

RESEARCH

Open Access



# Strategies for informed sample size reduction in adaptive controlled clinical trials

Ognjen Arandjelović

## Abstract

Clinical trial adaptation refers to any adjustment of the trial protocol after the onset of the trial. The main goal is to make the process of introducing new medical interventions to patients more efficient. 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. Our key idea is to select the sample subset for removal in a manner which minimizes the associated loss of information. We formalize this notion and describe three algorithms which approach the problem in different ways, respectively, using (i) repeated random draws, (ii) a genetic algorithm, and (iii) what we term pair-wise sample compatibilities. Experiments on simulated data demonstrate the effectiveness of all three approaches, with a consistently superior performance exhibited by the pair-wise sample compatibilities-based method.

**Keywords:** RCT, Bayesian, Information

## 1 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 [5–7].

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 [8]. 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 [9–11].

### 1.1 Contrast with previous work

Before introducing the proposed method in detail, it is worthwhile emphasising 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. Most existing work on trial adaptation by sample size adjustment adopts the frequentist paradigm. These methods follow the same pattern: a particular null hypothesis is formulated which

Correspondence: ognjen.arandjelovic@gmail.com  
School of Computer Science, University of St Andrews, North Haugh, St Andrews, UK

is then rejected or accepted using a suitable statistic and the desired confidence requirement (a good review is provided by Jennison and Turnbull [12]). 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. All 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 the any of the previously proposed methods but rather as a complementary element of the same framework.

## 2 Targeted removal sample selection

Previous research on clinical trial adaptation by sample size reduction has universally focused on the question of when such reduction should be performed. In contrast, no consideration has been given to the question of which specific samples should be removed from the trial and which should be retained when sample size reduction is performed. Indeed the current practice is to remove a random subset of samples. More formally, if the sample size before adaptation is  $n$ , the samples are  $\{x_1, x_2, \dots, x_n\}$ , and  $m$  of them are to be removed, the first sample to be removed  $x_{r_1}$  is selected by drawing  $r_1$  from the set  $\{1, 2, \dots, n\}$  with the uniform probability of  $1/n$ . Similarly, the second sample  $x_{r_2}$  is selected by drawing  $r_2$  from  $\{1, 2, \dots, n\} \setminus \{r_1\}$  with the uniform probability of  $1/(n-1)$ . This proceeds until all  $m$  random samples are selected, in each step selecting the  $i$ -th sample  $x_{r_i}$  by drawing  $r_i$  from  $\{1, 2, \dots, n\} \setminus \{r_1, \dots, r_{i-1}\}$  with the uniform probability of  $1/(n-i+1)$ . The work described in this paper is motivated by the observation that in general this strategy is not optimal. To see why this is the case, let us first observe that the described selection procedure is inherently uninformed in the sense that all samples are treated in exactly the same manner. What is being ignored is the fact that trial adaptation, by its very nature, takes place some time after the commencement of the trial. During this time differentiation between samples take place by virtue of their (in general) different responses to the interventions administered in the trial. This differentiation can be used to make the process of sample selection informed. In the remainder of this section, we examine different means by which this can be achieved. Specifically, we first formalize the aim of informed sample selection when a specific sample size reduction is required, and then follow this up by

a description of three different approaches which address the said aim.

### 2.1 Information preservation criterion

As explained previously, our goal is to perform the selection of samples which are to be removed in a manner which minimizes the amount of information loss, i.e. preserves the amount of information retained. Before this problem can be tackled, it is necessary to ascertain what the relevant information is.

Recall that our framework comprises two sets of samples. In keeping with the terminology of clinical trials these are the set of  $n_t$  'treatment' samples which are being administered the treatment of interest with the corresponding trial observations  $D_t = \{x_1, x_2, \dots, x_{n_t}\}$ , and the set of  $n_c$  'control' samples which are being administered an alternative control intervention with the corresponding trial observations  $D_c = \{x_{n_t+1}, x_{n_t+2}, \dots, x_n\}$ , where  $n_t + n_c = n$  is the total number of samples. Let the inherent statistics of the two sets of data be described by respectively the random variables  $X_t$  and  $X_c$ , which are governed by the underlying probability density functions  $p_t$  and  $p_c$  parameterized by the sets of latent variables  $\Theta_t$  and  $\Theta_c$  so that we can write  $p_t \equiv p_t(x; \Theta_t)$  and  $p_c \equiv p_c(x; \Theta_c)$ . Adopting the Bayesian methodology for inference, the observed trial data can be used to estimate the corresponding probability density functions as follows:

$$\begin{aligned}\hat{p}_t &= \int_{\Theta_t} p(D_t; \Theta_t) p(\Theta_t) d\Theta_t, \\ \hat{p}_c &= \int_{\Theta_c} p(D_c; \Theta_c) p(\Theta_c) d\Theta_c.\end{aligned}\quad (1)$$

Similarly, after the removal of samples from  $D_t$  and  $D_c$ , resulting in truncated sets  $D'_t$  and  $D'_c$ :

$$\begin{aligned}\hat{p}'_t &= \int_{\Theta_t} p(D'_t; \Theta_t) p(\Theta_t) d\Theta_t, \\ \hat{p}'_c &= \int_{\Theta_c} p(D'_c; \Theta_c) p(\Theta_c) d\Theta_c.\end{aligned}\quad (2)$$

Thus, it may seem reasonable to attempt to select samples for removal in a way which minimizes the difference between the estimates of  $p_t$  and  $p_c$  before and after the removal of said samples:

$$D'_t = \arg \min_{\hat{p}'_t} \mathcal{D}(\hat{p}_t, \hat{p}'_t), \quad (3)$$

where  $\mathcal{D}$  may be a divergence measure such as the Kullback-Leibler divergence or a distance such as the Bhattacharyya or the Hellinger distance.

