of Bayesian foundations in clinical trials. Most clinical investigators and scientific review panels are not against the use of Bayesian designs. However, there are barriers to a rapid diffusion of Bayesian methods in clinical trial designs. Here we only focus on one of them, perhaps an important one, by illustrating that the results reported at completion of a Bayesian trial do not necessary need to be linked and influenced by the choice of the prior π.

Clinicians, scientific review panels, and other stakeholders in the clinical trials arena, in most cases, are familiar with key statistical concepts from the frequentist literature; type I error rates, hypothesis testing, and confidence intervals to name a few. These concepts are accepted standards for reporting results in clinical trials and to communicate evidence of positive effects or futility of novel treatments. Bayesian designs are often perceived as incompatible with these established metrics for reporting results, in particular p-values and hypotheses testing. This is the perceived barrier that we will discuss. We illustrate the use of methods to combine Bayesian models π, utility functions u, and frequentist analyses, including the control of type I errors rates and confidence intervals.

We include frequentist constraints into a Bayesian decision theoretic framework [2, 14]. These constraints reflect desiderata from collaborators and other stakeholders. Examples include the control of type I error below 0.05 or minimal bias in the effect estimates. The first panel of Figure 1 illustrates graphically the application of the decision theoretic framework. The action space D, that is, the set of all trial designs, is shown on the right. A point d in D is a candidate trial design, and typically, it includes sample size, stopping rules, and also a plan on how to analyze the data and communicate the final results of the trial. Estimators and procedures to report evidence of treatment effects or futility are components of the trial design d. Importantly, the plan for final analyses can vary substantially across candidate designs in D. The space D is mapped to the range of expected utilities $U(D)$. The Bayesian statistician selects the trial design $d_{\max}$ that maximizes the expected utility $U(d)$. In the first panel of Figure 1, $u_{\max}$ is the maximum of the expected utility surface, which is achieved by the design $d_{\max}$.

We can now describe the strategy of our Bayesian biostatistician to select a trial design, by including its interactions with the scientific community, clinicians, editors of scientific journals, and review committees. We model these interactions by adding constraints to the operating characteristics of the trial (Panel B of Figure 1). Examples, as we mentioned, include the requirement to bound type I/II error rates below explicit thresholds or to limit the expected enrollment below a prespecified threshold under the hypothesis of a detrimental or toxic treatment. These are well defined frequentist constraints, and a candidate design d can satisfy the requirements or not, irrespective of the prior distribution π. In Figure 1, we indicate these constraints through the set V. The subset of designs that satisfies them $OC^{-1}(V) \subset D$ is identified by the map OC, which links designs d to their operating characteristics $OC(d)$. The choice is now constrained to the selection of a possibly suboptimal design within $OC^{-1}(V)$. Our Bayesian decision maker selects the design $d_{CO-\max}$ that optimizes the expected utility surface within the subset $OC^{-1}(V)$. We discuss algorithms to approximately select the constrained optimum $d_{CO-\max}$ in Sections 2.1 and 3.

We list a few properties of our constrained decision theoretic (CDT) framework:

- It includes an explicit unambiguous utility function u as the primary criterion to select a trial design.
- It makes effective use of prior estimates and scientific knowledge.
- It allows dissemination of the major findings using an established scientific language, including frequentist concepts such as hypothesis testing and power.
- It facilitates communication of the design characteristics with multiple stakeholders and non-statisticians.
- It is straightforward to extend the approach to prediction-based adaptive strategies and algorithms that remain similar in spirit and share a similar interpretation.

In the sections that follow, we first provide an example of a direct application the CDT framework.

We will then move to examples of more complex sequential designs where, similarly to the standard decision-theoretic framework, exact solutions become computationally unfeasible, and it is necessary to replace the optimization strategy with heuristic algorithms.

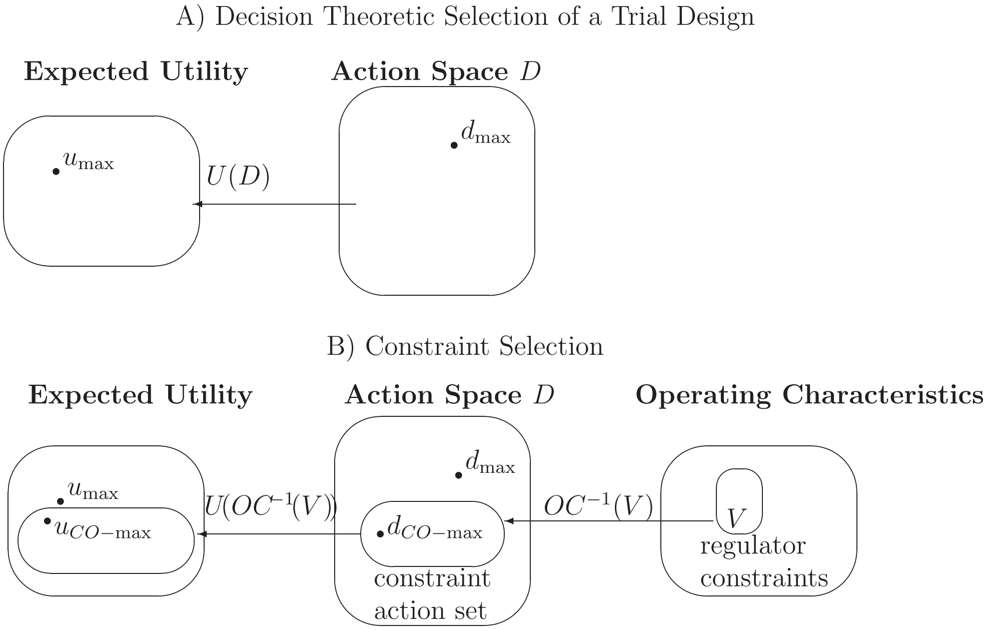

Figure 1. Graphical representation of the optimal Bayesian design $d_{\max}$ and of the constrained optimal Bayesian design $d_{CO-\max}$. In this diagram, V, $OC^{-1}(V)$, and $U(OC^{-1}(V))$ denote the regulator constraints, the subset of designs with operating characteristics in V, and the corresponding expected utilities. The expected utilities of $d_{\max}$ and $d_{CO-\max}$ are $u_{\max}$ and $u_{CO-\max}$, respectively.

