Determination of the optimal sample size for a clinical trial accounting for the population size
Nigel Stallard, Frank Miller, Simon Day, Siew Wan Hee, Jason Madan, Sarah Zohar, Martin Posch

To cite this version:

HAL Id: inserm-02456649
https://inserm.hal.science/inserm-02456649
Submitted on 27 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Determination of the optimal sample size for a clinical trial accounting for the population size

Nigel Stallard∗,1, Frank Miller2, Simon Day3, Siew Wan Hee1, Jason Madan4, Sarah Zohar5, and Martin Posch6

1 Statistics and Epidemiology, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK
2 Department of Statistics, Stockholm University, Stockholm, Sweden
3 Clinical Trials Consulting and Training Limited, Buckingham, UK
4 Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, UK
5 INSERM, U1138, team 22, Centre de Recherche des Cordeliers, Université Paris 5, Université Paris 6, Paris, France
6 Section of Medical Statistics, CeMSIIS, Medical University of Vienna, Austria

Received 28 October 2015; revised 10 February 2016; accepted 9 March 2016

The problem of choosing a sample size for a clinical trial is a very common one. In some settings, such as rare diseases or other small populations, the large sample sizes usually associated with the standard frequentist approach may be infeasible, suggesting that the sample size chosen should reflect the size of the population under consideration. Incorporation of the population size is possible in a decision-theoretic approach either explicitly by assuming that the population size is fixed and known, or implicitly through geometric discounting of the gain from future patients reflecting the expected population size. This paper develops such approaches. Building on previous work, an asymptotic expression is derived for the sample size for single and two-arm clinical trials in the general case of a clinical trial with a primary endpoint with a distribution of one parameter exponential family form that optimizes a utility function that quantifies the cost and gain per patient as a continuous function of this parameter. It is shown that as the size of the population, N, or expected size, N∗, in the case of geometric discounting, becomes large, the optimal trial size is $O(N^{1/2})$ or $O(N^{1/2})$. The sample size obtained from the asymptotic expression is also compared with the exact optimal sample size in examples with responses with Bernoulli and Poisson distributions, showing that the asymptotic approximations can also be reasonable in relatively small sample sizes.

Keywords: Bayesian; Clinical trial design; Decision theory; Exponential family form; Optimal sample size.

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

The problem of determining the sample size for a clinical trial is a very common one. For large-scale definitive phase III clinical trials, a frequentist approach is usually adopted, with the sample size chosen so as to control the type I error rate at a specified level, $\alpha$, and to give specified power $1 - \beta$, to detect some appropriately chosen size of treatment effect (see, e.g. Pocock, 1983, for details). Choice of $\alpha = 0.05$ and $\beta = 0.1$ or 0.2 is typical.

*Corresponding author: e-mail: n.stallard@warwick.ac.uk
The sample sizes obtained using the frequentist approach do not always seem appropriate. In particular they do not reflect the size of the population to which the results of the trial apply. The population size is relevant, however, when considering decisions made on the basis of trial results. This is particularly true for clinical trials conducted in rare diseases or other small populations, when the population size means that a large trial would be infeasible, or even impossible.

One way in which the size of the population can influence the sample size is to use a decision theoretic approach in which the benefits to future patients in the population, sometimes called the “patient horizon”, are explicitly considered so that future benefit depends on the size of this population. Such an approach has been proposed and discussed by numerous authors over the last 50 years (see, e.g. Anscombe, 1963; Colton, 1963; Sylvester, 1988; Berry et al., 1994; Cheng et al., 2003; Kikuchi and Gittins, 2009 and reviews by Pezeshk et al., 2013; Hee et al., 2016). Although this approach has very rarely been implemented in practice, it can nevertheless provide important insight into an appropriate choice of sample size for a clinical trial.

The role of the population size in determination of the optimal sample size for a clinical trial has been considered by Cheng et al. (2003). They considered single-arm and two-arm clinical trials with a primary endpoint following a Bernoulli distribution indicating either success or failure of treatment for each patient. Adopting a decision-theoretic approach, they obtained designs that maximize the total number of successes. Denoting the population size by $N$, they show that asymptotically as $N \to \infty$, the optimal sample size for a clinical trial is $O(N^{1/2})$ and give an expression for the asymptotically optimal sample size that depends on the prior distributions for the unknown probability of success for each trial arm.