Rather, here we argue that information and therefore the loss of information should be understood in the context of and relative to the ultimate aim of the trial. Invariably this aim is to estimate the probability that

the treatment of interest is more effective than the alternative, control treatment:

$$\rho(p_t(x; \Theta_t), p_c(y; \Theta_c)) = \int_{-\infty}^{\infty} \int_{-\infty}^x p_t(x; \Theta_t) \times p_c(y; \Theta_c) dy dx. \quad (4)$$

Using the Bayesian methodology as before allows us to write:

$$\begin{aligned} \rho^* &= \int_{\Theta_c} \int_{\Theta_t} \underbrace{\rho(p_t(x; \Theta_t), p_c(x; \Theta_c))}_{\text{Target probability}} \\ &\quad \times \underbrace{p(\Theta_t | D_t) p(\Theta_c | D_c)}_{\text{Parameter pdf's conditioned on observations}} \\ &\quad \times \underbrace{p(\Theta_t) p(\Theta_c)}_{\text{Parameter priors}} \times d\Theta_c d\Theta_t. \end{aligned} \quad (5)$$

Let the trial observation data in two matching subgroups be drawn from the random variables  $X_c$  and  $X_t$ , which are appropriately modelled using normal distributions [13]:

$$\begin{aligned} X_t &\sim \frac{1}{\sigma_t \sqrt{2\pi}} \exp \left\{ -\frac{(x - m_t)^2}{2\sigma_t^2} \right\}, \text{ and} \\ X_c &\sim \frac{1}{\sigma_c \sqrt{2\pi}} \exp \left\{ -\frac{(x - m_c)^2}{2\sigma_c^2} \right\}. \end{aligned} \quad (6)$$

Using uninformed priors on  $m_c$ ,  $m_t$ ,  $\sigma_c$ , and  $\sigma_t$  (just as in [14]) leads to the following expression:

$$\begin{aligned} \rho^* &\propto \int_{m_t, m_c \in [-\infty, \infty]} \int_0^{\infty} \int_0^x \sigma_t^{-1} e^{-\frac{(x-m_t)^2}{2\sigma_t^2}} \sigma_t^{-N} e^{-\frac{\sum_{i=1}^{n_t} (x_i^{(t)} - m_t)^2}{2\sigma_t^2}} \times \\ &\quad \sigma_c^{-1} e^{-\frac{(x-m_c)^2}{2\sigma_c^2}} \sigma_c^{-N} e^{-\frac{\sum_{i=1}^{n_c} (x_i^{(c)} - m_c)^2}{2\sigma_c^2}} dx dm_t d\sigma_t dy dm_c d\sigma_c \end{aligned} \quad (7)$$

$$= \int_0^{\infty} \int_0^x I_t(x) I_c(y) dy dx = \int_0^{\infty} I_t(x) \int_0^x I_c(y) dy dx, \quad (8)$$

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

$$\begin{aligned} I(x) &= \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma} \exp \left\{ -\frac{(x - m)^2}{2\sigma^2} \right\} \\ &\quad \times \frac{1}{\sigma^n} \exp \left\{ -\frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2} \right\} dm d\sigma, \end{aligned} \quad (9)$$

$\{x_i\}$  and  $\hat{n}$  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 logarithmically 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 [14] so that:

$$(x - m)^2 + \sum_{i=1}^{\hat{n}} (x_i - m)^2 \equiv (am + b)^2 + c, \quad (10)$$

which leads to the following simplification of Eq. (8):

$$\begin{aligned} I &\propto \int_0^{\infty} \frac{1}{\sigma^{\hat{n}+2}} \exp \left\{ -\frac{c}{2\sigma^2} \right\} \sigma d\sigma = \int_0^{\infty} \frac{1}{\sigma^{\hat{n}+1}} \\ &\quad \exp \left\{ -\frac{c}{2\sigma^2} \right\} d\sigma, \end{aligned} \quad (11)$$

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

$$\begin{aligned} c &= x^2 + \sum_{i=1}^{\hat{n}} x_i^2 - \frac{\left( x + \sum_{i=1}^{\hat{n}} x_i \right)^2}{\hat{n} + 1} \\ &= \frac{(\hat{n} + 1) \left( x^2 + \sum_{i=1}^{\hat{n}} x_i^2 \right) - \left( x + \sum_{i=1}^{\hat{n}} x_i \right)^2}{\hat{n} + 1}. \end{aligned} \quad (12)$$

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

$$\text{Gamma}(z; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp\{-\beta/z\}. \quad (13)$$

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:

$$I \propto c^{-\frac{n-1}{2}} = \left[ \frac{(n+1) \left( x^2 + \sum_{i=1}^n x_i^2 \right) - \left( x + \sum_{i=1}^n x_i \right)^2}{n+1} \right]^{-\frac{n-1}{2}}. \quad (14)$$