2. Constrained optimal designs

2.1. Constrained optimal bridging trials

In Ventz and Trippa [2], we previously explored the use of the CDT framework for the design of bridging trials [15]. Here we provide a summary of the results obtained by applying the CDT framework. A bridging trial assesses whether a drug recently approved in a region A, say Europe, can be marketed in a different region B, for instance Japan. Clinical data from region A should guarantee that the drug is effective and safe, and the bridging trial is a supplementary study to test whether the drug has a similar treatment effect and safety profile in population B [15]. In this setting, we have historical data from randomized trials and information, which the investigator can incorporate in the prior π. The use of the CDT framework requires the specification of three components π, u, and V. One can argue that the available data allow straightforward specification of the first component π. Additionally, the investigator can specify sound utility functions based on estimates of relevant utility parameters, such as the potential number of prescriptions per year in region B. Regulators and other stakeholders, such as patient representatives, can express the need for controlling type I error rates and/or other characteristics of the study from a frequentist viewpoint. We indicate the constraints by V as before.

For each patient i, the primary endpoint Y_i, say the reduction of blood pressure, conditional on the treatment $C_i = 1$, or the placebo $C_i = 0$, is assumed normal distributed with mean θ_k, $k = 0, 1$. The trial has to test $H_0 : \gamma_B = 0$ versus $H_1 : \gamma_B > 0$, where $\gamma_B = \theta_0 - \theta_1$. A group-sequential trial with a possible early termination at interim analyses $t = 1, \cdots, T-1$ in favor of H_1 is used, and N patients will be randomized to each arm between consecutive interim analyses. We use the summary $Z_t = (\hat{\theta}_{0,t} - \hat{\theta}_{1,t})/\sqrt{(2\sigma^2)/(Nt)}$ and, without loss of generality, assume a common variance σ^2 for the two arms. Here $\hat{\theta}$ denotes maximum likelihood estimates. The vector $Z_{1:T} = (Z_1, \dots, Z_T)$ is Gaussian with mean $\mu = (\gamma_B\sqrt{Nt/2\sigma^2})_{t \leqslant T}$ and covariance matrix $W = (W_{t,t'})_{1 \leqslant t,t' \leqslant T}$, where $W_{t,t'} = \sqrt{t/t'}$ for $t \leqslant t'$. A design is characterized by the parameter N and stopping boundaries $z_{1:T} = (z_1, \cdots, z_T)$ at interim and final analyses $t = 1, \cdots, T$.

Low power could delay patients' access to an effective drug. We therefore assume that the regulator requires types I and II error rates, at H_0 and $\gamma_B = \gamma_B^* > 0$, to be controlled at suitably chosen α and β levels, respectively. The information from region A can be summarized by a Gaussian prior for the parameter γ_B. Power priors [16], for example, are directly applicable to specify π on the basis of information from region A. We use an interpretable utility function, with costs linear in the number of randomized patients ($h > 0$) and, in case of a true positive finding, a fixed payoff at termination of the trial

$$\begin{aligned} u(Y,\theta,d) = \sum_{t=1}^{T} I(Z_t \geqslant z_t, Z_{1:(t-1)} \leqslant z_{1:(t-1)}, \gamma_B > 0) \\ -hN\left(1 + \sum_{t=1}^{T-1} I(Z_{1:t} \leqslant z_{1:t})\right). \end{aligned} \tag{1}$$

We solved the constrained optimization problem and computed optimal thresholds $z_{1:T}$ for the summary statistics $Z_{1:T}$ to allow early termination of the study with several variations on the prior π and utility function u. The solution showed negligible departures from linear thresholds. As expected, by varying utility parameters and prior distribution, we obtained considerably different thresholds.

To compute the optimal stopping rules $z_{1:T}$, we leverage on monotonicity properties [2]. First, the expectation of the utility function (1) can be written as the difference between two monotone functions $U(d) = U_1(z_{1:T}) - U_2(z_{1:T})$. Here $U_1(z_{1:T})$ is the expected payoff for discovering an effective treatment during the trial, while $U_2(z_{1:T})$ represents the expected cost of the trial, which is proportional to the expected sample size. Second, assuming that N is fixed, the operating characteristic $OC(z_{1:T}) = \mathbb{P}_{\gamma_k=0}[\cup_t\{Z_t \geqslant z_t\}]$ is a monotone function of $z_{1:T}$. The algorithm used to compute the optimal design partitions the space of designs and computes lower and upper expected utility bounds for each partition set. Figure 2 is a graphical representation of the optimization algorithm. In this case, we have one interim analysis and it is therefore necessary to compute $T = 2$ thresholds. The two panels show the status of the algorithm at two consecutive iterations. The current status consists of a list of rectangles $\{r^i\}$, where $r^i = [z^i_{L,1}, z^i_{U,1}) \times [z^i_{L,2}, z^i_{U,2})$ that could potentially harbor the constrained optimum. For each value of $z_{1:2}$ in the rectangle r^i the expected utility can be bounded from below by $LB(r^i)$ and from above by $UB(r^i)$. Here LB and UB correspond to the difference between the expected payoff and costs computed at different combinations of the lower and upper boundary points $(r^i_{\ell,1}, r^i_{\ell,2})$, $\ell = L, U$, of the rectangle r^i. The boundaries LB and UB are defined by exploiting the monotonicity of the cost and payoff component of the utility function. Similarly, the operating characteristic $OC(d(z_{1:2}))$ is bounded by exploiting monotonicity.

At each iteration, a single rectangle r^i is either (i) removed from the list because it does not contain $d_{OC-\max}$ or (ii) r^i is divided in two sub-rectangles. A rectangle r^i is removed from the list either because the operating characteristics of all the designs in r^i do not satisfy the constraint V or because the upper bound $UB(r^i)$ is dominated by the lower bound $LB(r^j)$

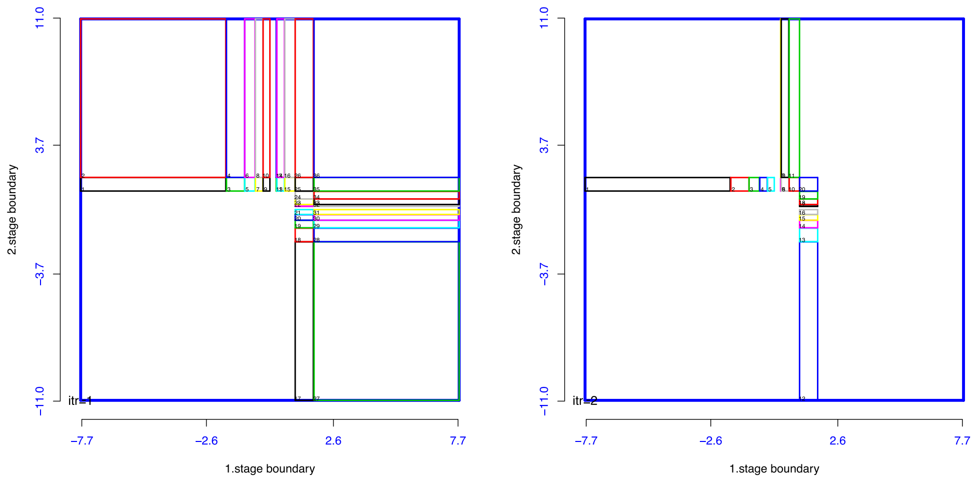

Figure 2. Cut-and-zoom-in algorithm for computing the constrained optimal design $d_{CO-\max}$. The left and right panels show consecutive iterations of the algorithm. At each iteration, the algorithm (i) removes rectangles and (ii) splits rectangles into disjoined components. [Colour figure can be viewed at wileyonlinelibrary.com]

