

RESEARCH ARTICLE

# Clinical Trial Adaptation by Matching Evidence in Complementary Patient Sub-groups of Auxiliary Blinding Questionnaire Responses

Ognjen Arandjelović\*

Centre for Pattern Recognition and Data Analytics, School of Information Technology, Deakin University, Geelong, Victoria, Australia

\* [ognjen.arandjelovic@gmail.com](mailto:ognjen.arandjelovic@gmail.com)



**OPEN ACCESS**

**Citation:** Arandjelović O (2015) Clinical Trial Adaptation by Matching Evidence in Complementary Patient Sub-groups of Auxiliary Blinding Questionnaire Responses. PLoS ONE 10(7): e0131524. doi:10.1371/journal.pone.0131524

**Academic Editor:** Massimo Pietropaolo, Baylor College of Medicine, UNITED STATES

**Received:** May 13, 2014

**Accepted:** June 3, 2015

**Published:** July 10, 2015

**Copyright:** © 2015 Ognjen Arandjelović. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** The author has no support or funding to report.

**Competing Interests:** The author has declared that no competing interests exist.

## Abstract

Clinical trial adaptation refers to any adjustment of the trial protocol after the onset of the trial. Such adjustment may take on various forms, including the change in the dose of administered medicines, the frequency of administering an intervention, the number of trial participants, or the duration of the trial, to name just some possibilities. The main goal is to make the process of introducing new medical interventions to patients more efficient, either by reducing the cost or the time associated with evaluating their safety and efficacy. The principal challenge, which is an outstanding research problem, is to be found in the question of how adaptation should be performed so as to minimize the chance of distorting the outcome of the trial. In this paper we propose a novel method for achieving this. Unlike most of the previously published work, our approach focuses on trial adaptation by sample size adjustment i.e. by reducing the number of trial participants in a statistically informed manner. We adopt a stratification framework recently proposed for the analysis of trial outcomes in the presence of imperfect blinding and based on the administration of a generic auxiliary questionnaire that allows the participants to express their belief concerning the assigned intervention (treatment or control). We show that this data, together with the primary measured variables, can be used to make the probabilistically optimal choice of the particular sub-group a participant should be removed from if trial size reduction is desired. Extensive experiments on a series of simulated trials are used to illustrate the effectiveness of our method.

## Introduction

Robust evaluation is a crucial component in the process of introducing new medical interventions. Amongst others, these include newly developed medications, novel means of administering known treatments, new screening procedures, diagnostic methodologies, physio-

therapeutical manipulations, and many others. Such evaluations usually take on the form of a controlled clinical trial (or a series thereof), the framework widely accepted as best suited for a rigorous statistical analysis of the effects of interest [1–3] (for a related discussion and critique also see [4]). Driven both by legislating bodies, as well as the scientific community and the public, the standards that the assessment of novel interventions are expected to meet continue to rise. Generally, this necessitates trials which employ larger sample sizes and which perform assessment over longer periods of time. A series of practical challenges emerge as a consequence. Increasing the number of individuals in a trial can be difficult because some trials necessitate that participants meet specific criteria; volunteers are also less likely to commit to participation over extended periods of time. The financial impact is another major issue—both the increase in the duration of a trial and the number of participants result in additional cost to an already expensive process. In response to these challenges, the use of adaptive trials has emerged as a potential solution [5–9].

The key idea underlying the concept of an adaptive trial design is that instead of fixing the parameters of a trial before its onset, greater efficiency can be achieved by adjusting them as the trial progresses [10]. For example, the trial sample size (e.g. the number of participants in a trial), treatment dose or frequency, or the duration of the trial may be increased or decreased depending on the accumulated evidence [11–13].

## Proposed method overview

The method for trial adaptation we describe in this paper extends the analysis presented in [14] which has been greatly influenced by recent work on the analysis of imperfectly blinded clinical trials [15, 16]. Its key contribution was to introduce the idea of trial outcome analysis by patient sub-groups which comprise trial participants matched by the administered intervention (treatment or control) and their responses to an auxiliary questionnaire in which the participants are asked to express their belief regarding their assignment intervention in the closed-form (see *Auxiliary data collection* for a summary of the adopted sub-group stratification method and the original paper [16] for full detail). This framework was shown to be suitable for robust inference in the presence of “unblinding” in a trial [16, 17]. The method proposed in the present paper emerges from the realization that the same framework can be used for trial adaptation by providing information which can be used to make a statistically informed selection of the trial participants which can be dropped from the trial before its completion, without significantly affecting the trial outcome. Thus, the proposed approach falls under the category of trial adaptations by “amending sample size”, in contrast to “dose finding” or “response adapting” methods which dominate previous work [13].

In [16] it was shown that the analysis of a trial’s outcome should be performed by aggregating evidence provided by matched participant sub-groups, where two sub-groups are matched if they contain participants who were administered different interventions but nonetheless had the same responses in the auxiliary questionnaire. Therefore, our idea advanced here is that an informed trial sample size reduction can be made by computing which matched sub-group pair’s contribution of useful information is affected the least with the removal of a certain number of participants from one of its groups.

**Contrast with previous work.** Before introducing the proposed method in detail, it is worthwhile emphasizing two fundamental aspects in which it differs from the methods previously described in the literature. The first difference concerns the nature of the statistical framework which underlies our approach. While the use of Bayesian techniques has become increasingly common in medical statistics [18–22], and perhaps particularly so in the context of clinical trial design and analysis [18, 23–28], with few notable exceptions [29] the existing

work on trial adaptation by sample size adjustment adopts the frequentist paradigm [30–32]. These methods follow the following general pattern: a particular null hypothesis is formulated which is then rejected or accepted using a suitable statistic and the desired confidence requirement (a good review is provided by Jennison and Turnbull [33]). In contrast, the method described in this paper is thoroughly Bayesian in nature.

The second major conceptual novelty of the proposed method lies in the question it seeks to answer. Previous work on trial adaptation by sample size adjustment addresses the question of *whether* the sample size can be reduced while maintaining a certain level of statistical significance of the trial's outcome. In contrast, the present work is the first to ask a complementary question of *which* particular individuals in the sample should be removed from the trial once the decision of sample size reduction has been made. Thus, the proposed method should not be seen as an alternative to any of the previously proposed methods but rather as a complementary element of the same framework.

### Auxiliary data collection

The type of auxiliary data collection we utilize in this work was originally proposed for the assessment of blinding in clinical trials [34]. Since then it has been adopted for the same purpose in a number of subsequent works [16, 35–37] (also see [38] for related commentary).

With the exception of [16], in all previous work the questionnaire is employed after the trial has ended (for discussions on the timing of the questionnaire see [16, 39, 40]). The questionnaire allows the trial participants to express their belief on the nature of the intervention they have been administered (control or treatment) using a fixed number of choices. The most commonly used, coarse-grained questionnaire admits the following three choices:

Choice 1: the patient believes that he/she was administered the control intervention (i.e. control group membership),

Choice 2: the patient believes that he/she was administered the treatment intervention (i.e. treatment group membership), and

Choice 3: the patient is undecided about the nature of the treatment he/she was administered (the “don't know” response).

Extensions of this scheme which attempt to harness more detailed information have also been used, for example allowing the participants to quantify the conviction of their belief as “weak” or “strong”. In that case, the questionnaire would offer five choices:

Choice 1: the patient *strongly* believes that he/she was administered the control intervention,

Choice 2: the patient *weakly* believes that he/she was administered the control intervention,

Choice 3: the patient is undecided about the nature of the treatment he/she was administered,

Choice 4: the patient *weakly* believes that he/she was administered the treatment intervention, and

Choice 5: the patient *strongly* believes that he/she was administered the treatment intervention.

More granular auxiliary data choices have the potential of providing a more accurate picture of the extent of blinding. However, depending on the statistical model used, this advantage may come at the cost of reduced statistical significance for each of the response sub-groups. The evidence on the capacity of a human's working memory [41] suggests that the number of comparisons a typical person can distinguish at any one time is approximately  $7 \pm 2$ , which leads to the conclusion that it is best to limit the number of choices to fewer than ten.

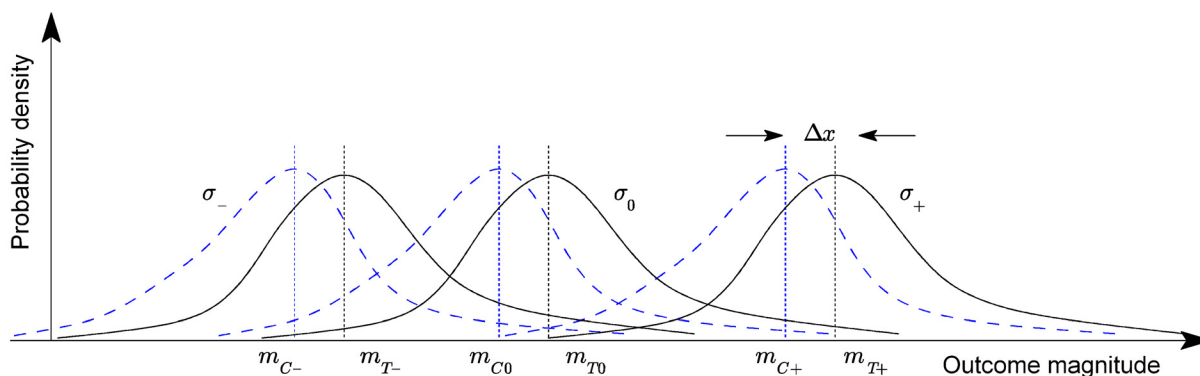
### Matching sub-groups outcome model

In the general case, the effectiveness of a particular intervention in a trial participant depends on the inherent effects of the intervention, as well as the participant's expectations (conscious or not). Thus, as in [16], in the interpretation of trial results, we separately consider each population of participants which share the same combination of the type of intervention and the expressed belief regarding this group assignment. For example, when a 3-tier questionnaire is used in a trial comparing the administration of the treatment of interest and control, we recognize  $3 \times 2 = 6$  auxiliary data sub-groups:

- Sub-group 1: participants of the control group who believe they were assigned to the control group (sub-group  $G_{C-}$ ),
- Sub-group 2: participants of the control group who are unsure of their group assignment (sub-group  $G_{C0}$ ),
- Sub-group 3: participants of the control group who believe they were assigned to the treatment group (sub-group  $G_{C+}$ ),
- Sub-group 4: participants of the treatment group who believe they were assigned to the control group (sub-group  $G_{T-}$ ),
- Sub-group 5: participants of the treatment group who are unsure of their group assignment (sub-group  $G_{T0}$ ), and
- Sub-group 6: participants of the treatment group who believe they were assigned to the treatment group (sub-group  $G_{T+}$ ).