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. (14) back into Eq. 8 gives the following expression for the distance function:

$$\rho^* = \int_0^{\infty} \int_0^x p_t(x) p_c(y) dy dx \propto \int_0^{\infty} \int_0^x c_t^{-\frac{n_t-1}{2}} c_c^{-\frac{n_c-1}{2}} dy dx \quad (15)$$

$$= \int_0^{\infty} c_t^{-\frac{n_t-1}{2}} \int_0^x c_c^{-\frac{n_c-1}{2}} dy dx. \quad (16)$$

The double integration can be performed numerically and hereafter our goal is to minimize the effects of sample removal on this value. Specifically, if we denote by  $\rho^*(D)$  the estimate of Eq. (16) after the removal of samples comprising the set  $D$ , we aim to minimize:

$$\Delta(D) = \left| \ln \frac{\rho^*(\emptyset)}{\rho^*(D)} \right| = |\ln \rho^*(\emptyset) - \ln \rho^*(D)|. \quad (17)$$

## 2.2 Repeated random draws based sample selection

The simplest way of informing sample size reduction can be achieved by performing repeated uninformed random draws of sample sets of the desired size, followed by the post hoc selection of the set corresponding to least information loss, as per Eq. (17). Specifically, we generate a hypothesis in the form of a set of samples considered for removal one by one. In each iteration  $k + 1$  (where  $k = 0, \dots, n_r - 1$ ), the probability of drawing  $x_i \in D \setminus D_s^{(k)}$  is:

$$p_{k+1}(x_i) = \frac{1}{n - k}, \quad (18)$$

where  $D_s^{(k)}$  is the subset of samples selected by the preceding  $k$  draws. This drawing rule ensures that in each iteration all samples which have not yet been selected have the same probability of being drawn next. After  $n_r$  iterations the entire hypothesis has been generated and its quality can be assessed by the associated information loss given by Eq. (17). If  $n_h$  hypotheses (i.e. sample subsets) are drawn, the sample subset corresponding to the hypothesis associated with the least loss is chosen as the best subset for removal from the subsequent stages of the trial.

## 2.3 An evolutionary approach

The selection method described in the preceding section is informed in the sense that it makes use of the available trial data to select preferentially one subset of samples over another. Nevertheless, this information is exploited rather inefficiently because the repeated draws themselves are entirely random and independent of one another—discriminative information is applied post hoc. In this section, we describe a method in which the said information is applied proactively, that is, in a manner whereby previously evaluated solutions direct the choice of future hypotheses.

We have already noted that if the optimal solution to the sample selection problem was to be guaranteed it would be necessary to evaluate the fitness of every possible sample subset of the desired size, which is in general unfeasible in practice. The key feature of the problem at hand is that it does not possess optimal substructure, that is, it cannot be solved (optimally) by an efficient combination of optimal solutions of its smaller subproblems. Put differently, the selection of the best  $(n_1 + n_2)$ -sized subset of samples to be removed cannot be efficiently constructed from the knowledge of the best selections of  $n_1$ -sized and  $n_2$ -sized subsets. The method we describe in this section is based on the idea that notwithstanding this inherent computational hardness of the problem, it is reasonable to expect that if good  $n_1$ -sized and  $n_2$ -sized solutions are known, a  $(n_1 + n_2)$ -sized solution better than one which would on average be obtained by a random draw can be hypothesized. We put this idea into practice by employing a genetic algorithm. For the benefit

of the reader, we briefly review the key ideas underlying genetic algorithms next, and then follow up by a specific implementation engineered by us for the sample selection problem at hand.

### 2.3.1 Genetic algorithms

Genetic algorithms belong to a class of adaptive heuristic search algorithms which are inspired by the evolutionary concepts from genetics and the theory of natural selection. In particular, they make use of concepts such as heritable characteristics and fitness-based selection and have been applied with success in a number of diverse domains [15–17].

The key elements of a genetic algorithm are ‘genes’, ‘chromosomes’ (in this context also called and synonymous with ‘individuals’), and ‘population’, whose functions bear resemblance to their biological namesakes. Genes<sup>1</sup> are the elementary units of heredity. A chromosome is used to encode a solution to the search problem and it comprises a sequence of a fixed number of genes. The population is a set of chromosomes which exists at a point in time, i.e. in a specific iteration (also referred to as a generation) of the algorithm. Each chromosome, which is to say a solution to the problem, has associated with it a fitness value, i.e. a measure of the quality of a solution, assessed using a suitable fitness function.

The algorithm is typically initialized with a random selection of chromosomes (noting the obvious constraint that each chromosome should encode a valid solution to the problem). Subsequently, in each generational transition, three processes take place: (i) survival of the fittest, (ii) sexual reproduction, and (iii) mutation. The first of these, the survival of the fittest, refers to the passing of the fittest chromosomes from one generation to another. The survival rate, as the proportion of the population, is governed by the parameter  $p_{\text{surv}}$  of the genetic algorithm. Sexual reproduction is a method of creating a new, offspring chromosome from two parent chromosomes of the previous generation. A chromosome with greater fitness is preferentially selected as a parent—this process is shaped by the function which maps a chromosome’s fitness to a probability value. An offspring is created by combining a random selection of genes from its parents. Lastly, mutation effects a random change of genes in a chromosome. Mutation rate  $p_{\text{mut}}$  is another free parameter of the algorithm. As always, the specific sexual reproduction and mutation rules should be designed in a manner which ensures that their result encodes a valid solution to the problem at hand.