of a different rectangle $j \neq i$. Otherwise, we split the rectangle r^i into two disjoined rectangles r^{i_1} and r^{i_2}, with $r^i = r^{i_1} \cup r^{i_2}$. The two rectangles are defined by selecting one of the two dimensions at random, say the second dimension, and selecting one point $z^i_{C,2}$ between $z^i_{L,2}$ and $z^i_{U,2}$. Then $r^{i_1} = [z^i_{L,1}, z^i_{U,1}) \times [z^i_{L,2}, z^i_{C,2})$ and $r^{i_2} = [z^i_{L,1}, z^i_{U,1}) \times [z^i_{C,2}, z^i_{U,2})$. In our optimization procedure, we used $2\Phi(z^i_{C,2}) = \Phi(z^i_{L,2}) + \Phi(z^i_{U,2})$, where $\Phi(\cdot)$ is the standard normal distribution function.

By computations of the expected utilities and operating characteristics at the extremes of the rectangles and exploiting monotonicity, the algorithm progressively and iteratively removes candidate designs d and zooms into regions of the action space with comparable operating characteristics that include the constrained optimum.

2.2. A multi-arm response-adaptive design in glioblastoma

Glioblastoma is a brain cancer associated with a poor prognosis. Numerous treatments, in recent years, have shown promise in preclinical models, but translation into tangible treatment effects and survival improvement has been slow and nearly negligible [17]. Current trial designs and more generally pipelines for developing new treatments have been severely criticized for being inefficient [18]. Most of the current early-phase trials for patients with glioblastoma are single-arm studies. In contrast, Trippa *et al.* [4] proposed and evaluated potential benefits of using controlled, response-adaptive multi-arm trials in this context.

Adaptive randomization schemes are developed to obtain a more desirable assignment of patients in the trial to competing treatments compared with balanced designs. Several contributions considered two-arm and multi-arm controlled trials and provide motivations for adaptively tuning the randomization probabilities during the study on the basis of the accumulated outcome data [3, 10, 11, 19]. Response adaptive randomization can be defined as the application of a map, used each time a patient is enrolled in the trial, which transforms the available data into suitable randomization probabilities. Frequentist approaches are direct, in that, intuitive and heuristic rules are used to map the available data into randomization probabilities [20–23]. These maps have been assessed using asymptotic theoretical analysis and in simulation studies [22, 24–26]. In contrast, Bayesian randomization methods are indirect, model based and exploit Bayesian predictions during the trial. The prior distribution π models jointly the primary outcome distributions $\theta_0, \theta_1, \ldots, \theta_K$ for control and experimental treatments. Most Bayesian adaptive strategies map posterior probabilities of treatment effects, say $(\theta_k - \theta_0)$, into randomization probabilities [3, 27, 28].

Bayesian adaptive randomization procedures typically do not maximize an explicit utility function. The computational burden to optimize a sequential multi-arm study motivates the use of heuristic procedures. In different words, we will discuss procedures that replace the decision theoretic paradigm. Zhang *et al*. [29] compare heuristics and decision theoretic optimal designs within the context of biomarker-subgroup trials. The development of nearly optimal assignment procedures tailored to explicit utility functions u remains an attractive area of research.

In Trippa *et al.* [4], we considered a controlled four-arm trial. The response to treatments is evaluated using progression-free survival (PFS) endpoints, and (S_0, S_1, S_2, S_3) denote the unknown time to event distributions for the control arm, $k = 0$, and experimental therapies, $k = 1, 2$ and 3. These are assume to follow a proportional hazards model, with unknown positive hazard ratios $\theta = (\theta_1, \theta_2, \theta_3)$, such that the equalities $S_k(t) = [S_0(t)]^{\theta_k}$ hold, for every $t \geqslant 0$ and $k = 1, 2, 3$.

We use identical symmetric prior distributions with mean zero for the log-hazard ratios $\log(\theta_1), \log(\theta_2)$ and $\log(\theta_3)$. In different words, the prior π assigns symmetric probabilities to scenarios where treatment k has a positive or negative effect compared with the control. We consider time varying randomization probabilities

$$R_i^k = p(i\text{-th enrolled patient is randomized to treatment } k|\text{available DATA}),$$

with $\sum_{k=0}^{3} R_i^k = 1$, defined by the following expressions:

$$\begin{aligned} R_i^k &\propto \frac{p\left(\theta_k < 1 \mid \text{available DATA}\right)^{\gamma(i)}}{\sum_{\ell=1,2,3} p\left(\theta_\ell < 1 \mid \text{available DATA}\right)^{\gamma(i)}} \quad \text{if } k = 1, 2, 3, \text{ and} \\ R_i^0 &\propto \exp\left\{\eta(i) \times \left(\max_{\ell=1,2,3} \#[\text{assignments to arm } \ell] - \#[\text{assignments to control}]\right\}\right. /3 \end{aligned} \qquad (2)$$

The aforementioned two expressions have a clear interpretation. The first one shows that for any choice of the tuning function $\gamma(i) > 0$ the algorithm assigns patients with higher probabilities to experimental arms with evidence of a positive treatment effect $\theta_k < 1$. Natural candidates for the parameters $\gamma(i)$ are non-decreasing functions with values close to zero during the initial stage of the trial. The second expression aims at approximately matching patient accrual to the control treatment and the number of patients on the experimental arm with the highest patient accrual. In our experience, values of η close to 0.25 during the final stage of the trial suffice to obtain the desired balance without making treatment assignment highly predictable.

Thall and Wathen [3] suggested $\gamma(i) = a \times i^b$, with $a, b \geqslant 0$ and recommended using $(a, b) = (1/(2 \times \text{Sample size}), 1)$ for binary endpoints. Trippa *et al.* [4] used $\gamma(i)$ linear with respect to the index i and recommended the tuning $\gamma(i)$ using simulations under a set of plausible scenarios. We refer to [4, 30] for details.