This is illustrated conceptually in Fig 1. In the general case, for an  $N$ -tier questionnaire and  $M$  different intervention types, we can distinguish between  $N \times M$  distinct sub-groups of participants.

The key idea underlying the method proposed in [16] is that because the outcome of an intervention depends on both the inherent effects of the intervention and the participants' expectations, the effectiveness should be inferred in a like-for-like fashion. In other words, the response observed in, say, the sub-group of participants assigned to the control group whose feedback professes belief in the control group assignment should be compared with the response of only the sub-group of the treatment group who equally professed belief in the



**Fig 1. Adopted statistical model for a three-tier feedback questionnaire (conceptual illustration)—the probability densities of the measured trial outcome across the three control (solid lines) and treatment sub-groups (dotted lines).** Diagram reproduced with permission.

doi:10.1371/journal.pone.0131524.g001

control group assignment. Similarly, the “don’t know” sub-groups should be compared only with each other, as should the sub-groups corresponding to the belief in the treatment assignment. Ideas similar in spirit were expressed by Berger in the consideration of the related problem of so-called selection bias and specifically the Berger-Exner test [42].

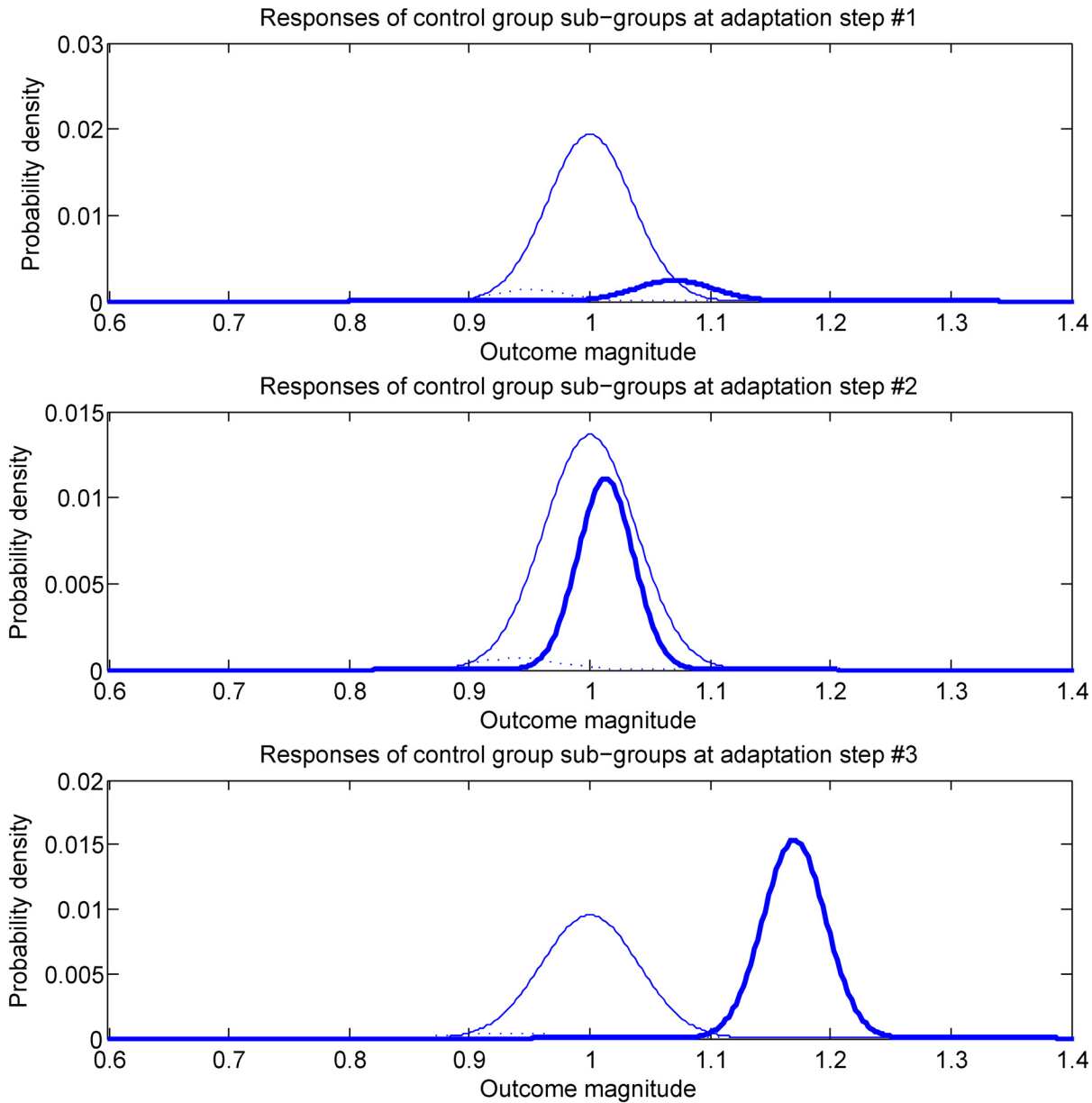
## Sub-group selection

The primary aim of the statistical framework described in [16] is to facilitate an analysis of trial data robust to the presence of partial or full unblinding of patients, or indeed patient preconceptions which too may affect the measured outcomes. Herein we propose to exploit and extend this framework to guide the choice of which patients are removed from the trial after its onset, in a manner which minimizes the loss of statistical significance of the ultimate outcomes.

At the onset of the trial, the trial should be randomized according to the current best clinical practice; this problem is comprehensively covered in the influential work by Berger [42]. If a reduction in the number of trial participants was attempted at this stage, by the very definition of a properly randomized trial, statistically speaking there is no reason to prefer the removal of any particular subject (or indeed a set of subjects) over another. Instead, any trial size adaptation must be performed at a later stage after some meaningful differentiation between subjects takes place [43–45].

The most obvious measurable differentiation that takes place between patients as the trial progresses is that of the outcomes of primary interest in the trial (the “response”). This differentiation may allow for a statistically informed choice to be made about which trial participants can be dropped from the trial in a manner which minimizes the expected distortion of the ultimate findings. For example, this can be done by seeking to preserve the distribution of measured outcomes within a group (treatment or control) but with the constraint of a smaller number of participants; indeed, our approach partially exploits this idea as will be explained in detail shortly. However, our key contribution lies in a more innovative approach, which exploits additional, yet readily collected discriminative information. The proposed approach not only minimizes the effect of smaller participant groups but also ensures that no unintentional bias is injected due to imperfect blinding. Recall that the problem of inference robust to imperfect blinding should always be considered—as stated earlier, blinding can only be attempted with respect to those variables of the trial which have been identified as revealing of the administered treatment, and even for the explicitly identified variables it is fundamentally impossible to ensure that absolute blinding is achieved.

Our idea is to administer an auxiliary questionnaire of the form described in [34, 35] (which is normally administered after the trial in the work on blinding assessment) every time an adaptation of the trial group size (i.e. reduction thereof) is sought. Just like in [16], this leads to the differentiation of each group of participants (control or treatment) into sub-groups, based on their belief regarding their group assignment. In general, this means that even if no participants are removed from the trial, a participant may change his/her sub-group membership status. This is illustrated conceptually in Fig 2. The first time an auxiliary questionnaire is administered (top-most plot), most of the treatment group participants are still unsure of their assignment (solid blue line); a smaller number of participants have correctly guessed (or inferred) their assignment (bold blue line); lastly, an even smaller number holds the incorrect belief that they are in fact members of the control group (dotted blue line). All of the sub-groups show a spread of responses to the treatment, such as may be expected due to various personal variations of their members. At the time of the second snapshot (plot in the middle), at the next instance when auxiliary data is collected, the proportions of participants in each



**Fig 2. A conceptual illustration on a hypothetical example of the phenomenon whereby trial participants change their sub-group membership (recall that each sub-group is defined by its members' intervention assignment and auxiliary questionnaire responses). This is quite likely to occur when the effects of the treatment are very readily apparent but various other mechanisms can act so as to cause a non-zero and changing sub-group flux.**

doi:10.1371/journal.pone.0131524.g002

sub-group has changed, as do the associated treatment response statistics. A similar observation can be made with respect to the third and the last snapshot pictured in the figure (bottom-most plot). This sort of a development would not be unexpected—if the treatment is effective, as the trial progresses there will be an increase in the number of treatment group participants who observe and correctly interpret these changes (note that this also means that there will be an associated increase in the number of participants who may exhibit an additional positive effect from the fortunate realization that they are receiving the studied treatment intervention, rather than the control intervention). That being said, it should be emphasized that no



assumption on the statistics of sub-group memberships or their relative sizes is made in the proposed method. The example in Fig 2 is merely used for illustrative purposes.

The question is: how does this differentiation of patients by auxiliary data sub-groups help us make a statistically robust choice of which participants in the trial should be preferentially dropped if a reduction in the trial size is sought? To answer this question, recall the main premise of [16]:

“The key idea [ . . . ] is that it is meaningful to compare only the corresponding treatment and control participant sub-groups, that is, sub-groups matched by their auxiliary responses.”

Each sub-group comparison contributes information used to infer the probability density of the differential effects of the treatment. We can then reformulate the original question as: from which matching sub-group pair should participants be preferentially dismissed from further consideration so as to best preserve the information contribution from all sub-groups?

Consider how the information on the differential effects between a single pair of matching sub-groups is inferred. In its general form, we can estimate some distance between the distributions of the two sub-groups using a Bayesian approach. Indeed, Bayesian analysis based methods have in recent years been continually gaining acceptance across the clinical community [44]. In brief, the key idea is to “integrate out” the unknown latent parameters of the two distributions. Expressed formally:

$$\rho^* = \int_{\Theta_c} \int_{\Theta_t} \underbrace{\rho(p_c(x; \Theta_c), p_t(x; \Theta_t))}_{\substack{\text{Distance between distributions} \\ \text{for specific parameter values}}} \times \underbrace{p(\Theta_c|D_c) p(\Theta_t|D_t)}_{\substack{\text{Probability of parameters} \\ \text{conditioned on observations}}} \times \underbrace{p(\Theta_c) p(\Theta_t)}_{\text{Parameter priors}} \times d\Theta_t d\Theta_c \quad (1)$$

where  $\Theta_c$  and  $\Theta_t$  are the sets of variables parameterizing the two corresponding distributions  $p_c(x; \Theta_c)$  and  $p_t(x; \Theta_t)$ ,  $p(\Theta_c)$  and  $p(\Theta_t)$  the parameter priors,  $\rho(p_c(x; \Theta_c), p_t(x; \Theta_t))$  a particular distance function (e.g. the Kullback-Leibler divergence [46], Bhattacharyya [47] or Hellinger distances [48], or indeed the posterior of the difference of means used in [16]), and  $D_c$  and  $D_t$  the measured trial outcomes (such as the amount of fat loss in a fat loss trial, the reduction in blood plasma LDL in a statin trial etc).