In this paper, we extend the work of Cheng et al. (2003), both to more general distributional forms for the primary endpoint and to situations in which the aim is to maximize some general utility expressed as a function of a parameter of these distributions. We show that the result that the optimal sample size is $O(N^{1/2})$ applies for any continuous utility function and for responses with a distribution of any one-parameter exponential family form assuming a conjugate prior distribution. We also consider the case where no finite patient horizon is assumed, but gains from future patients are geometrically discounted, that is the gain from the $j$th patient is multiplied by $\lambda^j$ for some discounting parameter, $\lambda < 1$ (Berry and Fristedt, 1985). As considered in the discussion section, the size of $\lambda$ can also reflect the size of the patient population in this setting, via an effective number of patients, $(1 - \lambda)^{-1}$, which will be denoted $N^*$. We show that in this case as $\lambda \to 1$ so that $N^*$ is large, the optimal sample size is $O(N^{1/2})$. We also investigate through exact calculation the small-sample accuracy of the large sample approximations. Although the results obtained depend on asymptotics, we show that, depending on the exact form of the utility function chosen, these may be reasonable even for extremely rare diseases, for example for patient populations of 1000 or less when the optimal sample size can be less than 50.

2 Detailed problem description and notation

2.1 Outline of the decision problem

Suppose that a clinical trial is to be conducted to choose between two treatments with $n_1$ and $n_2$ patients receiving treatment 1 and treatment 2, respectively, where treatment 2 may be the current standard treatment included as a control. Note that taking $n_2 = 0$ corresponds to a single-arm trial, though the decision to be taken at the end of such a trial remains comparative, with a choice being made regarding treatment of future patients.

It is assumed that the gain associated with treatment of patients in the trial or patients outside the trial who receive each treatment can be specified as a function of a parameter of the distribution for the response for patients receiving that treatment. It is noted that here “gain” is to be interpreted widely to include any kind of costs, losses, gains, or benefits associated with treatment. Following the trial,
the most preferable treatment, that is the treatment for which the posterior expected gain given the observed data, is highest, will be selected. The remaining patients will then receive this treatment. We wish to determine the optimal values, \( n_1^\ast \) and \( n_2^\ast \), of the sample sizes \( n_1 \) and \( n_2 \) and to determine how \( n_1^\ast \) and \( n_2^\ast \) depend on the population size.

### 2.2 Decision problem formulation and notation

We will assume that responses for patients follow some distribution of natural one-parameter exponential family form. In detail, let \( Y_{ij} \) denote the response for patient \( j \) receiving treatment \( i \) and assume \( Y_{i1}, \ldots, Y_{in} \) are i.i.d. with density \( f_j(y \mid \psi_i) = a_i(y) \exp(y \psi_i - b_i(\psi_i)) \) for some \( a_i(y) \) and \( b_i(\psi_i) \). Typically responses from patients in the two treatment groups will follow distributions of the same form, that is functions \( a_i \) and \( b_i \) will not depend on \( i \), with \( \psi_1 \) and \( \psi_2 \) differing, though this is not assumed.

We will assume that \( \psi_1 \) and \( \psi_2 \) are taken to have independent prior distributions of conjugate form, that is with \( \psi_i \) having density \( \pi(\psi_i \mid \eta_i, \gamma_{i0}) = c_i(\eta_i, \gamma_{i0}) \exp(\eta_i \gamma_{i0} - \eta_i b_i(\psi_i)) \) for some \( \gamma_{i0} \) and \( \eta_i \) and normalising constant \( c_i(\eta_i, \gamma_{i0}) \), \( i = 1, 2 \). The values \( \gamma_{i0} \) and \( \eta_i \) can be interpreted respectively as the prior mean of \( \xi_i = b_i(\psi_i) = E(Y_{ij}) \) and the number of observations to which the prior information is equivalent, so that following a trial with \( n_i \) patients receiving treatment and observation of \( Y = (Y_{i1}, \ldots, Y_{in}, Y_{21}, \ldots, Y_{2n_2}) \), the posterior mean for \( \bar{\xi}_i \) given \( Y \) is equal to

\[
\frac{n_{i0}Y_{i0} + n_i \bar{Y}_i}{n_{i0} + n_i}, \quad i = 1, 2
\]

with \( \bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i \) (see Bernado and Smith, 2000).