Alternative definitions of randomization probabilities have been suggested in the literature, for instance, $R^k \propto p(\cap_{k'=0}^{3}\{\theta_k \leqslant \theta_{k'}\}|\text{ available DATA})^{\gamma(i)}$, for $k = 1, \cdots, 3$, [31, 32]. In the case of multiple effective arms, with different treatment effects sizes, this randomization scheme tends to assign most patients only to the arm with the largest treatment effect. In contrast, with the adaptive randomization algorithm (2), the posterior probabilities $p(\theta_k \leqslant 1|\text{ available DATA})$ tend to become close to 1 for all the effective experimental arms, and adaptive randomization will therefore assign more patients to all the effective arms compared with balanced randomization.

Figure 3 shows the distribution of the arm specific sample sizes under a fixed scenario with one effective arm across simulated trials. Trippa *et al.* [4] showed, using simulations under several scenarios, significant gains of the Bayesian design in statistical power and in the number of patients assigned to effective therapies compared with balanced treatment assignment. Also, Wason and Trippa [33] compared the Bayesian outcome-adaptive design with alternative multi-arm multi-stage trial designs. The authors showed that the Bayesian design is more efficient than multi-arm multi-stage trial designs when there are effective experimental treatments; while if none of the experimental treatments is effective, the designs have similar operating characteristics.

2.3. The endTB trial: an adaptive study in tuberculosis

In 2010, there were an estimated 650,000 prevalent cases of multi-drug resistant tuberculosis (MDR-TB), and nearly 500,000 new cases emerge annually through acquisition of resistance during treatment and through airborne transmission [34]. The need for new regimens is therefore indisputable. The recent conditional approval by regulatory authorities of two new anti-TB drugs, bedaquiline and delamanid, presents the first opportunity of a significant improvement in the treatment of MDR-TB since half a century.

The endTB study is a phase III trial that seeks to evaluate five novel treatments for MDR-TB. The study is sponsored by Médecins sans Frontiéres, planned in conjunction with Partners In Health, Harvard Medical School, Epicentre, and the Institute for Tropical Medicine, and supported by UNITAID. It will generate evidence on efficacy and recommendations for those arms that will show treatment effects. The endTB is estimated to have a final enrollment of 750 patients. We designed the trial using Bayesian outcome-adaptive randomization [30], adapting on surrogate and primary endpoints based on joint modeling of the binary culture conversion outcome after 8 and 39 weeks of treatment

$$\theta_{39,k} = \theta_{39,PR,k} \times \theta_{8,k} + \theta_{39,NR,k} \times (1 - \theta_{8,k}).$$

Here $\theta_{8,k}$ denotes the probability of a positive early response to therapy k after 8 weeks of treatment, while $\theta_{39,PR,k}$ and $\theta_{39,NR,k}$ are the response rates after 39 weeks given a positive (PR) or negative (NR) 8-week response. We include interim analyses at regular intervals; after a total of 100, 200, and so on, primary outcomes become available.

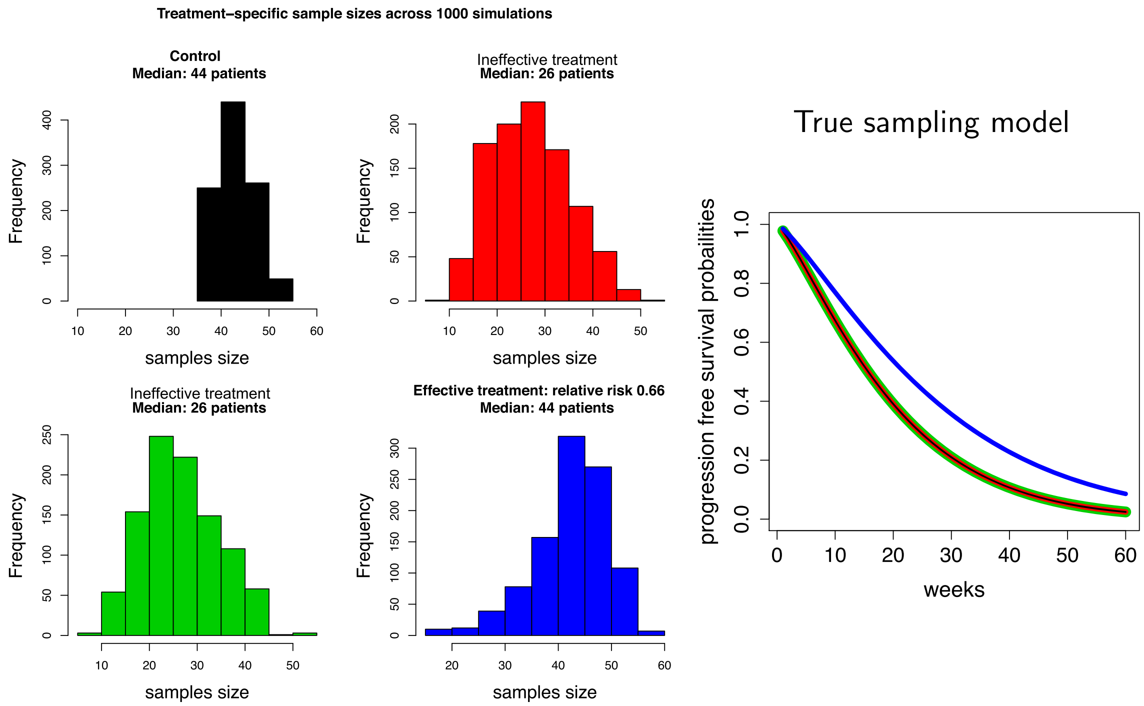

Figure 3. Distribution of patients accrual to the control and experimental arms across simulated trials using Bayesian adaptive randomization. The left panel shows the distribution of patients accrual for each therapy. The right panel shows the true progression-free survival for the control and experimental arms. [Colour figure can be viewed at wileyonlinelibrary.com]

Experimental arms are dropped for futility if the available data, and posterior probabilities suggest no treatment effect on the primary outcome. Arm $k > 1$ is dropped for futility if

$$p(\theta_{39,k} > \theta_{39,0}|\text{ available DATA}) \leqslant b_{\mathrm{f}},$$

where $b_{\mathrm{f}} \in [0, 1]$ and $\theta_{39,0}$ denotes the response rate of the control treatment. The endTB trial uses outcome adaptive randomization, followed, at the end of the trial by frequentist analyses using a rigorous control of prespecified targeted type I error rates. In Sections 3.2 and 3.3 later, we discuss algorithms for the control of type I error rates of adaptive trials.