Note that by changing (reducing) the number of participants in one of the groups, the only affected term on the right hand side of Eq (1) is one of the likelihood terms,  $p(\Theta_c|D_c)$  or  $p(\Theta_t|D_t)$ . Seen another way, a change in the number of participants in the trial changes the *weighting* of the product of the distance term  $\rho(p_c(x; \Theta_c), p_t(x; \Theta_t))$  and the priors  $p(\Theta_c) p(\Theta_t)$ . Our idea is then to choose to remove a trial participant from that sub-group which produces the smallest change in the estimate  $\rho^*$ . However, it is not clear how this may be achieved, since it is the size of the set  $D_c$  that is changing (so, for example, treating  $D_c$  and  $D_t$  as vectors and  $f$  as a function of vectors would not achieve the desired aim). Examining the sensitivity of  $\rho^*$  with the removal of each datum (i.e. trial participant) from  $D_c$  and  $D_t$  is also unsatisfactory since the problem does not lend itself to a greedy strategy: the optimal choice of which  $n_{rem}$  trial participants to drop from the trial cannot be made by making  $n_{rem}$  optimal choices of which *one* participant to drop. An approach following this direction but attempting to examine all possible sets of size  $n_{rem}$  would encounter computational tractability obstacles since this problem is NP-complete. The alternative which we propose is to consider and compare the magnitudes of partial derivatives of  $\rho^*$  with respect to the sizes of data sets  $D_c$  and  $D_t$ , but with an important constraint—the derivatives are taken of the *expected* functional form of  $\rho^*$  over different members

of  $D_c$  and  $D_t$ . Formalizing this, we compute:

$$E \left[ \frac{\partial \rho^*}{\partial n_c} \right]_{D_c} \quad \text{and} \quad E \left[ \frac{\partial \rho^*}{\partial n_t} \right]_{D_t}, \tag{2}$$

where  $E[\rho^*]_{D_c}$  and  $E[\rho^*]_{D_t}$  are respectively the expected values of  $\rho^*$  across the space of possible observations in  $D_c$  and  $D_t$  (each with a uniform prior, as before). Thus  $E[\rho^*]_{D_c}$  and  $E[\rho^*]_{D_t}$  are functions of two scalars, the sizes  $n_c$  and  $n_t$  of sets  $D_c$  and  $D_t$  i.e. the numbers of members of the corresponding sub-groups.

The proposed solution is not only theoretically justified but it also lends itself to simple and efficient implementation. Since the expected values  $E[\rho^*]_{D_c}$  and  $E[\rho^*]_{D_t}$  are evaluated over sets  $D_c$  and  $D_t$ , in Eq (1) the only term affected is  $p(\Theta_c|D_c) p(\Theta_t|D_t)$ , so the solution is readily obtained as a closed form expression. Equally, the integration is readily performed using one of the standard Markov chain Monte Carlo integration methods [49].

### Application example

In order to illustrate how the described idea could be applied in practice, we now consider one specific example of the distance function discussed in the previous section and show how the mathematical results needed to implement the proposed methodology can be derived or estimated. Readers who are not interested in full technical detail of this nature can skip this section and proceed to *Note on practical application* without loss in continuity.

Let the trial observation data in two matching sub-groups be drawn from the random variables  $X_c$  and  $X_t$ , which are appropriately modelled using log-normal distributions [50]:

$$X_t \sim \frac{1}{x \sigma_t} \exp \left\{ -\frac{\ln x - m_t}{2\sigma_t^2} \right\} \tag{3}$$

and

$$X_c \sim \frac{1}{x \sigma_c} \exp \left\{ -\frac{\ln x - m_c}{2\sigma_c^2} \right\} \tag{4}$$

The next step is to choose an appropriate distance function  $\rho$  in Eq (1). In practice, this choice would be governed by the goals of the study. Herein, for illustrative purposes we choose  $\rho$  to be the probability that a patient will do better when the treatment rather than the control intervention is administered:

$$\rho(p_t(x; \Theta_t), p_c(y; \Theta_c)) = \int_0^\infty \int_0^x p_t(x; \Theta_t) p_c(y; \Theta_c) dy dx$$

where  $\Theta_c = (m_c, \sigma_c)$  and  $\Theta_t = (m_t, \sigma_t)$  are the mean and standard deviation parameters specifying the corresponding log-normal distributions.

$$\rho^* \propto \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_0^x \left\{ p_t(x|m_t, \sigma_t) p(m_t, \sigma_t) dx dm_t d\sigma_t \times p_c(y|m_c, \sigma_c) p(m_c, \sigma_c) dy dm_c d\sigma_c \right\} \tag{5}$$

Making the usual substitution whereby  $e^x$  is substituted for  $x$  (and similarly  $e^y$  for  $y$ ), which normalizes the log-normal distribution, and assuming uninformed priors on  $m_c, m_t, \sigma_c$ , and  $\sigma_t$



(just as in [16]) leads to the following expression

$$\begin{aligned} \rho^* &\propto \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_0^x \sigma_t^{-1} e^{-\frac{(x-m_t)^2}{2\sigma_t^2}} \sigma_t^{-N} e^{-\frac{\sum_{i=1}^{n_t} (x_i^{(t)} - m_t)^2}{2\sigma_t^2}} dx dm_t d\sigma_t \\ &\times \sigma_c^{-1} e^{-\frac{(x-m_c)^2}{2\sigma_c^2}} \sigma_c^{-N} e^{-\frac{\sum_{i=1}^{n_c} (x_i^{(c)} - m_c)^2}{2\sigma_c^2}} dy dm_c d\sigma_c \end{aligned} \tag{6}$$

$$= \int_0^\infty \int_0^x I_t(x) I_c(y) dy dx = \int_0^\infty I_t(x) \int_0^x I_c(y) dy dx \tag{7}$$

where each of the integrals  $I_t(x)$  and  $I_c(y)$  has the form:

$$I = \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \times \frac{1}{\sigma^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}\right\} dm d\sigma, \tag{8}$$

$\{x_i\}$  and  $n$  stand for either  $\{x_i^{(c)}\}$  and  $n_c$  or  $\{x_i^{(t)}\}$  and  $n_t$ , and  $\{x_i^{(c)}\}$  ( $i = 1 \dots n_c$ ) and  $\{x_i^{(t)}\}$  ( $i = 1 \dots n_t$ ) are exponentially transformed measured trial variables. This integral can be evaluated by combining the two exponential terms and completing the square of the numerator of the exponent as in [16] so that:

$$(x - m)^2 + \sum_{i=1}^n (x_i - m)^2 \equiv (am + b)^2 + c, \tag{9}$$

which leads to the following simplification of Eq (7):

$$I \propto \int_0^\infty \frac{1}{\sigma^{n+2}} \exp\left\{-\frac{c}{2\sigma^2}\right\} \sigma d\sigma \tag{10}$$

$$= \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{c}{2\sigma^2}\right\} d\sigma, \tag{11}$$

where the value of the only non-constant term,  $c$ , is:

$$c = x^2 + \sum_{i=1}^n x_i^2 - \frac{(x + \sum_{i=1}^n x_i)^2}{n + 1} \tag{12}$$

$$= \frac{(n + 1)(x^2 + \sum_{i=1}^n x_i^2) - (x + \sum_{i=1}^n x_i)^2}{n + 1}. \tag{13}$$

Observing that the form of the integrand in Eq (11) matches that of the inverse gamma distribution:

$$\text{Gamma}(z; \alpha, \beta) = \frac{\beta^z}{\Gamma(\alpha)} z^{-\alpha-1} \exp\{-\beta/z\}. \tag{14}$$

where  $\Gamma(\alpha)$  is the value of the gamma function at  $\alpha$ . The variable  $z$  and the two parameters of the distribution,  $\alpha$  and  $\beta$ , can be matched with the terms in Eq (11) and the density integrated

out, leaving the integral proportional to a single non-constant term:

$$I \propto c^{-\frac{n-1}{2}}. \tag{15}$$

Notice that the form of this result is very similar to that obtained in [16].

Remembering that the functional form of  $c$  is different for the control and the trial groups (since it is dependent on  $x_i$  which stands for either  $x_i^{(c)}$  or  $x_i^{(t)}$ ), and substituting the result from Eq (15) back into Eq (7) gives the following expression for the distance function:

$$\rho^* = \int_0^\infty \int_0^x p_t(x) p_c(y) dy dx \tag{16}$$

$$\propto \int_0^\infty \int_0^x c_t^{-\frac{n_t-1}{2}} c_c^{-\frac{n_c-1}{2}} dy dx \tag{17}$$

$$= \int_0^\infty c_t^{-\frac{n_t-1}{2}} \int_0^x c_c^{-\frac{n_c-1}{2}} dy dx \tag{18}$$

Our goal now is to evaluate  $S_c(\rho^*)$  and  $S_t(\rho^*)$ , the sensitivities of the distance function to the change in the size of respectively the control and the treatment groups. Without loss of generality, let us consider  $S_t(\rho^*)$ —the symmetry of the expression in Eq (18) makes it trivial to apply the same process to the computation of  $S_c(\rho^*)$ . From Eq (18):

$$S_t(\rho^*) \propto \int_{-\infty}^\infty \int_{-\infty}^x S[I_t(x)] I_c(y) dy dx \tag{19}$$

To evaluate  $S[I_t(x)]$  we will employ the standard chain rule and perform differentiation with respect to  $n_t$  when the corresponding term is a function of the number of treatment participants but not any  $x_i^{(t)}$ . On the other hand, as proposed in *Sub-group selection*, to handle those terms which do depend on  $x_i^{(t)}$  (through  $c_t$ ), we will use the expected value of the change in the term, averaged over all possible  $x_i^{(t)}$  that a unitary decrease in  $n_t$  can be achieved. Applying this idea on the expression in Eq (8) leads to the following expansion:

$$S[I_t(x)] = - \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \times \left[ \frac{\ln\sigma}{\sigma^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}\right\} + \frac{1}{\sigma^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}\right\} \frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2 n^2} \right] dm d\sigma \tag{20}$$

noting that we used the standard result:

$$\frac{d}{dn} \frac{1}{\sigma^n} = -\frac{\ln\sigma}{\sigma^n} \tag{21}$$

without including its derivation with intermediary steps shown explicitly. The resulting

expression can be further rearranged, yielding the following, more concise expression:

$$S[I_t(x)] = - \int_0^\infty \int_{-\infty}^\infty \frac{\ln \sigma}{\sigma^n} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{\sum_{i=1}^n (x_i-m)^2}{2\sigma^2}\right\} dm d\sigma \quad (22)$$

$$- \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma^{n+3} n^2} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{\sum_{i=1}^n (x_i-m)^2}{2\sigma^2}\right\} \sum_{i=1}^n (x_i-m)^2 dm d\sigma. \quad (23)$$