### 2.3.2 Genetic algorithm-based sample selection

Having explained the key ideas behind the design of genetic algorithms, we now explain how we utilized these to search the space of possible sample subsets for the one

which can be removed with the least loss of information, as described in Section 2.1.

In our design, each gene in a chromosome corresponds to a particular sample in a trial. Thus, if the trial has  $n$  samples, all chromosomes have the length  $n$ . Furthermore, in the proposed approach genes are binary—they are either ‘on’ or ‘off’. An ‘on’ gene indicates that in the solution encoded by that chromosome the corresponding sample is selected for removal; an ‘off’ gene indicates that it is not. Therefore, for a chromosome to encode a valid solution it is necessary to ensure that it contains exactly  $n_r$  ‘on’ genes.

When two selected parent chromosomes sexually reproduce, offspring are generated as follows. First, we compute a temporary chromosome which can be described as the union of two parents if they are understood as representing sets. Specifically, in this new chromosome a gene is ‘on’ if and only if it is ‘on’ in one of the parent chromosomes. Note that a chromosome produced in this manner does not in general represent a valid solution as it may have more than  $n_r$  ‘on’ genes (but not fewer). However, this is only an intermediate result. From this chromosome, a child chromosome is generated by selecting randomly an  $n_r$ -sized subset of the intermediate chromosome’s ‘on’ chromosomes (with the uniform probability of  $n_r! (n_t - n_r)! / n_t!$  where  $n_t$  is the number of ‘on’ genes in the intermediate chromosome) and setting them ‘on’ in the child too, with the remaining genes staying ‘off’. This process is illustrated in Fig. 1. Clearly, offspring generated in this way have exactly  $n_r$  ‘on’ genes and are thus valid solutions.

Equation 17 provides a ready means of assessing the relative fitness of two solutions—a chromosome encoding a solution associated with lesser loss is fitter than the one associated with a greater loss. However we still need a way of accounting for this in the random selection of chromosomes which mate to generate offspring. Clearly the form of the fitness measure needs to be a monotonically decreasing function of the loss  $\Delta(D)$ . In this work we

map the loss to a quasi-probability value using a decaying exponential:

$$f_p(D) = \exp \{-0.1 \times \Delta(D)\}, \quad (19)$$

and perform normalization over the entire solution set in a generation to ensure that the corresponding probabilities sum to unity.

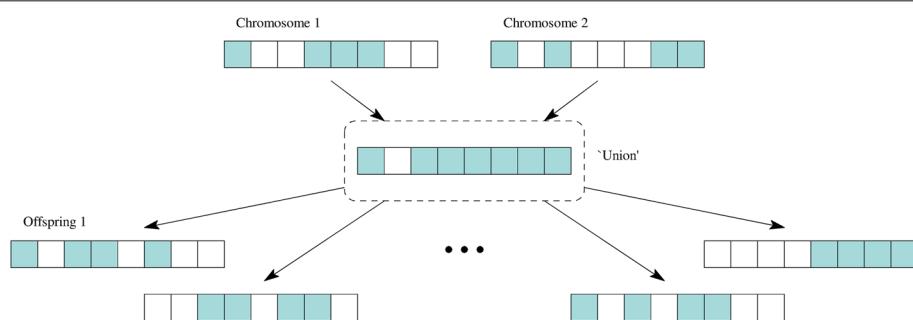
In genetic algorithms the operation of mutation is usually implemented as a change to a gene at a randomly selected locus in a chromosome. As in our case this would produce a chromosome which is not a valid solution, we implement mutation as the swapping of genes at two randomly selected loci in a chromosome. Clearly, this leaves the numbers of ‘on’ and ‘off’ genes unchanged so a valid solution always produces a valid solution too. In effect, the described operation either leaves a chromosome unchanged (when the two genes have the same value) or it changes the values of exactly two genes (when the two genes have different values).

The values of the parameters of the genetic algorithm used in our experiments are detailed in Section 3.

#### 2.4 Pair-wise compatibility-based selection

In contrast to the repeated random draws approach described in Section 2.2 which uses the available information about samples in a post hoc manner, the genetic algorithm-based method introduced in the previous section actively uses this information in guiding the search for the optimal sample subset. At the same time, as is always the case when genetic algorithms are used, although simple to implement the behaviour of this method is difficult to understand. Thus, in this section our aim is to develop a sample set selection algorithm which inherits the key advantageous aspects of the genetic algorithm-based method, while attaining higher interpretability of action.

As we noted earlier, the sample selection problem does not possess the property of optimal substructure. At the same time we argued that it is reasonable to expect (and



**Fig. 1** Pair-wise chromosomal mating as implemented in the proposed genetic algorithm-based sample selection. From a temporary chromosome inheriting all ‘on’ genes from both parents, offspring (children) chromosomes are generated by selecting randomly  $n_r$ -sized subsets of ‘on’ genes

indeed we will shortly demonstrate this empirically) that a combination of good sub-solutions will yield a better solution than one which would be obtained by a random draw. The method we describe now uses pair-wise sample compatibility to drive sample selection, that is, it builds an  $n_r$ -sized solution by adding individual samples to the solution, the preference being guided by the previously included samples and their compatibility with the as of yet not selected samples. We formalize this method next.