A detailed study of the design is provided in [30]; here we provide a brief summary of the results. When we compare power under adaptive randomization to several alternative designs with realistic simulation scenarios, we observe that Bayesian outcome-adaptive randomization requires fewer patients than alternative designs to achieve the targeted power. See Figure 4 for two examples. Moreover, Bayesian adaptive randomization consistently allocates more participants to effective arms compared with alternative designs.

2.4. Combining progression-free and overall-survival outcomes in glioblastoma

We recently considered the use of a surrogate PFS outcome jointly with the primary overall survival (OS) outcome in glioblastoma [36]. One potential way to shorten the time from trial initiation to early results of efficacy is to use imaging-based assessments of progression, such as PFS, with earlier times to event than OS [37]. Furthermore, because experimental therapies most directly influence the time until progression, it can be easier to detect effects on PFS, especially if there is long and heterogeneous treatment post progression [38]. There has been some concern, however, regarding the use of progression-based endpoints for clinical trials in neuro-oncology. While outcomes, such as OS, may have clear clinical relevance, endpoints based on imaging assessments, such as response or progression status, are not as clearly linked to patients benefit [39]. It is not trivial to anticipate how positive effects on overall response rates or PFS translate to effects in OS [40]. The approach that we followed is similar to the one illustrated for the endTB trial.

Trippa *et al.* [36] defined an adaptive randomization procedure for multi-arm trials based on a joint Bayesian model for PFS and OS outcomes. The model includes $(K + 1)$ PFS distributions $(S_{PFS,0}, \cdots, S_{PFS,K})$ and $(K + 1)$ OS distributions $(S_{OS,0}, \cdots, S_{OS,K})$, one for each of the K experimental arms and the control arm $k = 0$. Survival distributions are assumed to follow a proportional hazard model $S_{x,k} = S_{x,0}^{\theta_{x,k}}$ for both $x = PFS, OS$ and $k = 1, 2, 3$ with joint prior distribution for the unknown hazard ratios $\pi(\theta_{PFS}, \theta_{OS})$. Adaptation based on OS leverages on early PFS information through the joint model

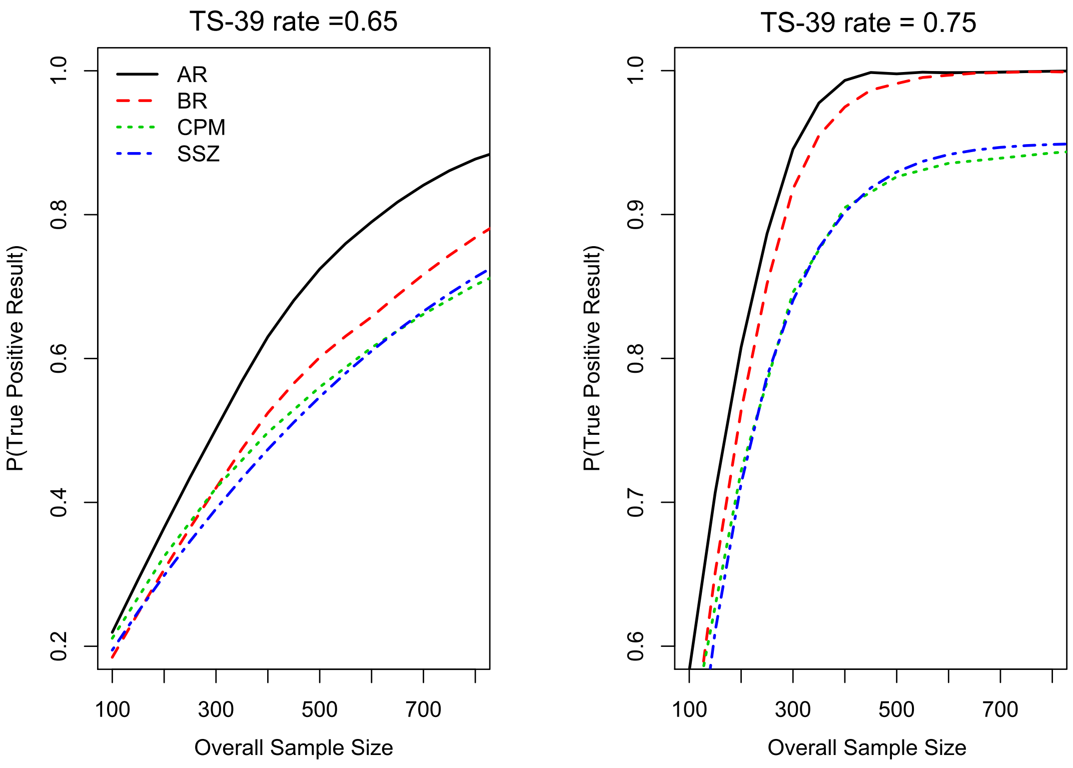

Figure 4. Power comparison under Bayesian adaptive-randomization (AR) and three alternative balanced randomized designs. The first alternative design is a group-sequential balanced design with stopping boundaries defined by the conditional-power method (CPM) [35], whereas the second alternative design has identical stopping rules as the adaptive AR design but uses balanced randomization (BR). The SSZ (single-stage z test) design has no interim analyses and uses balanced treatment assignment. The panels correspond to scenarios with two effective treatments and a 39-week response rate of 0.65 (left penal) and 0.75 (right panel) for both effective therapies compare with 0.5 for the control. [Colour figure can be viewed at wileyonlinelibrary.com]

$\pi(\theta_{PFS}, \theta_{OS})$. At each patients' enrollment, the posterior distributions of θ_{OS} given available PFS and OS outcomes are translated into randomization probabilities.

Advantages of joint modeling in this setting can be summarized by two properties. First, when treatment effects $\theta_{PFS,k}$ and $\theta_{OS,k}$ for PFS and OS are concordant, the proposed approach results in efficiency gains compared with randomization based on OS alone while sacrificing minimal efficiency compared with using PFS as the primary endpoint. Second, if treatment effects are limited to PFS, our approach provides randomization probabilities that become similar to adaptive randomization probabilities defined using only OS data. The alternative to our composite model would be to use OS only. Results in Trippa *et al*. [36] showed that the OS-only adaptive design still results in efficiency gains compared with balanced randomization.

3. Computational methods

We discuss computational approaches, which helped us to evaluate and monitor frequentist operating characteristics for Bayesian designs. We illustrate (i) a stochastic search algorithm for the optimal constrained design $d_{CO-\max}$, followed by (ii) a bootstrap procedure and (iii) an importance sampling algorithm, which we used for the endTB and glioblastoma trial designs to control frequentist operating characteristics.