Full double integration of the integrands in Eq (23) is difficult to perform analytically. However, one level of integration—with respect to  $m$ —is readily achieved. Note that the first term, as a function of  $m$ , has the same form as the integral in Eq (8) which we already evaluated. The same procedure which uses the completion of the square in the exponential term can be applied here as well (note that unlike in Eq (8) here it is important to keep track of the multiplicative constants as these will be different for the second term in Eq (23)). The integrand in the second term can be expressed in the form  $\propto (z-\lambda)^2 \exp -z^2 dz$ . This integration is also readily performed using the standard results:

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} z^2 \exp -\frac{z^2}{2} dz = 1 \quad (24)$$

and

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp -\frac{z^2}{2} dz = 1, \quad (25)$$

and by noting that the integrand is an odd function

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} z \exp -\frac{z^2}{2} dz = 0 \quad (26)$$

A straightforward application of these results to Eq (23) leads to the following expression for the sensitivity  $S[I_t(x)]$  of the integral  $I_t$  to changes in the size of the corresponding sub-group (the treatment sub-group in a matching pair):

$$S[I_t(x)] = - \int_0^\infty \frac{\ln \sigma}{\sigma^{n+1}} \frac{\sigma}{a} \sqrt{2\pi} \exp\left\{-\frac{c}{2\sigma^2}\right\} d\sigma - \int_0^\infty \frac{\exp\left\{-\frac{c}{2\sigma^2}\right\}}{2\sigma^{n+3} n^2} \left[ n \left(\frac{\sigma}{a}\right)^3 + \sqrt{2\pi} \frac{\sigma}{a} \sum_{i=1}^n x_i \right] d\sigma \quad (27)$$

This result, together with the expression in Eq (15), can be substituted into Eq (19) and the remaining integration performed numerically.

### Note on practical application

As demonstrated in the preceding section, it is worth observing that none of the calculations involved make any strong assumptions on the nature of the phenomenon studied i.e. the nature of the outcome data. There are only two elements of the described framework which are subject to change depending on the trial. These are the distance function  $\rho$  and the form of the likelihood terms  $p(\Theta_c|D_c)$  and  $p(\Theta_t|D_t)$ , see Eq (1). In practice, all of these terms are confined to a small number of reasonable options, which means that all of the derivations described in some

detail in the previous section can be performed *a priori*, with only the final numerical computations (which are not computationally demanding) performed “on the fly”. Consequently, while the implementation of the proposed approach clearly requires expertise in mathematics and computation, no mathematical expertise or thorough understanding of the details of the algorithm need to be expected from the user (e.g. a clinician or a biostatistician). The user would merely need to select one of a number of possible outcomes of interest (e.g. “effect greater than”) and set the value of the relevant parameter according to the aims of the study (e.g. “effect greater than 0.1”), and be automatically presented with the best sample selection and the corresponding posterior distributions quantifying the uncertainty associated with the resultant outcome. This will be illustrated shortly on simulated examples in *Evaluation and analysis*.

### From target sub-groups to specific participants

Adopting the framework proposed in [16] whereby the analysis of a trial takes into account sub-groups of trial participants, which emerge from grouping participants according to their assigned intervention and auxiliary data, thus far we focused on the problem of choosing the sub-group from which participants should be preferentially removed if a reduction in trial size is sought. The other question which needs to be considered is how *specific* sub-group members are to be chosen, once the target sub-group is identified. Fortunately, the proposed framework makes this a simple task. Recall that the observed trial data within each sub-group is assumed to comprise an identically and independently distributed sample from the underlying distribution, i.e.  $x_i^{(c)} \sim X_c$  (or indeed  $x_i^{(t)} \sim X_t$ ). This means that it is sufficient to randomly sample the set of target sub-group members to select those which can be removed.

The simplicity of the selection process that our approach allows has an additional welcome consequence. Recall that in the proposed method the choice of the target sub-group is made by comparing differentials in Eq (2). It is important to observe that their values are computed for the initial values of  $n_c$  and  $n_t$ . Thus, as the number of participants in either of the sub-groups is changed, so do the values of the differentials, and thus possibly the optimal sub-group choice. This is why the removal of participants should proceed sequentially as summarized in Fig 3.

### Evaluation and Analysis

The primary novelty introduced in this paper is of a methodological nature. In the previous section we explained in detail the mathematical process involved in applying the proposed methodology in practice. Pertinent results were derived for a specific distance function used to

- ```

1 repeat
2   | Select the optimal sub-group by evaluating the differentials in (19);
3   | Randomly select and remove a participant from the selected
   | sub-group
   | (the corresponding data set  $D = \{x_i\}$  and its size  $n$  are updated);
4 until desired number of participants removed;
```

**Fig 3. A ‘high-level’ summary of the key steps in the proposed method.**

doi:10.1371/journal.pone.0131524.g003

quantify the difference in the outcomes between the control and treatment groups in a trial. The choice of the distance function—which would in practice be made by the clinicians to suit the aims of a specific trial—governs the relative loss of information when participants are removed from a specific sub-group, and consequently dictates the choice of the optimal sub-group from which the removal should be performed if the overall trial sample size needs to be reduced.

In this section we apply the derived results on experimental data, and evaluate and discuss the performance of the proposed methodology. We adopt the evaluation protocol standard in the domain of adaptive trials research, and obtain data using a simulated experiment.

### Experimental setup

We simulated a trial involving 180 individuals, half of which were assigned to the control and the other half to the treatment group. For each individual we maintain a variable which describes that person’s belief regarding his/her group assignment. Thus, for the control group we have  $n_c$  beliefs  $\{b_i^{(c)}\}$  ( $i = 1 \dots n_c$ ) and similarly for the treatment group  $n_t$  beliefs  $\{b_i^{(t)}\}$  ( $i = 1 \dots n_t$ ). Belief is expressed by a real number,  $\forall i. b_i^{(c)}, b_i^{(t)} \in (-\infty, +\infty)$ , with 0 indicating true undecidedness. Negative beliefs express a preference towards the belief in control group assignment, and positive towards the belief in treatment group assignment. The greater the absolute value of a belief variable is, the greater is the person’s conviction. We employ a three-tier questionnaire. To simulate a participant’s response, we map the corresponding belief to one of the three possible questionnaire responses according to the following thresholding rule:

$$b < -1 \rightarrow \text{Belief in control group assignment} \tag{28}$$

$$-1 \leq b \leq 1 \rightarrow \text{Uncertain (“don’t know”)} \tag{29}$$

$$1 < b \rightarrow \text{Belief in treatment group assignment} \tag{30}$$

The starting beliefs of participants, i.e. their beliefs before the onset of the trial, are initialized as follows:

$$b_i^{(c)} = b_i^{(t)} \begin{cases} -1 & \text{for } i = 1 \dots 8 \\ 0 & \text{for } i = 9 \dots 81 \\ 1 & \text{for } i = 82 \dots 90 \end{cases} \tag{31}$$

Put in informal terms, this initialization reflects the conservative belief of most individuals, and the tendency of a smaller number of individuals to exhibit either “pessimistic” or “optimistic” expectations. Also, notice that the same distribution was used both for the control and the treatment groups, reflecting a well performed randomization in the group assignment process.

**Effect accumulation.** As the trial progresses the effects of the treatment accumulate. These are modelled as positive i.e. the treatment is modelled as successful in the sense that on average it produces a superior outcome in comparison with the control intervention. We model this using a stochastic process which captures the variability in participants’ responses to the same treatment. Specifically, at the discrete time step  $k+1$  (the onset of the trial corresponding to  $k = 0$ ), the effect of the  $i$ -th treatment group participant at the preceding time step  $k$ ,  $e_i^{(t)}(k)$ , is updated in the following manner:

$$e_i^{(t)}(k+1) = e_i^{(t)}(k) + w_i^{(t)}(k+1) \times \exp\left\{-\frac{k+1}{10}\right\} \tag{32}$$

where  $w_i^{(t)}(k + 1)$  is drawn from a normal distribution:

$$W_t \sim \mathcal{N}(0.02, 0.05). \tag{33}$$

Notice that this progression has a ‘ground truth’ asymptote at:

$$\lim_{k \rightarrow \infty} E[e_i^{(t)}(k)] = e_i^{(t)}(0) + 0.02 \times \frac{\exp\{-\frac{1}{10}\}}{1 - \exp\{-\frac{1}{10}\}} \approx e_i^{(t)}(0) + 0.19 \tag{34}$$

Similarly, for the control group participants:

$$e_i^{(c)}(k + 1) = e_i^{(c)}(k) + w_i^{(c)}(k + 1) \times \exp\left\{-\frac{k + 1}{10}\right\} \tag{35}$$

where  $w_i^{(c)}(k + 1)$  is drawn from a normal distribution:

$$W_c \sim \mathcal{N}(0.00, 0.05) \tag{36}$$

By definition, at the onset of the trial there is no effect of the treatment; thus:

$$\forall .i = 1 \dots n_t. e_i^{(t)}(0) = 0 \tag{37}$$

$$\forall .i = 1 \dots n_c. e_i^{(c)}(0) = 0 \tag{38}$$

**Belief refinement.** As the effects of the respective interventions are exhibited, the trial participants have increasing amounts of evidence available guiding them towards forming the correct belief regarding their group assignment. In our experiment this process is also modelled using a stochastic process which is dependent on the magnitude of the effect that an intervention has in a particular participant, as well as uncertainty and differences in people’s inference from observations. At the discrete time step  $k+1$ , the belief of the  $i$ -th control group participant previous at the time step  $k$ ,  $b_i^{(c)}(k)$ , is updated using the following update equation:

$$b_i^{(c)}(k + 1) = b_i^{(c)}(k) + \phi \times e_i^{(c)}(k + 1) + \omega_i^{(c)}(k + 1) \tag{39}$$

where  $\omega_i^{(c)}(k + 1)$  is drawn from a normal distribution:

$$\Omega \sim \mathcal{N}(0.00, 0.005). \tag{40}$$

The model parameter  $\phi$  captures the ease with which an improvement (or equivalently, a deterioration) in the outcome is observed by a patient. For example in a trial where the outcome of interest is, say, muscular strength or mobility [51], observability is high. In contrast, in a trial which examines the effects of different interventions on, say, bone mineral density [52], the patient is entirely or virtually entirely unable to gauge relevant changes. As expected from the theory presented in the previous sections and as we shall demonstrate empirically, this aspect of a clinical trial under consideration has important consequences on the benefit of sample selection proposed in this work. This is investigated in detail in the next section.