As in the simplest method we described in Section 2.2, we generate a hypothesis in the form of a set of samples considered for removal one by one. The first sample in the set is selected by a random draw where the probability of drawing  $x_i$  is:

$$p_1(x_i) = \frac{f_p(\Delta(\{x_i\}))}{\sum_{x \in D} f_p(\Delta(\{x\}))}, \quad (20)$$

where as before  $f_p$  is the function that maps the information loss quantified by Eq. (17) to a quasi-probability value as per Eq. (19). In each subsequent iteration  $k + 1$  until the entire  $n_r$ -sized set is generated (i.e. until  $k = n_r$ ), a sample  $x_i$  from the set of not yet selected samples is drawn by weighting its individual probability-mapped loss by its compatibility with each of the already selected samples:

$$p_{k+1}(x_i) = \frac{\hat{p}_{k+1}(x_i)}{\sum_{x \in D \setminus D_s} \hat{p}_{k+1}(x)}, \quad (21)$$

where

$$\hat{p}_{k+1}(x_i) = f_p(\Delta(x_i)) \times \prod_{x \in D_s} f_p(\Delta(\{x_i, x\})), \quad (22)$$

and  $D_s$  is the subset of samples selected by the preceding  $k$  draws. This drawing rule captures a weighting of the initial probability in Eq. (18). Specifically, the weighting is done by the compatibility  $f_p(\Delta(\{x_i, x_j\}))$  of a not yet selected sample  $x_i$  with each of the previously selected samples  $x$ . As in the sample random draws based method, after  $n_r$  iterations the entire hypothesis has been generated and its quality can be assessed in the same manner as before by the associated information loss given by Eq. (17). After  $n_h$  hypotheses (i.e. sample subsets) are drawn, the best subset for removal from the subsequent stages of the trial is made. Specifically, this is the sample subset corresponding to the hypothesis associated with the least loss.

### 3 Experiments and discussion

We now turn our attention to the empirical analysis of the methods proposed in the previous section, and report and discuss their performance. Like most previous work in this area, we adopt the evaluation protocol standard in the domain of adaptive trials research and obtain data using a simulated experiment [14, 18–22]. The first experiment investigates the effectiveness of the proposed methods on the proximal aims of information loss minimization.

The second experiment examines their performance on the ultimate goal of targeted sample removal. The precise methodology is explained in detail next.

#### 3.1 Methodology

In the first set of experiments our aim was to investigate the performance of the three sample subset selection strategies described in Section 2. This was done in the context of their ability to minimize information loss expressed by Eq. (17). We approach this by generating a set of synthetic trial outcomes and evaluate the loss associated with each of the algorithms across 100 repeated runs to minimize variability caused by the stochastic nature of the algorithms. We generate the sets of  $n_t = n_c = 200$  trial outcomes of the treatment and control groups by random draws from respectively  $\mathcal{N}(1, 5)$  and  $\mathcal{N}(0, 5)$ . To ensure a fair comparison of methods which are governed by different parameters, we compare a genetic algorithm with the population size  $n_p$ , the maximal number of generations  $n_g$ , and the fittest survival rate  $r_f$  with a repeated random draws algorithm which makes  $n_h$  draws computed as follows:

$$n_h = (1 - r_f) \times n_p \times n_g + r_f \times n_p. \quad (23)$$

This way in a single run, in general, each of the algorithms generates the same number of unique hypotheses. The term  $(1 - r_f) \times n_p$  is the number of chromosomes in a generation generated as offspring from the previous generation; these in general correspond to unique hypotheses. Thus,  $(1 - r_f) \times n_p \times n_g$  is the number of all such hypotheses over the entire run of the algorithm (this includes  $(1 - r_f) \times n_p$  hypotheses of the first generation which although not generated as offspring are unique as they are randomly drawn). The second term in the equation,  $r_f \times n_p$ , is simply the remaining number of hypotheses of the first generation which, again, are unique by design. In our experiments, we used the following parameter values: the survival rate  $r_f = 0.2$ , the population size in a generation  $n_p = 50$ , and the number of generations computed from Eq. 23 for a specific value of  $n_h$  which we varied from  $n_h = 100$  to  $n_h = 800$ . Lastly, to examine the effect that the proportion of samples removed has on relative performances of different selection methods, we repeated the experiment with different values for  $n_r$ : 50, 100, 150, and 200 (i.e. 12.5, 25, 37.5, and 50% of the total number of samples).

For the second set of experiments, we extend the simulation to model the entire duration of a trial and examine the effect of different sample size reduction strategies on the ultimate outcome of the trial. Specifically, we simulated a trial involving 400 samples, half of which were assigned to the control and the other half to the treatment group. For each sample, we maintain a variable which

describes the associated effect of the assigned intervention. Thus, for the treatment group, we have  $n_t$  variables  $\{x_1, \dots, x_{n_t}\}$  and similarly for the control group  $n_c$  variables  $\{x_{n_t+1}, \dots, x_{n_t+n_c}\}$ . 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 associated with the  $i$ -th treatment sample at the preceding time step  $k$ ,  $x_i(k)$ , is updated in the following manner:

$$x_i(k + 1) = x_i(k) + w_i^{(t)}(k + 1) \times \exp \left\{ -\frac{k + 1}{10} \right\}, \quad (24)$$