3.1. Simulated annealing for constrained optimal designs

Finding constrained optimal designs analytically is infeasible in most cases. Stochastic search procedures can be used to approximate $d_{CO-\max}$.

We describe a simulated annealing algorithm for finding $d_{CO-\max}$ [2,41]. The procedure is summarized in Algorithm 2. The algorithm approximately identifies the constrained optimum within a compact set of candidate designs. The procedure starts from a candidate design d_1, for instance, by generating random designs d from the set of designs, which satisfy the desired regulatory constraints V, and then selecting the design with the highest expected utility as starting value d_1. In the Bridging trial, in Section 2.1, a design is represented by thresholds $z_{1:T}$, while V specifies a bound on type II/I error rates for these thresholds.

The simulated annealing algorithm generates an inhomogeneous Markov sequence of designs d_t within $OC^{-1}(V)$. We build on and modify the simulation of a Markov chain with transition probabilities targeting $\exp(U(d)\lambda_t)$ at time t, where λ_t is an increasing sequence of multipliers [2]. At each iteration, the algorithm generates a design $d^\star \in OC^{-1}(V)$ from a proposal distribution g_t with value $d^\star$ in a neighborhood of the current state of the Markov chain d_t. The chain selects $d_{t+1} = d^\star$ with probability $w_t = \min\{1, \exp[(U(d^\star) - U(d_t)) * \lambda_t]\}$ and otherwise sets $d_{t+1} = d_t$ with probability $1 - w_t$. The acceptance probability w_t is an increasing function of the difference of expected utility $(U(d^\star) - U(d_t)) * \lambda_t$. Under regularity conditions [41], the chain will eventually converge to a (local) maximum. We approximate $d_{CO-\max}$ with the simulated design d_t that attained the highest expected utility. Details are outlined in Algorithm 2.

Algorithm 2 Simulated annealing for constrained optimal designs

```
set $t = 1$
Select an initial design $d_t \in D$ with operating characteristics $OC(d_t)$ in $V$
Compute the expected utility $U(d_t)$ of the design $d_t$
**for** $t$ equal to $1, \cdots, T$ **do**
    Generate a design $d^\star \sim g_t$ from a neighborhood $B(d_t, \epsilon_t) \cap OC^{-1}(V)$ of $d_t$
    Compute $\Delta_t = U(d^\star) - U(d_t)$
    Compute the acceptance probability $w_t = \min(1, \exp\{\Delta_t \lambda_t\})$
    Generate $U \sim U(0, 1)$ and select $d_{t+1} = d^\star$ if $\Delta_t \leqslant U$ and $d_{t+1} = d_t$ otherwise.
**end for**
**Return:** $\widehat{d}_{CO-\max} = \arg\max_{d \in \{d_1, \ldots, d_T\}} U(d)$
```

3.2. *A bootstrap scheme for controlling type I error rates*

We describe a bootstrap algorithm for combining Bayesian designs with frequentist analyses. The algorithm is a variation of the bootstrap scheme proposed in [42] for computing confidence intervals and is summarized below in Algorithm 3. The procedure is implemented separately for each treatment arm k of a response-adaptive trial and tests the presence of a treatment effect for experimental arm k, with null hypothesis H_k. The null hypothesis H_k is tested using a statistic Z_k that summarizes the observed outcomes of arm k compared with the control. The bootstrap procedure estimates the distribution of the test statistic under the response-adaptive design and assuming that H_k holds. It generates replicates $Z_k^{(t)} \sim P_{\hat{F}}(Z_k^{(t)})$ of the test statistics under the response-adaptive design. Here $\hat{F} = (\hat{F}_0, \cdots, \hat{F}_K)$ denote estimates of the outcome distributions of different arms.

First, based on the data generated by the adaptive trial $\mathcal{T}$, and for a fixed arm of interest k, the test statistics Z_k is computed. In the TB trial [30], we use the standardized difference between the culture conversions proportions of experimental arm k and control therapy. Second, we compute for every arm $k' = 0, \ldots, K$, consistent estimates of the outcomes distribution $\hat{F}_{k'}$ assuming that the null hypothesis H_k of no treatment effect for arm k holds. In the TB study, for example, this includes estimation of (i) the response probabilities for the surrogate endpoint and (ii) the conditional response probabilities for the primary endpoint, given a positive and negative early outcome. A consistent estimator of the accrual rate is also computed. Third, to test H_k, we simulate $t = 1, \cdots, T$ adaptive trials $\mathcal{T}_t$, with the same stopping rules and tuning parameters as used in the actual trial. For the t-th simulated trial, patients are randomized accordingly to the estimated accrual rate, and each patient assigned to arm k' responds to treatment with probability identical to the estimate $\hat{F}_{k'}$. Note that patients respond to treatment k across the T simulations with probabilities that might be different from those observed in the actual trial because simulations have to be consistent with the null hypothesis H_k that we test. For each simulated trial $\mathcal{T}_t$, we then obtain a statistic $Z_k^{(t)}$, which represents approximately a draw from the null distribution under H_k. We finally estimate the p-value as the proportion of simulated trials with statistic $Z_k^{(t)}$ larger than the observed Z_k. Last, the null hypothesis for arm k is rejected when the p-value is below the prespecified α level. See Algorithm 3 for details.

3.3. *Control of type I error rates with importance sampling*

Importance sampling has been recently used as an alternative approach to control the type I error under a prespecified threshold α in [33]. To simplify the presentation, we assume binary outcomes with response probabilities $\theta = (\theta_0, \ldots, \theta_K)$, one for each of the $K + 1$ therapies. Let Z be a summary statistic that, similarly to a p-value, evaluates evidence against a generic null hypothesis H_0, with large values indicating strong evidence against it. The approach is applicable to both Bayesian and non-Bayesian adaptive randomization schemes, and it is summarized in Algorithm 4.