The changes in the beliefs of the treatment group are modelled in the identical manner to those of control group participants:

$$b_i^{(t)}(k + 1) = b_i^{(t)}(k) + \phi \times e_i^{(t)}(k + 1) + \omega_i^{(t)}(k + 1) \tag{41}$$

where  $\omega_i^{(t)}(k+1)$  is again drawn from a normal distribution:

$$\Omega_c \sim \mathcal{N}(0.00, 0.005) \quad (42)$$

## Experiment 1: Baseline performance

Using the model for generating longitudinal patient data described in *Experimental setup*, in this experiment we compared the proposed method with random sample selection in the context of a single sample size reduction step. In particular we simulated the reduction in the total number of participants of  $n = 70$  (i.e. approximately 39% of the total cohort size of 180) after  $k = 50$  time steps; please see Eqs (32)–(41). Following the removal time step, data accrual of the selected participants was discontinued. Note that in this experiment the data collected from the removed patients up to the time step  $k = 50$  was used in the final outcome analysis. The trial was simulated for the total number of  $k_{\max} = 100$  time steps after which the outcomes were used to infer the differential effects of the two simulated interventions. The final outcomes both in the case of the proposed method and random selection were analysed using the Bayesian method proposed in [16]. The observability parameter in this experiment was set to  $\phi = 0.01$ , please see Eq (39). Finally, to facilitate robust analysis, the experiment was repeated 100 times, resulting in different patient states, effects of the two interventions (treatment and control), as well as different removal choices for both selection procedures.

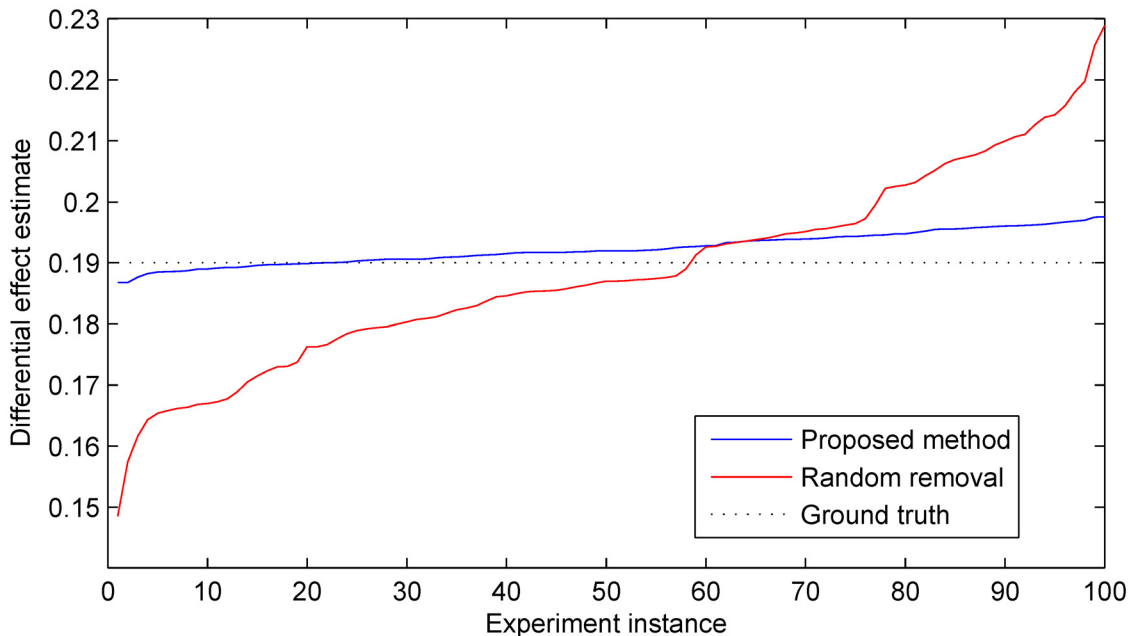
As expected, averaged over 100 instances of the simulated experiment there was no significant difference in the accuracy of the estimate of differential effects in the treatment group relative to the control group between the two sample size reduction methods. In both cases the average accuracy was correct to within 1% with a statistically insignificant error from the ground truth. However an examination of the corresponding estimate precisions readily reveals a different picture. In particular across different experiment instances the proposed method results in approximately 70% lower standard deviation of the target estimate; this is illustrated in Fig 4. This finding illustrates with clarity the argument we put forward in the preceding sections. By performing sample selection in a manner which takes into account all available information, the statistically best decision can be made about which specific samples can be removed so as to ensure that in any particular case the associated information loss is minimized. In the experiments which follow, we sought to examine in detail how different parameters of a trial affect the precision improvement achieved with the proposed method.

## Experiment 2: Effects of timing

As in the previous experiment, we used the data generation model from *Experimental setup* and compared the proposed method with random sample selection in the context of a single sample size reduction step. Our aim in this experiment was to investigate the effect that the timing of participant removal, that is, the cessation of the corresponding data accrual, has on the precision gained by using the proposed method. Hence we performed simulations for three different time steps at which sample size reduction was performed,  $k = 30, 50, 70$ . All the other parameters were kept constant—the number of participants removed was set to  $n = 50$  (i.e. approximately 28% of the total cohort size of 180), and the observability parameter which was set to  $\phi = 0.01$ . As before the final outcomes were analysed after  $k_{\max} = 100$  time steps using the method described in [16], and for each combination of parameters 100 simulations were performed.

Much like before and conforming to predictions from theory, we found no differential effect between the average estimates produced by the proposed method and the naïve random





**Fig 4. Estimates of the differential effect of treatment produced by the proposed method and the random selection-based baseline across 100 simulated trials.** The estimates are shown ordered in magnitude for easier visualization. In all trial instances the reduction in the total number of participants of  $n = 70$  (i.e. approximately 39% of the total cohort size of 180) was performed after  $k = 50$  time steps. Following the removal time step, data accrual from the selected participants was discontinued but the data collected from them before that was still used in the final outcome analysis. Both the proposed method and the random selection-based baseline exhibited no bias, with a statistically insignificant error and the accuracy of within 1% from the ground truth. However the proposed method demonstrated far superior precision as readily observed from the plot; also see Figs 5–7.

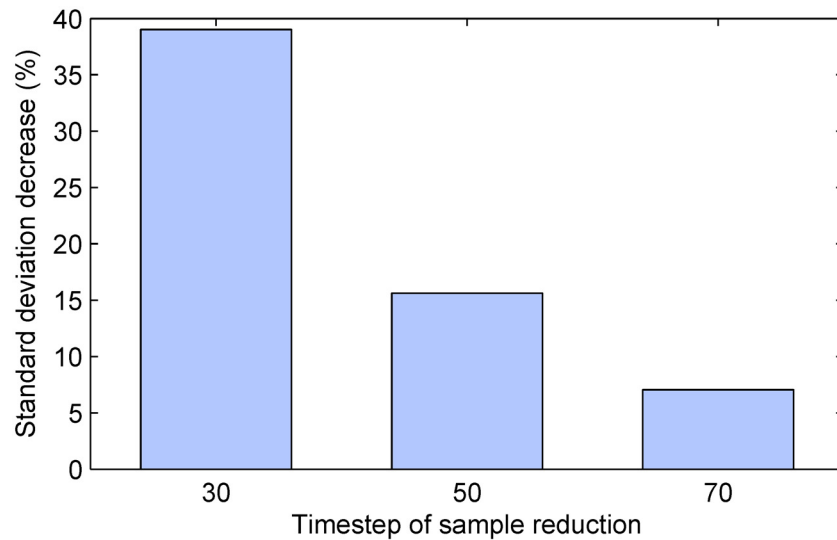
doi:10.1371/journal.pone.0131524.g004

selection-based approach—in both cases the average accuracy was correct to within 1% with a statistically insignificant error from the ground truth. On the other hand the difference in precision was again stark and offered interesting insight into the proposed method. Our findings are summarized by the plot in Fig 5 which shows the reduction in the standard deviation of the estimate of the differential effect of treatment, across different simulation instances. Firstly, observe that in all cases, that is, regardless of the time step at which sample size reduction was performed, the proposed method achieved superior results over the naïve alternative. Secondly, it is immediately apparent that the greatest benefit was found when sample size reduction was performed earliest (i.e. after  $k = 30$  time steps). In this case the standard deviation of the estimate was reduced extensively, by approximately 39%. Lower but still substantial reduction of approximately 16% was achieved when sample removal was performed half-way through the duration of the simulated trials. When the removal was delayed to after 70% of the duration of the trial, the standard deviation reduction was still significant at approximately 7%.

The observed behaviour of our algorithm is unsurprising and could be readily predicted from theory.

### Experiment 3: Effects of size adjustment

In this experiment our aim was to investigate the effect that the magnitude of sample size reduction (i.e. the number of participants for whom data accrual is discontinued) has on the precision gained by using the proposed method. The experiment was set up in the same manner as *Experiment 2* with the exception that now the time step at which sample size reduction



**Fig 5. The impact of sample size reduction timing on the precision of the estimate of the differential effect of treatment.** Plotted is the decrease in the standard deviation of the estimate across different simulated trial instances (in %, relative to the random selection-based baseline). In all cases the proposed method demonstrates a substantial improvement over the baseline. Greatest improvement is achieved for early sample size adjustment. Also see Fig 4.

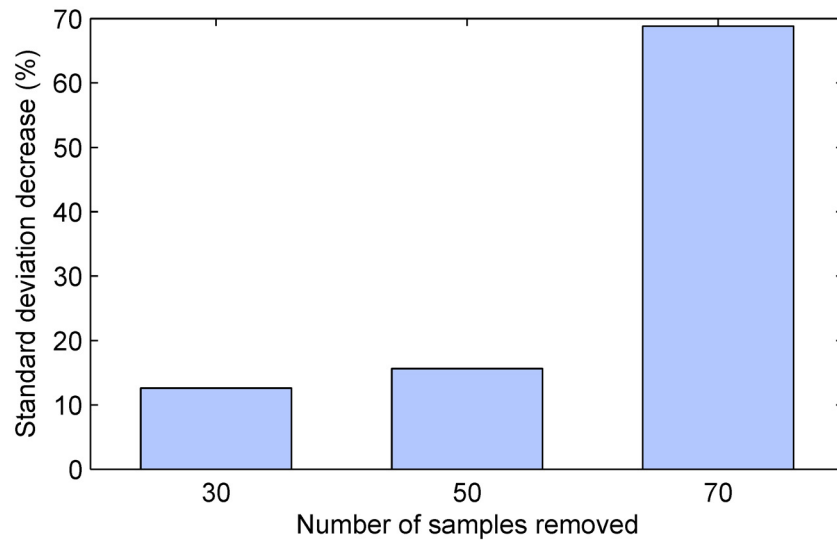
doi:10.1371/journal.pone.0131524.g005

was performed was set to  $k = 50$  and kept at this value in all simulation instances, while the number of participants who were removed from the continued accrual of data was treated as an independent variable,  $n = 30, 50, 70$ . As before 100 simulations were performed with each combination of parameters.