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

$$W_t \sim \mathcal{N}(0.01, 0.05). \quad (25)$$

Notice that this progression has a 'ground truth' asymptote at:

$$\begin{aligned} \lim_{k \rightarrow \infty} E[x_i(k)] &= x_i(0) + 0.1 \times \frac{\exp \left\{ -\frac{1}{10} \right\}}{1 - \exp \left\{ -\frac{1}{10} \right\}} \\ &\approx x_i(0) + 0.95. \end{aligned} \quad (26)$$

Similarly, the effect of the control intervention on the  $i$ -th control sample is:

$$x_{n_t+i}(k + 1) = x_{n_t+i}(k) + w_i^{(c)}(k + 1) \times \exp \left\{ -\frac{k + 1}{10} \right\}, \quad (27)$$

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

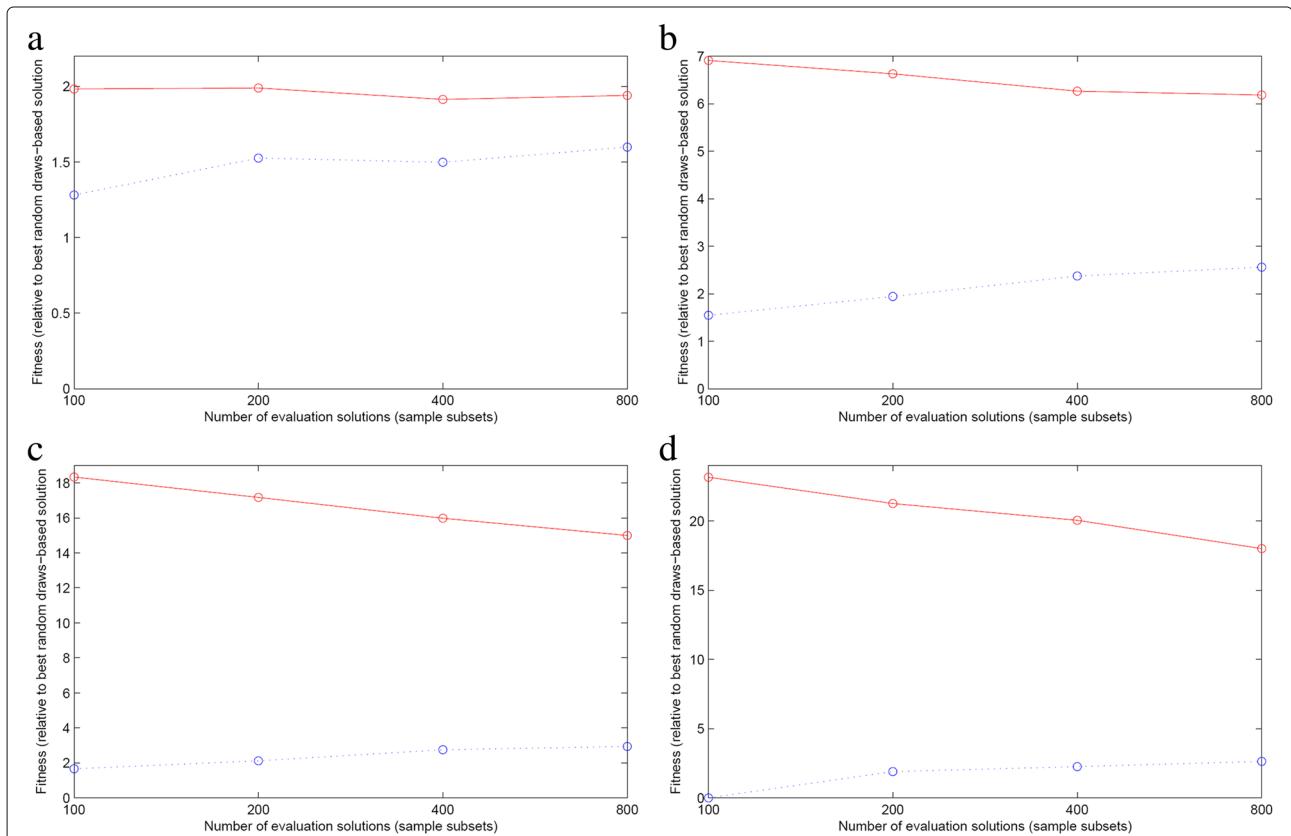
$$W_c \sim \mathcal{N}(0.00, 0.05). \quad (28)$$

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

$$\forall i = 1 \dots n_t + n_c. x_i(0) = 0. \quad (29)$$

### 3.1.1 Results and discussion

The key results of the first set of experiments are summarized in Fig. 2. Each plot in this figure shows the average fitness of the best sample selection solutions produced



**Fig. 2** Comparative performance of different sample selection methods on information loss. Each plot shows the average fitness of the methods described in Section 2.4 (solid red line) and Section 2.3.2 (dotted blue line) relative to the simplest, baseline described in Section 2.2, and its variation as a function of the number of generated hypotheses. Different plots correspond to different sizes of the sample set selected for removal (from the total initial number of samples equal to 400). **a**  $n_r = 50$  (12.5% of data). **b**  $n_r = 100$  (25% of data). **c**  $n_r = 150$  (37.5% of data). **d**  $n_r = 200$  (50% of data)

by the methods described in Section 2.4 (solid red line) and Section 2.3.2 (dotted blue line) relative to the simplest, baseline described in Section 2.2. The variation in relative fitness is shown as a function of the number of generated hypotheses  $n_h$ . Regardless of the proportion of samples which are removed, the plots demonstrate a consistent behaviour of the described selection methods. As expected both informed approaches, the genetic algorithm based and the pair-wise compatibility-based method, outperform the simple baseline. What is more, the pair-wise compatibility-based method consistently exhibited superior performance. This was particularly pronounced with the increase in the proportion of samples which were to be removed. When only 12.5% of samples were removed, the average fitness of the solution produced by the pair-wise compatibility-based method was approximately 1.6 and 2 times greater than respectively the fitness of the average genetic algorithm based and the repeated random draws based methods. When the proportion was increased to 50% of the samples, the corresponding gains in fitness were approximately 10 and 20.