Algorithm 3 A Bootstrap algorithm for testing treatment efficacy of therapy k.

```
Input: A design $d$ and a trial $\mathcal{T}$
Input: The experimental arm $k$ and hypothesis $H_k$ which should be tested
Compute the statistics $Z_k$ for arm $k$
Estimate the accrual rate of the trial by $\hat{\lambda}$
Estimate the outcome distributions for each arm $k'$ under $H_k$ by $\hat{F}_{k'}$
for $t$ in 1 to $T$ do
    Simulate a trial $\mathcal{T}_t$ under $d$ with accrual rate $\hat{\lambda}$ and outcome distributions $\hat{F}_{k'}$
    Compute the statistics $Z_k^{(t)} = Z_k(\mathcal{T}_t)$
end for
reject $H_k$ at level $\alpha$ if $\hat{p}_k = \frac{1}{T}\sum_1^T I(Z_k^{(t)} > Z_k) \leqslant \alpha$
```

The algorithm estimates the distribution $p(Z \geqslant \cdot|\theta)$ for varying θ values using a single sample of simulated trials $\mathcal{T}_t,\ t = 1, 2, \ldots, T$. Each of these trials, $\mathcal{T}_t$ is generated under the adaptive design with random $\theta^{(t)} \sim g(\theta)$. Then, we determine for a fixed $\alpha \in (0, 1)$ an estimate of the smallest threshold z_α such that $p(Z > z_\alpha|\theta) \leqslant \alpha$ for all θ in H_0.

The procedure iteratively simulates $t = 1, \cdots, T$ adaptive clinical trials varying $\theta^{(t)}$ at each iteration. The response probabilities $\theta^{(t)}$ at each simulation t are generated independently from a continuous distribution g, for instance, a beta distribution. Each simulated trial $\mathcal{T}_t$ is based on a different set of response probabilities $\theta^{(t)} \sim g$, where g is a conveniently selected distribution. Let $p(\mathcal{T}|\theta)$ be the probability of a trial $\mathcal{T}$ under the adaptive scheme and θ; this is the probability of a specific sequence of outcomes and treatment assignments at a fixed value of the vector $\theta = (\theta_0, \ldots, \theta_K)$. We chose the distribution g so that for each generated trial, the importance weights

$$w(\mathcal{T}; \theta) = \frac{p(\mathcal{T}|\theta)}{\int p(\mathcal{T}|\theta')g(\theta')d\theta'}$$

can be straightforwardly computed. Standard importance sampling, using the aforementioned weights, can now be used to approximate the distribution of Z at any fixed θ value (see step 5 of Algorithm 4). Importantly, we can use the same draws $\{\mathcal{T}_t\}$ to approximate $p(Z \geqslant \cdot|\theta)$ for different θ values.

The second part of the algorithm estimates the cutoff point z with the constraint that for every value θ consistent with the null hypothesis H_0, the inequality $p(Z \geqslant z|\theta) < \alpha$ holds. That is, the cutoff point z controls the type I error at the α level. The algorithm uses a grid of values for θ and selects z such that the estimated type I error rate – obtained by importance sampling – across possible θ values is bounded by a prespecified α.

Both importance sampling and the bootstrap algorithm can be used for frequentist analyses in complex adaptive trials, with or without explicitly using a utility function to quantify investigators preferences. The bootstrap algorithm is simpler to implement and approximates the sampling distribution of test statistics. Importance sampling may be computationally more demanding, and it can provide arbitrarily precise estimates of the distribution of the test statistics by increasing the number of generated trials.

Algorithm 4 Importance Sampling for the control of type I error rates

```
1: Simulate $T$ response probabilities $\theta^{(t)} = (\theta_0^{(t)}, \cdots, \theta_K^{(t)}) \sim g(\theta), t = 1, \cdots, T$
2: Generate a trial $\mathcal{T}_t$ under design $d$ with patients response rates $\theta^{(t)}$ for each $t = 1, \cdots, T$
3: Compute the statistics $Z^{(t)}, t = 1, \cdots, T$
4: For each trial $t$ compute the importance weight
       $w(\mathcal{T}_t; \theta) = \frac{p(\mathcal{T}_t|\theta)}{\int p(\mathcal{T}_t|\theta')g(\theta')d\theta'}$
5: Approximate the type I error for the threshold $z$ at $\theta$ by
       $\hat{p}(Z \geqslant z|\theta) = T^{-1} \sum_{t=1}^{T} \frac{w(\mathcal{T}_t; \theta)}{\sum_\ell w(\mathcal{T}_\ell; \theta)} \times I(Z^{(t)} \geqslant z)$
6: Compute $\hat{z}_\alpha = \min\{z : \hat{p}(Z \geqslant z|\theta) \leqslant \alpha$ for all $\theta in H_0\}$
```

4. Summary

Clinical trials are evolving from traditional two-arm studies in large heterogeneous patient populations towards studies with many subpopulations, multiple research questions, and substantial correlative analyses [43].

Traditional frequentist and Bayesian designs are often challenged by these new directions, which demand designs that are applicable in a variety of settings, and can be tailored towards specific research questions [44–46].

Bayesian designs, which enable the use of explicit or implicit utilities u and prior probabilities π to incorporate existing information in the design, can be tailored to specific study purposes [47, 48].

Clinical investigators and medical journals are typically familiar with frequentist measures of evidence. Bayesian testing using Bayesian factors or posterior probabilities, while based on coherent foundational axioms, can be difficult to communicate to these audiences. In addition, regulatory agencies, for instance the US Food and Drug Administration, continue to make systematic use of frequentists testing principles for drug approval and practice changing recommendations.

We described several Bayesian clinical trial designs and presented algorithms for frequentist data analyses. Bayesian designs are combined with frequentist analyses [49]. The use of Bayesian designs is motivated by the desire to optimize the acquisition of information about the clinical utility of therapies by incorporating available prior knowledge and using response-adaptive assignments rules. The use of frequentist analysis is motivated by the desire to communicate results of clinical trials to the medical community, pharmaceutical companies, and regulatory authorities using widely accepted frequentist metrics.

4.1. Supplementary materials

The web-based supplementary material contains R code with examples of the algorithms that we discussed and a pdf-animation of the cut-and-zoom-in algorithm. An additional R package that implement bootstrap analyses for various Bayesian adaptive designs can be found at http://bcb.dfci.harvard.edu/~steffen/software.html.

Acknowledgement

Lorenzo Trippa has been funded by the Claudia Adams Barr Program in Innovative Cancer Research and the Burroughs Wellcome Fund in Regulatory Science.