As in the previous experiments, the average accuracy of the estimate of the differential effects of the treatment intervention over control was the same regardless of the sample selection strategy employed, but the corresponding precisions differed significantly. Our findings are summarized by the plot in Fig 6. Again, it is worth starting with the most obvious observation, which is that in all cases the proposed methodology resulted in an improvement. The greatest improvement was observed in the case when the sample size reduction was most drastic, with the proposed algorithm achieving nearly 70% lower standard deviation than the random selection-based alternative. This is of course consistent with our theoretical argument and supportive of the propositions put forward in the preceding sections—the more the sample size is reduced, the greater becomes the advantage of using all available information to make the adjustment targeted to the specific cohort.

#### Experiment 4: Effects of perceptibility

In *Belief refinement* we highlighted the importance of  $\phi$ , the observability parameter of the model used to generate trial data. Recall that this parameter was introduced to model the ease with which a participant can gauge the effects of the administered treatment i.e. the participant's ability to observe an improvement or deterioration of the outcome of interest. The significance that this inherent nature of the studied phenomenon has in the context of the proposed sample size reduction methodology can be readily appreciated by observing that the proposed approach derives key power from the stratification of patients based on their responses to the auxiliary questionnaire (described in *Auxiliary data collection*). Consider the



**Fig 6. The impact of sample size reduction magnitude on the precision of the estimate of the differential effect of treatment.** Plotted is the decrease in the standard deviation of the estimate across different simulated trial instances (in %, relative to the random selection-based baseline). In all cases the proposed method demonstrates a substantial improvement over the baseline. Greatest improvement is achieved for the removal of a large number of participants. Also see [Fig 4](#).

doi:10.1371/journal.pone.0131524.g006

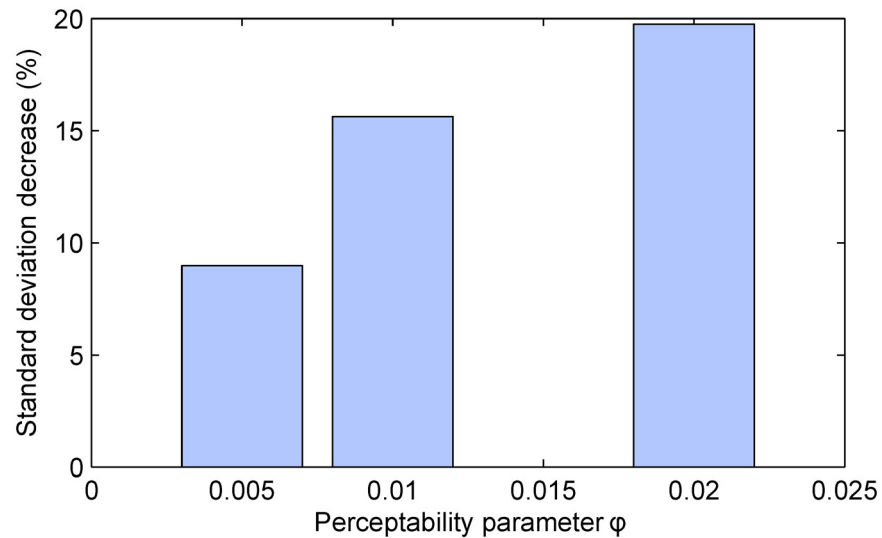
extreme case when perceptibility vanishes i.e. when  $\phi = 0$ . In this case the collected auxiliary data is effectively random and as such contributes no useful information to guide the selection of participants which are to be removed from the trial. More generally, it can be said that the lower the perceptibility of the outcome, the less there is to be gained from using the proposed algorithm as compared to simple random selection.

Following the design adopted in the previous experiments, we performed 100 simulated trials for each combination of parameter values, while keeping all of the trial variables fixed except for the observability parameter. In addition to the previously used value of  $\phi = 0.01$ , we examined the performance benefit of using our algorithm when observability was halved ( $\phi = 0.005$ ) and doubled ( $\phi = 0.02$ ). In all cases the number of removed participants was set to  $n = 50$  which was performed after  $k = 50$  time steps. The results are summarized in [Fig 7](#) and can be seen to match our theoretical expectations.

## Experiment 5: Sequential adjustment

In the final experiment we report, we adopted a different experimental design than in the previous experiments. Specifically, in the previous experiments sample size reduction was understood to result in the cessation of data accrual for the removed trial participants, but the data accrued before their removal was still used in the final analysis of the trial outcomes. In contrast, in this experiment all data for the removed participants, including the data accrued before removal, is discarded for the final analysis. This design captures those trial instances when the studied treatment exhibits cumulative effects which complicate the analysis of outcomes observed at different points in time.

Following the same practice as in the previous experiments, we used the data generation model described in *Experimental setup*. However unlike before in this experiment sample size



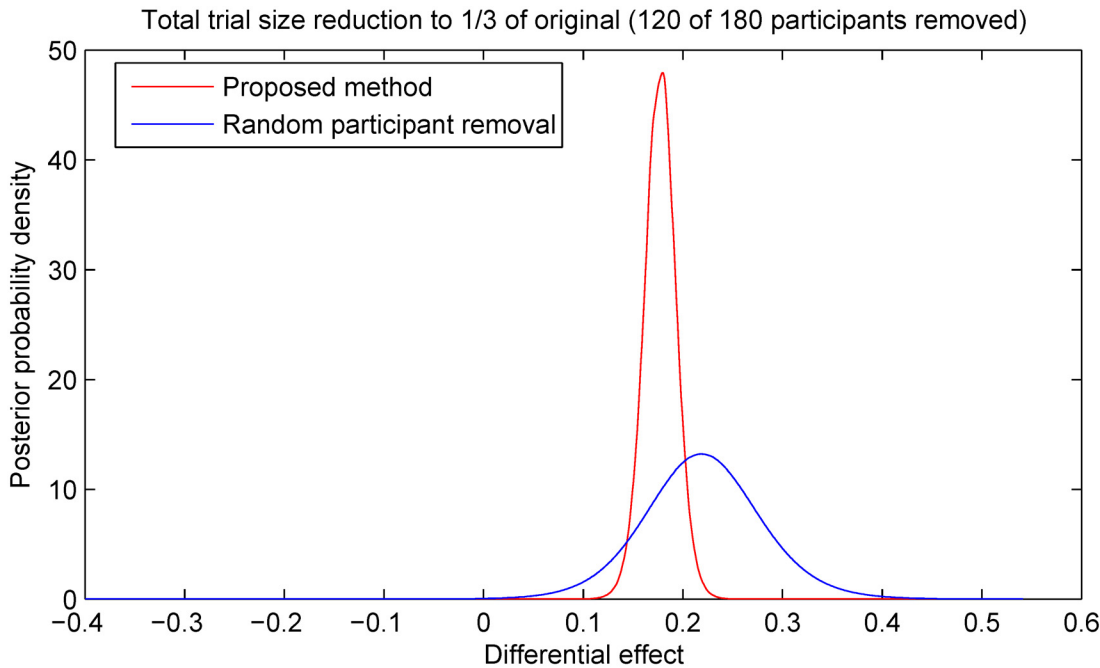
**Fig 7. The impact of outcome perceptibility on the precision of the estimate of the differential effect of treatment.** Plotted is the decrease in the standard deviation of the estimate across different simulated trial instances (in %, relative to the random selection-based baseline). In all cases the proposed method demonstrates a substantial improvement over the baseline. Greatest improvement is achieved for trials with high outcome perceptibility. Also see Fig 4.

doi:10.1371/journal.pone.0131524.g007

reduction was not performed at a single point in time. Rather a single participant was removed after each time step. Clearly this is not intended to replicate any realistic scenario which could be encountered in clinical practice. Though artificial in this sense, this experiment is useful as an analytical tool in the study of the behaviour of the proposed algorithm.

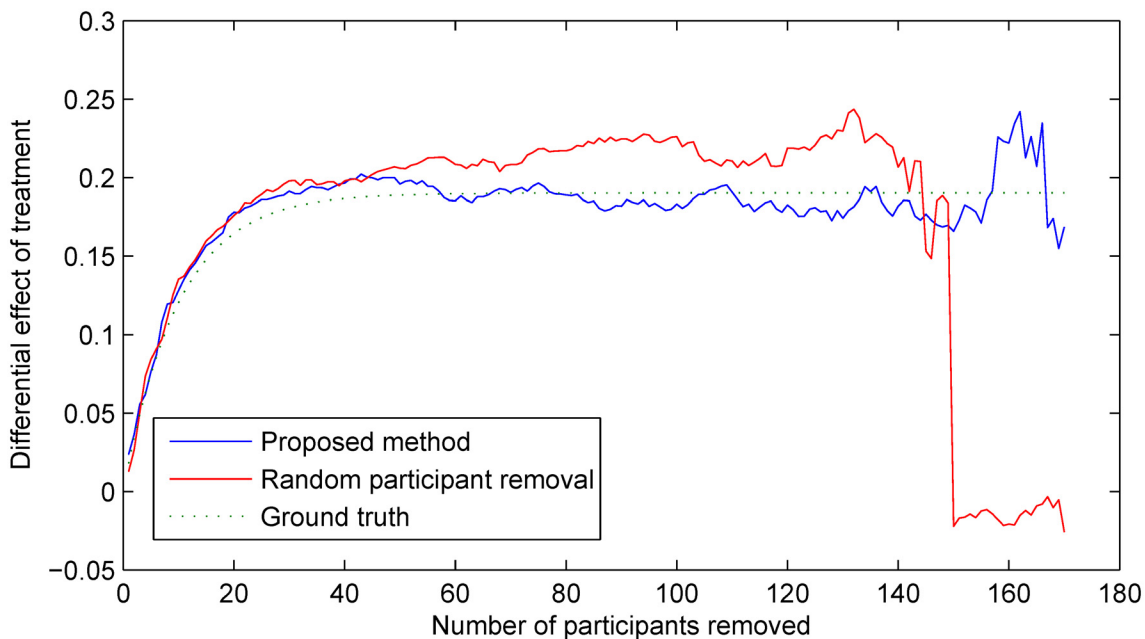
It is insightful to start with a look at the impact that the removal of a part of the cohort has on the estimate of the differential effect of the treatment intervention. A typical result is illustrated in Fig 8. This plot shows the posterior distributions of the differential effect of the treatment inferred after the removal of 120 individuals obtained using the proposed method (red line) and random selection (blue line). In both cases the posteriors were computed as described in [16]. The most notable difference between the two distributions is in their spreads i.e. the associated uncertainties. Specifically, the proposed method results in a much more peaked posterior, that is, a much more confident estimate. In comparison, the posterior obtained using random selection is much broader, admitting a lower degree of certainty associated with the corresponding estimate.