Suppose that the expected gain from a patient receiving treatment \( i \) in the trial is \( h_i(\bar{\xi}_i) \), and that the expected gain to a future patient receiving treatment \( i \) is \( g_i(\bar{\xi}_i) \) where \( h_i \) and \( g_i, i = 1, 2 \) are such that the expected values \( E_0(h_i(\bar{\xi}_i)) \), \( E_0(g_i(\bar{\xi}_i)) \) and \( E_0(\max(g_1(\bar{\xi}_i), g_2(\bar{\xi}_i))) \) where \( E_0(\cdot) \) denotes the expected value taken over the prior distribution of \( \bar{\xi}_i \), exist and \( h_i \) and \( g_i \) are assumed to be differentiable with \( g_i \) strictly increasing and with finite derivative. Assume further that

\[
E_0(h_i(\bar{\xi}_i)) \leq E_0(\max(g_1(\bar{\xi}_i), g_2(\bar{\xi}_i))), \quad i = 1, 2.
\]

This ensures that the gain from treating patients in the trial cannot exceed that from treating them outside the trial. This is considered further in the discussion section below.

We will consider two cases. In the first, the population is considered to be finite with known size, \( N \). The number of patients treated following the trial is thus \( N - n_1 - n_2 \). In the second case, no finite population size is assumed, but the gain from future patients is geometrically discounted, so that the gain from patient \( j \) if they receive treatment \( i \) is \( \lambda^{j-1}h_i(\bar{\xi}_i) \) if they are included in the trial and \( \lambda^{j-1}g_i(\bar{\xi}_i) \) if they are treated following the trial, for some \( \lambda < 1 \).

The geometric discounting of gains from future patients can be interpreted in a number of ways. One interpretation is that gains further in the future are reduced to reflect either opportunity loss or loss of financial interest on an investment (see, e.g. Fergusson, 2008). With this interpretation it might be appropriate to take \( \lambda < 1 \) and \( N \) finite. An alternative interpretation is to imagine that gain from each future patient is of constant value, as is assumed in the finite horizon case, but that the size of the population, \( N \), rather than being fixed in advance, is random, following a geometric distribution. This could be the case if, for example, the population is limited by some new treatment becoming available at which point the trial, or use of the recommended treatment following the trial, will be terminated, with the probability of this event constant over time (see Berry and Fristedt, 1985). In this interpretation \( N^* = (1 - \lambda)^{-1} \) is the expected population size prior to this new treatment becoming available. The size of \( \lambda \) and the resulting \( N^* \) can thus reflect the population size, and in some ways a smaller value of \( \lambda \), corresponding to fewer patients being available prior to the new treatment becoming available,
consider first the setting of a finite patient horizon of size $N$. Following observation of data $Y = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2})$, the total expected gain if treatment $i$ is recommended for all further $N - n_1 - n_2$ patients is $n_1 E_{\xi_1 \mid Y}(h_1(\xi_1) \mid Y) + n_2 E_{\xi_1 \mid Y}(h_2(\xi_2) \mid Y) + (N - n_1 - n_2) E_{\xi_1 \mid Y}(g(\xi) \mid Y)$ where $E_{\xi_1 \mid Y}(\cdot \mid Y)$ denotes the expected value taken over the posterior distribution of $\xi$ given $Y$.

The optimal action at the end of the trial is thus to select the treatment with the largest value of $E_{\xi_1 \mid Y}(g(\xi) \mid Y)$ and the expected gain assuming this action is taken is equal to $n_1 E_{\xi_1 \mid Y}(h_1(\xi_1) \mid Y) + n_2 E_{\xi_1 \mid Y}(h_2(\xi_2) \mid Y) + (N - n_1 - n_2) \max_{i=1,2} E_{\xi_1 \mid Y}(g_i(\xi) \mid Y)$.