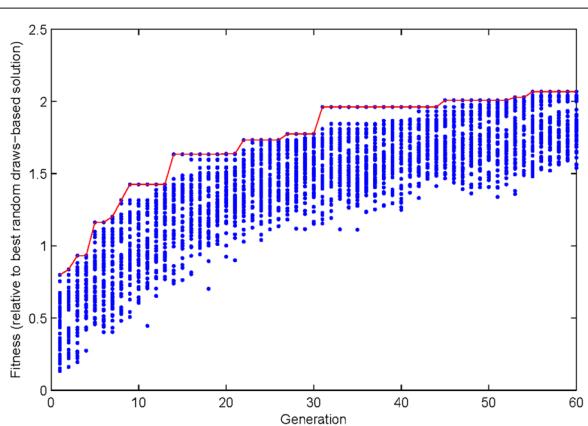
It is insightful to investigate the behaviour of the genetic algorithm-based method in further detail. Figure 3 shows the typical variation in the fitness of all samples in a generation across time as well as the fitness of the best solution, relative to the fitness of the solution produced by the baseline, repeated random draws based method using  $n_r = 100$  (i.e. 25% of the samples are to be removed). Blue dots show the relative fitness (ordinate) of each individual (solution) in a generation (abscissa). The red line shows the variation of the maximal fitness (i.e. the best solution) in a generation and is, by the design of the algorithm, non-decreasing. It can be readily seen that all

of the solutions generated in the first generation, and thus the best solution of this generation too, are worse than the baseline solution of the repeated random draws based method (as witnessed by their relative fitness being lower than unity). This is to be expected from theory; recall from the design of the experiment that in the first generation  $n_g$  random solutions are generated whereas the repeated random draws based method performs  $n_h = (1 - r_f) \times n_p \times n_g + r_f \times n_p$  solutions which is equal to the number of unique solutions created by the genetic algorithm across all generations. Indeed, the plot in Fig. 3 shows that the genetic algorithm finds a better than baseline solution after only five generations, i.e. after generating fewer than 10% of the number of solutions examined by the repeated random draws based method.

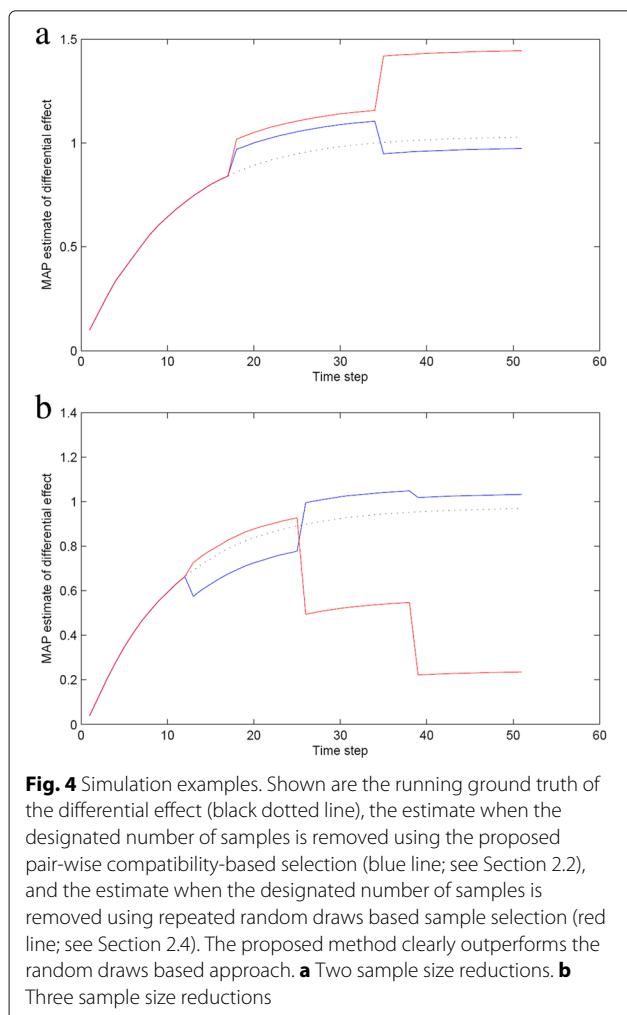
The same relative performances of different sample selection methods we observed in the first set of experiments were maintained when the experiment was extended to a simulation of an entire adaptive trial. The error of the estimate of differential outcome of treatment and control interventions was consistently lower using the proposed pair-wise compatibility-based method than the baseline repeated random-draws algorithm. Specifically, we found that the average error was approximately 2.9, 3.6, and 4.1 times lower for respectively one, two, and three sample size reductions. What is more, the consistency of the estimate was improved too as witnessed by the standard deviation of the error which was 3.3, 5.0, and 6.1 times lower for the proposed method. Representative simulations are illustrated in Fig. 4 which shows the running ground truth of the differential effect (black dotted line), the estimate when the designated number of samples is removed using the proposed pair-wise compatibility-based selection (blue line; see Section 2.2), and the estimate when the designated number of samples is removed using repeated random draws based sample selection (red line; see Section 2.4).