References

1. Yi Cheng DAB, Su F. Choosing sample size for a clinical trial using decision analysis. *Biometrika* 2003; **90**(4):923–936.
2. Ventz S, Trippa L. Bayesian designs and the control of frequentist characteristics: a practical solution. *Biometrics* 2015; **71**(1):218–226.
3. Thall PF, Wathen KJ. Practical Bayesian adaptive randomisation in clinical trials. *European Journal of Cancer* 2007; **5**:859–866.
4. Trippa L, Lee EQ, Wen PY, Batchelor TT, Cloughesy T, Parmigiani G, Alexander BM. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *Journal of Clinical Oncology* 2012; **30**:3258–3263.
5. Lee JJ, Chen N, Yin G. Worth adapting? revisiting the usefulness of outcome-adaptive randomization. *Clinical Cancer Research* 2012; **18**(17):4498–4507.
6. DeGroot MH. *Optimal Statistical Decisions*, vol. 82. McGraw-Hill Book Company: New York, U.S.A, 1970.
7. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine* 1995; **14**(3):231–246.
8. Berry DA, Fristedt B. *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall: New York, 1985.
9. Inoue LYT, Berry DA, Parmigiani G. Relationship between Bayesian and frequentist sample size determination. *The American Statistician* 2005; **59**(1):79–87.
10. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR, Tsao A, Stewart DJ, Hicks ME, Erasmus J, Gupta S, et al. The battle trial: personalizing therapy for lung cancer. *Cancer Discovery* 2011; **1**(1):44–53.
11. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-spy 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics* 2009; **86**(1):97–100.
12. Jack Lee J, Chu CT. Bayesian clinical trials in action. *Statistics in Medicine* 2012; **31**(25):2955–2972.
13. Freidlin B, Korn EL. Biomarker-adaptive clinical trial designs. *Pharmacogenomics* 2010; **11**(12):1679–1682.
14. Bowden J, Trippa L. Unbiased estimation for response adaptive clinical trials. *Statistical Methods in Medical Research* 2015. DOI: 10.1177/0962280215597716.
15. Chow S-C, Chiang C, Liu J-p, Hsiao C-F. Statistical methods for bridging studies. *Journal of Biopharmaceutical Statistics* 2012; **22**(5):903–915.
16. Chen M-H, Ibrahim JG, et al. The relationship between the power prior and hierarchical models. *Bayesian Analysis* 2006; **1**(3):551–574.
17. Quant EC, Drappatz J, Wen PY, Norden AD. Recurrent high-grade glioma. *Current Treatment Options in Neurology* 2010; **12**(4):321–333.
18. Berry DA. Adaptive clinical trials: the promise and the caution. *Journal of Clinical Oncology* 2011; **29**(6):606–609.
19. Lee JJ, Gu X, Liu S. Bayesian adaptive randomization designs for targeted agent development. *Clinical Trials* 2010; **7**:584 –596.
20. Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association* 1978; **73**(364): 840–843.

21. Eisele JR. The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference* 1994; **38**(2): 249–261.
22. Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML. Optimal adaptive designs for binary response trials. *Biometrics* 2001; **57**(3): 909–913.
23. Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*. John Wiley & Sons: Hoboken, New Jersey, 2016.
24. Hu F, Rosenberger WF. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* 2003; **98**(463):671–678.
25. Hu F, Zhang L-X. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Annals of Statistics* 2004; **32**(1): 268–301.
26. Zhang L-X, Hu F, Cheung SH, Chan WS. Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics* 2007:1166–1182.
27. Lee JJ, Liu DD. A predictive probability design for phase ii cancer clinical trials. *Clinical Trials* 2008; **5**(2):93–106.
28. Yin G, Chen N, Jack Lee J. Phase II trial design with Bayesian adaptive randomization and predictive probability. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2012; **61**(2):219–235.
29. Zhang Y, Trippa L, Parmigiani G. Optimal Bayesian adaptive trials when treatment efficacy depends on biomarkers. *Biometrics* 2016; **72**(2): 414–421.
30. Cellamare M, Ventz S, Boudin E, Mitnick C, Trippa L. A Bayesian response-adaptive trial in tuberculosis: The endTB trial. *Clinical Trials* 2017; **14**(1):17–28.
31. Berry SM, Carlin BP, Lee JJ, Muller P. *Bayesian Adaptive Methods for Clinical Trials*. CRC Press, New York, 2010.
32. Yuan Y, Guo B, Munsell M, Lu K, Jazaeri A. Midas: a practical Bayesian design for platform trials with molecularly targeted agents. *Statistics in Medicine* 2016; **35**(22):3892–3906. https://doi.org/10.1002/sim.6971.
33. Wason J, Trippa L. A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Statistics in Medicine* 2014; **33**(13):2206–2221.
34. Organization WH. *Global tuberculosis report 2015*, 2015. World Health Organization.
35. Betensky RA. Alternative derivations of a rule for early stopping in favor of h_0. *The American Statistician* 2000; **54**(1):35–39.
36. Trippa L, Wen PY, Parmigiani G, Berry DA, Alexander BM. Combining progression-free survival and overall survival as a novel composite endpoint for glioblastoma trials. *Neuro Oncology* 2015; **17**(8):1106–1113.
37. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute* 2009; **101**(23):1642–1649.
38. Terasaki M, Murotani K, Narita Y, Nishikawa R, Sasada T, Itoh AYK, Morioka M. Controversies in clinical trials of cancer vaccines for glioblastoma. *Journal of Vaccines & Vaccination* 2013; **4**(1):171–172.
39. Alexander BM, Trippa L. Progression-free survival: too much risk, not enough reward?. *Neuro-oncology* 2014; **16**(5):615–616.
40. Alexander BM, Galanis E, Yung WKA, Ballman KV, Boyett JM, Cloughesy TF, Degroot JF, Huse JT, Mann B, Mason W, et al. Brain malignancy steering committee clinical trials planning workshop: report from the targeted therapies working group. *Neuro Oncology* 2015; **17**(2):180–188.
41. Robert C, Casella G. *Monte Carlo Statistical Methods*. Springer Science & Business Media: New York, 2004.
42. Rosenberger WF, Hu F. Bootstrap methods for adaptive designs. *Statistics in Medicine* 1999; **18**(14):1757–1767.
43. Harrington D, Parmigiani G. I-spy 2 a glimpse of the future of phase 2 drug development? *New England Journal of Medicine* 2016; **375**(1):7–9.
44. Esserman LJ, Woodcock J. Accelerating identification and regulatory approval of investigational cancer drugs. *JAMA* 2011; **306**(23):2608–2609.
45. Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA* 2015; **313**(16):1619–1620.
46. Trippa L, Alexander BM. Bayesian baskets: a novel design for biomarker-based clinical trials. *Journal of Clinical Oncology* 2016; **35**(6): 681–687.
47. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science* 2004; **19**(1):175–187. DOI: https://doi.org/10.1214/088342304000000044.
48. Thall PF. Bayesian models and decision algorithms for complex early phase clinical trials. *Statistical Science* 2010; **25**(2):227–244. DOI: https://doi.org/10.1214/09-STS315.
49. Etzioni R, Kadane JB. Optimal experimental design for another's analysis. *Journal of the American Statistical Association* 1993; **88**:1404–1411.