Since, prior to the commencement of the trial, $Y$ is unknown, the expected gain from the trial is equal to $E_0(\mathcal{G})$ where the expectation is taken over the prior distribution for $\xi_1$ and $\xi_2$ and $\mathcal{G}$ is the function of $\xi_1, \xi_2, N, n_1, \text{and } n_2$ given by

$$
\mathcal{G} = n_1 E_Y(E_{\xi_1 \mid Y}(h_1(\xi_1) \mid Y)) + n_2 E_Y(E_{\xi_1 \mid Y}(h_2(\xi_2) \mid Y)) + (N - n_1 - n_2) E_Y(\max_{i=1,2} E_{\xi_1 \mid Y}(g_i(\xi) \mid Y))
$$

(3)

where $E_Y$ denotes the expectation taken over $Y$ for a given value of $\xi_1$ and $\xi_2$ so that expectations are taken first over the posterior distributions of $\xi_i$ given $Y$ and then over $Y$ given $\xi_1$ and $\xi_2$.

Since $E_0(\mathcal{G}) = E_0(E_Y(E_{\xi_1 \mid Y}(h_i(\xi_1) \mid Y)))$ is equal to the prior expectation $E(h_i(\xi_1))$ for any function $h_i$ for which the expectations exist, we get

$$
E_0(\mathcal{G}) = n_1 E_0(h_1(\xi_1)) + n_2 E_0(h_2(\xi_2)) + (N - n_1 - n_2) E_0(\max_{i=1,2} E_{\xi_1 \mid Y}(g_i(\xi) \mid Y)).
$$

(4)

We wish to find the optimal values of $n_1$ and $n_2$, that is the values for which $E_0(\mathcal{G})$ is maximized. For small $N$, when the optimal sample sizes, $n_1^*$ and $n_2^*$ will also be small, it may be feasible to evaluate the prior expected gain given by (4) directly, taking the expectation over the prior predictive distribution for $Y$ and to find $n_1^*$ and $n_2^*$ by a numerical search. For larger $N$, such an approach may be infeasible. In this case asymptotic expressions for the optimal sample sizes, $n_1^*$ and $n_2^*$, as the population size, $N$, becomes large are more useful.

For finite $n_{01}$ and $n_{02}$, the expectation $E_0(E_Y(\max_{i=1,2} E_{\xi_1 \mid Y}(g_i(\xi) \mid Y)))$ is increasing in $n_1$ and $n_2$. Thus for $N > n_1 + n_2$, $(N - n_1 - n_2)E_0(E_Y(\max_{i=1,2} E_{\xi_1 \mid Y}(g_i(\xi) \mid Y)))$ and hence $E_0(\mathcal{G})$ is also increasing in $n_1$ and $n_2$. Thus as $N \to \infty$ the optimal trial design has both $(n_{01} + n_1) \to \infty$ and $(n_{02} + n_2) \to \infty$. The case in which both $n_{01}$ and $n_{02}$ are both infinite corresponds to both $\xi_1$ and $\xi_2$ being known a priori. We will therefore consider the case in which, without loss of generality, $n_{01}$ is finite, and derive the optimal value $n_1^*$.

Note that as $N \to \infty$, so that the optimal $n_{01} + n_1$ and $n_{02} + n_2$ also approach infinity, $E_0(E_Y(\max_{i=1,2} E_{\xi_1 \mid Y}(g_i(\xi) \mid Y))) \to \max_{i=1,2} E_0(g_i(\xi))$ so that the optimal sample sizes are such that $n_i/N \to 0$, $i = 1, 2$. 

---

By the central limit theorem we have $\tilde{Y}_i \xrightarrow{d} N(\xi_i, v_i(\xi_i)/n_i)$ where $v_i(\xi_i)$ denotes the variance of $Y_{ij}$. Thus from (1), applying the delta method since $g_i$ is assumed to be differentiable and strictly increasing so that the derivative, $g_i'(\xi_i)$, is non-zero, we get

$$
E_{\xi|\gamma}(g_i(\xi_i) \mid \tilde{Y}_i) \xrightarrow{d} N \left( g_i \left( \frac{n_0 y_{0i} + n_i \xi_i}{n_0 + n_i} \right), \left( \frac{n_i}{n_0 + n_i} \right)^2 \frac{v_i(\xi_i)}{n_i} \right).
$$