The accuracy of the two methods is better assessed by observing their behaviour over time. The plot in Fig 9 shows the *maximum a posteriori* estimates of the differential effect of treatment obtained using the two methods during the course of the trial. Also shown is the ‘ground truth’, that is, the actual differential effect which we can compute exactly from the setup of the experiment (see *Experimental setup*). In the early stages of the trial, while the magnitude of the accumulated effect is small and the number of participants large, the two estimates are virtually indistinguishable, and they follow the ground truth plot closely. As expected, as the number of participants removed increases both estimates start to exhibit greater stochastic perturbations. However, both the accuracy (that is, the closeness to the ground truth) and the precision (that is, the magnitude of stochastic variability) of the proposed method can be seen to show superior performance—its *maximum a posteriori* estimate follows the ground truth more closely



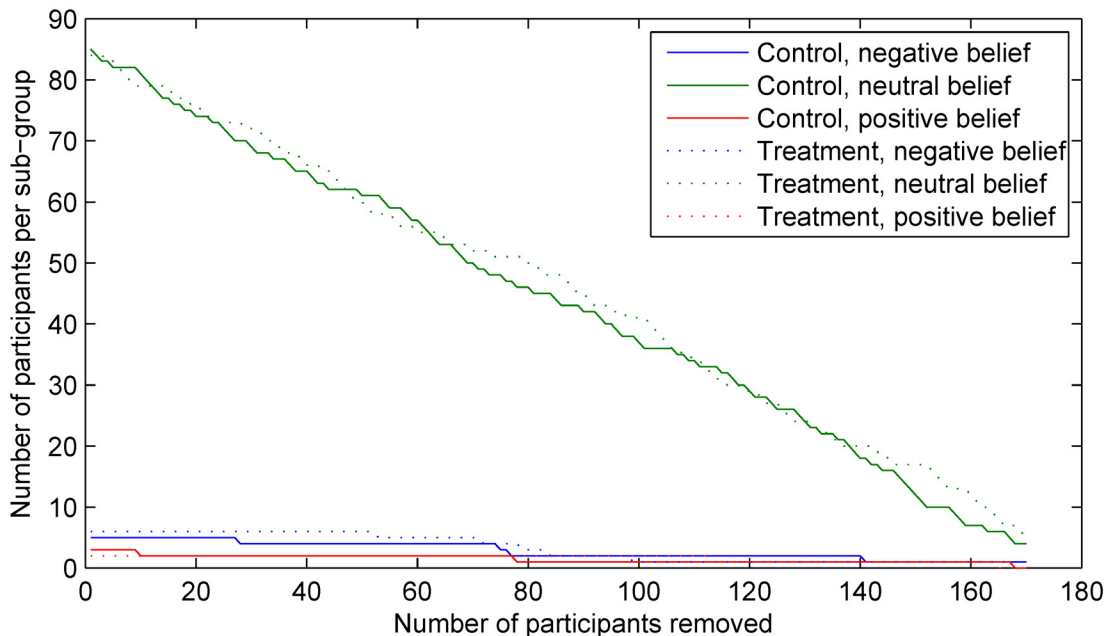
**Fig 8. Posterior distributions of the differential effect of treatment after the removal of 120 participants i.e. after 120 time steps; see Eqs (32)–(42).** The method introduced in this paper results in a much more certain estimate of the differential, as witnessed by the highly peaked distribution, than does random selection.

doi:10.1371/journal.pone.0131524.g008



**Fig 9. The maximum a posteriori estimates of the differential effect of treatment during the course of the trial as an increasing number of participants is removed.** The blue line shows the estimates obtained using the proposed method, while the red line shows the estimates obtained using random selection. Also shown is the ‘ground truth’ (green dotted line), that is, the actual differential effect computed exactly from the setup of the experiment (see *Experimental setup*).

doi:10.1371/journal.pone.0131524.g009



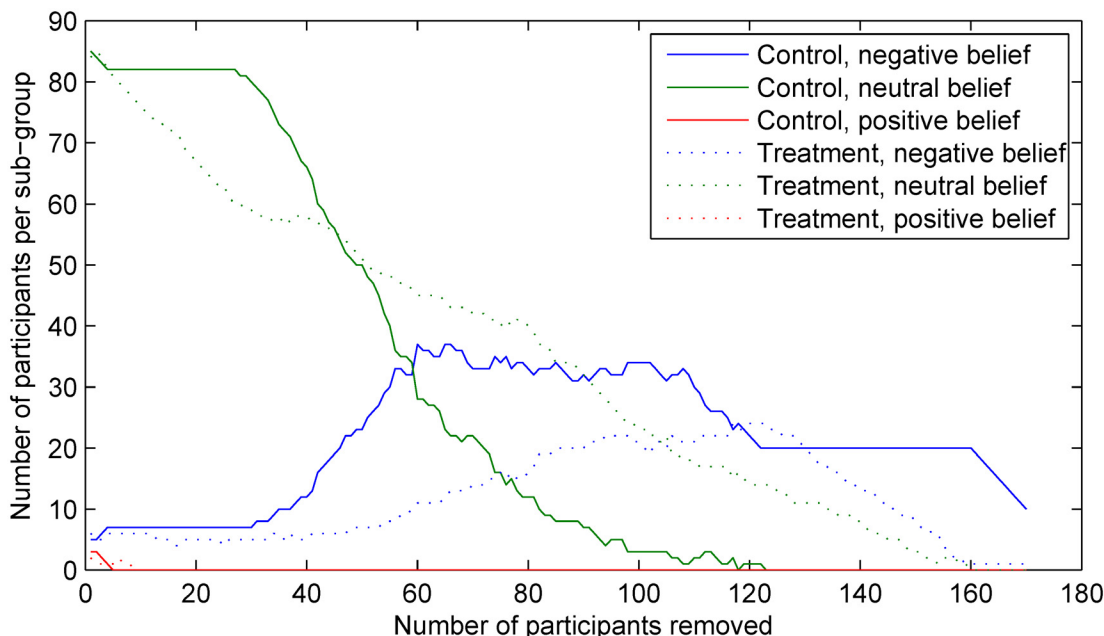
**Fig 10. The changes in the sample sizes within each of the six participant sub-groups (using the stratification described in *Matching sub-groups outcome model* and adopted from [16]) observed in our experiment using random selection based participant removal.** Note the predictable outcome of the method, which results in a linear decrease of sample size for all sub-groups, and the contrasting behaviour of our approach in Fig 11.

doi:10.1371/journal.pone.0131524.g010

and fluctuates less than the estimate obtained when random selection is employed instead. It is also important to observe the rapid degradation of performance of the random selection method as the number of remaining participants becomes small, which is not seen in the proposed method. This too can be expected from the theoretical argument put forward in *Sub-group selection*—the statistically optimal choice of the sub-group from which participants are removed ensures that the posterior is not highly dependent on a small number of samples which would make it highly sensitive to the change in sample size.

Lastly, it is interesting to observe the differences between the changes in the sample sizes within each sub-group using the two approaches. This is illustrated using the plots in Figs 10 and 11. As expected, when random participant removal is employed, the sizes of all sub-groups decrease roughly linearly (save for stochastic variability), as shown in Fig 10. In contrast, the sub-group size changes effected by the proposed method show more complex structure, governed by the specific values of the belief and effect variables in our experiment, as shown in Fig 11. It is particularly interesting to note that the size changes are not only non-linear, but also non-monotonic. For example, the relative size of the control sub-group which includes individuals which correctly identified their group assignment (i.e. the sub-group  $G_{C-}$ ) begins to increase notably after the removal of 30 participants and starts to decrease only after the removal of further 78 participants.

**Summary and conclusions.** In this paper we introduced a novel method for clinical trial adaptation. Our focus was on adaptation by amending sample size. In contrast to all previous work in this area, the problem we considered was not *when* sample size should be adjusted but rather *which* particular individuals should be removed from the trial once the decision of sample reduction is made. Thus, our method is not an alternative to the current state-of-the-art,



**Fig 11. The changes in the sample sizes within each of the six participant sub-groups (using the stratification described in *Matching sub-groups outcome model* and adopted from [16]) observed in our experiment using the proposed method.** Compare with Fig 10 and note the data-specific behaviour of our approach.

doi:10.1371/journal.pone.0131524.g011

but rather a complementary element of the same framework. Our approach is based on the adopted stratification recently proposed for the analysis of trial outcomes in the presence of imperfect blinding. This stratification is based on the trial participants’ responses to a generic auxiliary questionnaire that allows each participant to express belief concerning his/her intervention assignment (treatment or control). Extensive experiments on a simulated trial were used to illustrate the effectiveness of our method.

### Acknowledgments

I would like to express my gratitude to Prof. Chiappelli and the anonymous reviewers whose many questions, comments, and suggestions greatly helped me identify the best manner to present the key contributions of my work.

### Author Contributions

Conceived and designed the experiments: OA. Performed the experiments: OA. Analyzed the data: OA. Contributed reagents/materials/analysis tools: OA. Wrote the paper: OA.

### References

1. Meinert CL. *Clinical Trials: Design, Conduct, and Analysis*. 3rd ed. New York, New York: Oxford University Press; 1986.
2. Piantadosi S. *Clinical Trials: A Methodologic Perspective*. 3rd ed. Hoboken, New Jersey: Wiley; 1997.
3. Friedman L, Furberg C, DeMets D. *Fundamentals of Clinical Trials*. 3rd ed. New York, New York: Springer-Verlag; 1998.