#### 4 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, but rather a complementary element of the same framework. Our approach is based on the idea of selecting the sample subset for removal in a manner which minimizes the associated loss of information which we formalized using a Bayesian framework. Using the derived result we described three algorithms which approach the resulting optimization problem in different ways. Specifically,



**Fig. 3** The fitness of the generation in the genetic algorithm-based solution across time. Blue dots show the fitness (ordinate) of each individual (solution) in a generation (abscissa). The red line shows the variation of the maximal fitness (i.e. the best solution) in a generation and is, by the design of the algorithm, non-decreasing



**Fig. 4** Simulation examples. Shown are the running ground truth of the differential effect (black dotted line), the estimate when the designated number of samples is removed using the proposed pair-wise compatibility-based selection (blue line; see Section 2.2), and the estimate when the designated number of samples is removed using repeated random draws based sample selection (red line; see Section 2.4). The proposed method clearly outperforms the random draws based approach. **a** Two sample size reductions. **b** Three sample size reductions

we proposed sample selection methods using (i) repeated random draws, (ii) a genetic algorithm, and (iii) pair-wise sample compatibilities. Experiments on simulated data demonstrate the effectiveness of all three approaches, with a consistently superior performance exhibited by the pair-wise sample compatibilities-based method.

### Endnote

<sup>1</sup> Hereafter, we omit the use of inverted commas for the sake of reducing clutter, with the understanding that the terms are a part of the technical jargon of genetic algorithms unless stated otherwise.

### Competing interests

The author declares that he has no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 May 2017 Accepted: 17 October 2017

Published online: 30 October 2017

### References

- CL Meinert, *Clinical Trials: Design, Conduct, and Analysis*, 3rd edn. (Oxford University Press, New York, 1986)
- S Piantadosi, *Clinical Trials: A Methodologic Perspective*, 3rd edn. (Wiley, Hoboken, New Jersey, 1997)
- L Friedman, C Furberg, D DeMets, *Fundamentals of Clinical Trials*, 3rd edn. (Springer, New York, 1998)
- J Penston, Large-scale randomised trials—a misguided approach to clinical research. *Med. Hypotheses*. **64**(3), 651–657 (2005)
- LD Fisher, Self-designing clinical trials. *Stat. Med.* **17**, 1551–1562 (1998)
- 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)
- HMJ Hung, S-J Wang, RT O'Neill, Methodological issues with adaptation of clinical trial design. *Pharmaceut Statist.* **5**, 99–107 (2006)
- S-C Chow, M Chan, *Adaptive Design Methods in Clinical Trials*. (Chapman & Hall, 2011)
- L Cui, HMJ Hung, SJ Wang, Modification of sample size in group sequential clinical trials. *Biometrics*. **55**, 321–324 (1999)
- SE Nissen, ADAPT: The wrong way to stop a clinical trial. *PLoS Clin. Trials.* **1**(7), 35 (2006)
- T Lang, Adaptive trial design: could we use this approach to improve clinical trials in the field of global health? *Am. J. Trop. Med. Hyg.* **85**(6), 967–970 (2011)
- C Jennison, BW Turnbull, Mid-course sample size modification in clinical trials based on the observed treatment effect. *Stat. Med.* **22**(6), 971–993 (2003)
- J Aitchison, JAC Brown, *The Lognormal Distribution*. (Cambridge University Press, Cambridge, 1957)
- O Arandjelović, A new framework for interpreting the outcomes of imperfectly blinded controlled clinical trials. *PLoS ONE*. **7**(12), e48984 (2012)
- O Arandjelović, R Cipolla, Achieving robust face recognition from video by combining a weak photometric model and a learnt generic face invariant. *Pattern Recogn.* **46**(1), 9–23 (2013)
- B Conn, O Arandjelović, in *Proc. IEEE International Joint Conference on Neural Networks*. Towards computer vision based ancient coin recognition in the wild—automatic reliable image preprocessing and normalization (IEEE, 2017), pp. 1457–1464
- I Schlag, O Arandjelović, in *Proc. International Conference on Computer Vision Workshop on e-Heritage*. Ancient Roman coin recognition in the wild using deep learning based recognition of artistically depicted face profiles (IEEE, 2017)
- KE James, DA Bloch, KK Lee, HC Kraemer, RK Fuller, An index for assessing blindness in a multi-centre clinical trial: disulfiram for alcohol cessation—a va cooperative study. *Stat. Med.* **15**(13), 1421–1434 (1996)
- H Bang, L Ni, CE Davis, Assessment of blinding in clinical trials. *Contemp. Clin. Trials.* **25**(2), 143–156 (2004)
- O Arandjelović, Assessing blinding in clinical trials. *Adv. Neural Inf. Process. Syst.* **25**, 530–538 (2012)
- O Arandjelović, Clinical trial adaptation by matching evidence in complementary patient sub-groups of auxiliary blinding questionnaire responses. *PLoS ONE*. **10**(7), e0131524 (2015)
- O Arandjelović, Sample-targeted clinical trial adaptation. *Proc. AAAI Conf Artif Intell.* **3**, 1693–1699 (2015)