4. Penston J. Large-scale randomised trials—a misguided approach to clinical research. *Med Hypotheses*. 2005; 64(3):651–657. doi: [10.1016/j.mehy.2004.09.006](https://doi.org/10.1016/j.mehy.2004.09.006) PMID: [15617882](https://pubmed.ncbi.nlm.nih.gov/15617882/)
5. Fisher LD. Self-designing clinical trials. *Stat Med*. 1998; 17:1551–1562. doi: [10.1002/\(SICI\)1097-0258\(19980730\)17:14%3C1551::AID-SIM868%3E3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0258(19980730)17:14%3C1551::AID-SIM868%3E3.0.CO;2-E) PMID: [9699229](https://pubmed.ncbi.nlm.nih.gov/9699229/)
6. U S Department of Health and Human Services. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics. Food and Drug Administration Draft Guidance. 2010 Feb;.
7. Hung HMJ, Wang SJ, O'Neill RT. Methodological issues with adaptation of clinical trial design. *Pharmaceut Statist*. 2006; 5:99–107. doi: [10.1002/pst.219](https://doi.org/10.1002/pst.219)
8. Orloff J, Douglas F, Pinheiro J, Levinson S, Branson M, Chaturvedi P, et al. The future of drug development: advancing clinical trial design. *Nat Rev Drug Discov*. 2009; 8(12):949–957. PMID: [19816458](https://pubmed.ncbi.nlm.nih.gov/19816458/)
9. Berry DA. Adaptive clinical trials: the promise and the caution. *J Clin Oncol*. 2011; 29(6):606–609. doi: [10.1200/JCO.2010.32.2685](https://doi.org/10.1200/JCO.2010.32.2685) PMID: [21172875](https://pubmed.ncbi.nlm.nih.gov/21172875/)
10. Chow SC, Chan M. *Adaptive Design Methods in Clinical Trials*. Chapman & Hall; 2011.
11. Cui L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999; 55:321–324.
12. Nissen SE. ADAPT: The Wrong Way to Stop a Clinical Trial. *PLoS Clin Trials*. 2006; 1(7):e35. doi: [10.1371/journal.pctr.0010035](https://doi.org/10.1371/journal.pctr.0010035) PMID: [17111045](https://pubmed.ncbi.nlm.nih.gov/17111045/)
13. Lang T. Adaptive trial design: could we use this approach to improve clinical trials in the field of global health? *Am J Trop Med Hyg*. 2011; 85(6):967–970.
14. Arandjelović O. Sample-targeted clinical trial adaptation. *Proc Conf AAAI Artif Intell*. 2015; 3:1693–1699.
15. Arandjelović O. Assessing blinding in clinical trials. *Adv Neural Inf Proc Sys*. 2012; 25:530–538.
16. Arandjelović O. A new framework for interpreting the outcomes of imperfectly blinded controlled clinical trials. *PLOS ONE*. 2012; 7(12):e48984. doi: [10.1371/journal.pone.0048984](https://doi.org/10.1371/journal.pone.0048984) PMID: [23236350](https://pubmed.ncbi.nlm.nih.gov/23236350/)
17. Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials*. 2006; 3(4):360–365. PMID: [17060210](https://pubmed.ncbi.nlm.nih.gov/17060210/)
18. Berry SM, Carlin BP, Lee JJ, Müller P. *Bayesian adaptive methods for clinical trials*. CRC Press; 2010.
19. Kadane JB, editor. *Bayesian methods and ethics in a clinical trial design*. John Wiley & Sons; 2011.
20. Lawson AB. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC Press; 2013.
21. Oniško A, Druzdzel MJ. Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems. *Artif Intell Med*. 2013; 57(3):197–206. doi: [10.1016/j.artmed.2013.01.004](https://doi.org/10.1016/j.artmed.2013.01.004) PMID: [23466438](https://pubmed.ncbi.nlm.nih.gov/23466438/)
22. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. Chapman & Hall/CRC; 2014.
23. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons; 2004.
24. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov*. 2006; 5:27–36. doi: [10.1038/nrd1927](https://doi.org/10.1038/nrd1927) PMID: [16485344](https://pubmed.ncbi.nlm.nih.gov/16485344/)
25. Chen MH, Dey DK, Müller P, Sun D, Ye K. Bayesian clinical trials. In: *Frontiers of Statistical Decision Making and Bayesian Analysis*. Springer; 2010. p. 257–284.
26. Lee JJ, Chu CT. Bayesian clinical trials in action. *Stat Med*. 2012; 31(25):2955–2972. doi: [10.1002/sim.5404](https://doi.org/10.1002/sim.5404) PMID: [22711340](https://pubmed.ncbi.nlm.nih.gov/22711340/)
27. Berry DA. Adaptive clinical trials in oncology. *Nat Rev Clin Oncol*. 2012; 9(4):199–207. doi: [10.1038/nrclinonc.2011.165](https://doi.org/10.1038/nrclinonc.2011.165)
28. Yin G. *Clinical trial design: Bayesian and frequentist adaptive methods*. John Wiley & Sons; 2013.
29. Broglio KR, Connor JT, Berry SM. Not too big, not too small: A Goldilocks approach to sample size selection. *J Biopharm Stat Biopharm Stat*. 2014; 24(3):685–705. doi: [10.1080/10543406.2014.888569](https://doi.org/10.1080/10543406.2014.888569)
30. Gao P, Ware JH, Mehta C. Sample size re-estimation for adaptive sequential design in clinical trials. *J Biopharm Stat*. 2008; 18:1184–1196. doi: [10.1080/10543400802369053](https://doi.org/10.1080/10543400802369053) PMID: [18991116](https://pubmed.ncbi.nlm.nih.gov/18991116/)
31. Leon AC. Implications of clinical trial design on sample size requirements. *Schizophr Bull*. 2008; 34(4):664–669. doi: [10.1093/schbul/sbn035](https://doi.org/10.1093/schbul/sbn035) PMID: [18469326](https://pubmed.ncbi.nlm.nih.gov/18469326/)
32. Zang Y, Lee JJ. Adaptive clinical trial designs in oncology. *Chin Clin Oncol*. 2014; 3(4):49.
33. Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. *Stat Med*. 2003; 22(6):971–993. doi: [10.1002/sim.1457](https://doi.org/10.1002/sim.1457) PMID: [12627413](https://pubmed.ncbi.nlm.nih.gov/12627413/)

34. James KE, Bloch DA, Lee KK, Kraemer HC, Fuller RK. An index for assessing blindness in a multi-centre clinical trial: disulfiram for alcohol cessation—a VA cooperative study. *Stat Med*. 1996; 15(13):1421–1434. doi: [10.1002/\(SICI\)1097-0258\(19960715\)15:13%3C1421::AID-SIM266%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-0258(19960715)15:13%3C1421::AID-SIM266%3E3.0.CO;2-H) PMID: [8841652](https://pubmed.ncbi.nlm.nih.gov/8841652/)
35. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. *Contemp Clin Trials*. 2004; 25(2): 143–156. doi: [10.1016/j.cct.2003.10.016](https://doi.org/10.1016/j.cct.2003.10.016)
36. Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol*. 2007; 36(3):654–663. doi: [10.1093/ije/dym020](https://doi.org/10.1093/ije/dym020) PMID: [17440024](https://pubmed.ncbi.nlm.nih.gov/17440024/)
37. Kolahi J, Bang H, Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. *Community Dent Oral Epidemiol*. 2009; 37(6):477–484. doi: [10.1111/j.1600-0528.2009.00494.x](https://doi.org/10.1111/j.1600-0528.2009.00494.x) PMID: [19758415](https://pubmed.ncbi.nlm.nih.gov/19758415/)
38. Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol*. 2007; 36(3):664–665. PMID: [17675306](https://pubmed.ncbi.nlm.nih.gov/17675306/)
39. Hemilä H. Assessment of blinding may be inappropriate after the trial. *Contemp Clin Trials*. 2005; 26(4):512–514. doi: [10.1016/j.cct.2005.04.004](https://doi.org/10.1016/j.cct.2005.04.004) PMID: [15951244](https://pubmed.ncbi.nlm.nih.gov/15951244/)
40. Henneicke-von Zepelin HH. Letter to the Editor. *Contemp Clin Trials*. 2005; 26(4):512. doi: [10.1016/j.cct.2005.04.004](https://doi.org/10.1016/j.cct.2005.04.004) PMID: [15936250](https://pubmed.ncbi.nlm.nih.gov/15936250/)
41. Wickens CD. *Engineering Psychology and Human Performance*. 2nd ed. New York: Harper Collins Publishers Inc.; 1992.
42. Berger V. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. Hoboken, New Jersey: Wiley; 2005.
43. Biswas S, Liu DD, Lee JJ, Berry DA. Bayesian clinical trials at the University of Texas MD Anderson cancer center. *Clin Trials*. 2009; 6(3):205–216. doi: [10.1177/1740774509104992](https://doi.org/10.1177/1740774509104992) PMID: [19528130](https://pubmed.ncbi.nlm.nih.gov/19528130/)
44. Nelson NJ. Adaptive Clinical Trial Design: Has Its Time Come? *JNCI J Natl Cancer Inst*. 2010; 102(16):1217–1218.
45. Meurer WJ, Lewis RJ, Tagle D, Fetters MD, Legocki L, Berry S, et al. An overview of the adaptive designs accelerating promising trials into treatments (ADAPT-IT) project. *Ann Emerg Med*. 2012; 60(4):451–457. doi: [10.1016/j.annemergmed.2012.01.020](https://doi.org/10.1016/j.annemergmed.2012.01.020) PMID: [22424650](https://pubmed.ncbi.nlm.nih.gov/22424650/)
46. Kullback RA, Leibler S. On Information and Sufficiency. *Ann Math Stat*. 1951; 22(1):79–86. doi: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
47. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull Calcutta Math Soc*. 1943; 35:99–109.
48. Hellinger E. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *J Reine Angew Math*. 1909; 136:210–271.
49. Gilks WR. *Markov Chain Monte Carlo in Practice*. 1st ed. Chapman and Hall/CRC; 1995.
50. Aitchison J, Brown JAC. *The Lognormal Distribution*. Cambridge University Press; 1957.
51. Ly LP, Jimenez M, Zhuang TN, Celermajer DS, Conway AJ, Handelsman DJ. A double-blind, placebo-controlled, randomized clinical trial of transdermal dihydrotestosterone gel on muscular strength, mobility, and quality of life in older men with partial androgen deficiency. *J Clin Endocrinol Metab*. 2001; 86(9):4078–4088. doi: [10.1210/jcem.86.9.7821](https://doi.org/10.1210/jcem.86.9.7821) PMID: [11549629](https://pubmed.ncbi.nlm.nih.gov/11549629/)
52. Cohen SB, Dore RK, Lane NE, Ory PA, Peterfy CG, Sharp JT, et al. Denosumab treatment effects on structural damage, bone mineral density, and bone turnover in rheumatoid arthritis: A twelve-month, multicenter, randomized, double-blind, placebo-controlled, phase II clinical trial. *Arthritis Rheum*. 2008; 58(5):1299–1309. doi: [10.1002/art.23417](https://doi.org/10.1002/art.23417) PMID: [18438830](https://pubmed.ncbi.nlm.nih.gov/18438830/)