(5)

Using an expression for the expected value of the maximum of two normally distributed random variables given by Clark (1961), we have

$$
E_Y \left( \max_{i=1,2} E_{\xi|\gamma}(g_i(\xi_i) \mid Y) \right) \rightarrow \bar{E}_1 \Phi(\delta) + \tilde{E}_2 \Phi(-\delta) + \alpha \phi(\delta)
$$

(6)

where $\bar{E}_1 = g_1((n_0 y_{01} + n_1 \xi_1)/(n_0 + n_1)), a^2 = \sum_{i=1}^2 n_i^2 (g_i'(\xi_i))^2 v_i(\xi_i)/n_i(n_0 + n_i)^2$, $\delta = (\bar{E}_1 - \tilde{E}_2)/\alpha$ and $\phi$ and $\Phi$ denote standard normal density and distribution functions. Thus

$$
E_0(\bar{G}) \rightarrow n_1 E_0(h_1(\xi_1)) + n_2 E_2(h_2(\xi_2)) + (N - n_1 - n_2)E_0(\tilde{E}_1 \Phi(\delta) + \tilde{E}_2 \Phi(-\delta) + \alpha \phi(\delta)).
$$

(7)

In order to find $n_i^*$, we obtain the derivative $\partial E_0(\bar{G})/\partial n_i = E_0(\partial \bar{G}/\partial n_i)$. From (3),

$$
\frac{\partial \bar{G}}{\partial n_i} = h_1(\xi_i) - E_Y \left( \max_{i=1,2} E_{\xi|\gamma}(g_i(\xi_i) \mid Y) \right) + (N - n_1 - n_2) \frac{\partial E_Y(\max_{i=1,2} E_{\xi|\gamma}(g_i(\xi_i) \mid Y))}{\partial n_i}.
$$

(8)

We will thus find a large-sample approximation for this derivative.

Equation (6) gives an approximation for $E_Y(\max_{i=1,2} E_{\xi|\gamma}(g_i(\xi_i) \mid Y))$. The derivative of the right hand side of (6) is equal to

$$
\Phi(\delta) g_1'(\xi_1) n_{01} (\xi_i - y_{0i}) + \phi(\delta) \frac{(n_0 - n_1) v_i(\xi_i) (g_i'(\xi_i))^2}{2a(n_0 + n_1)^3}
$$

(see Web Appendix A for details), so that $\partial \bar{G}/\partial n_i$ can be approximated by

$$
h_1(\xi_i) - E_Y(\max_{i=1,2} E_{\xi|\gamma}(g_i(\xi_i) \mid Y)) +
$$

$$+(N - n_1 - n_2) \left( \Phi(\delta) g_1'(\xi_1) n_{01} (\xi_i - y_{0i}) + \phi(\delta) \frac{(n_0 - n_1) v_i(\xi_i) (g_i'(\xi_i))^2}{2a(n_0 + n_1)^3} \right).
$$

The limit of this as $N$, $n_1$ and $n_0 + n_2$ all approach infinity with $n_1/N$ and $n_2/N$ approaching 0 (again, see Web Appendix A for details), has expected value

$$
E_0(h_1(\xi_1) - \max_{i=1,2} g_i(\xi_i)) + \frac{N}{n_1^2} \int - \frac{v_i(g_i^{-1}(g_i'(\xi_2)))}{2} (g_i'(g_i^{-1}(g_i'(\xi_2)))) \pi(g_i^{-1}(g_i'(\xi_2)), \xi_2) d\xi_2
$$

(9)

where $\pi(\xi_1, \xi_2)$ denotes the joint prior distribution of $\xi_1$ and $\xi_2$. 

Optimal clinical trial sample size for population size

Setting this to zero and solving for \(n_1\), the maximum expected gain is found to be at

\[
 n_1^* = \sqrt{\frac{N \int v_1(g_1^{-1}(g_2(\xi_1)))g_1'(g_1^{-1}(g_2(\xi_2)))\pi(g_1^{-1}(g_2(\xi_2)), \xi_2)d\xi_2}{2(E_0(\max_{i=1,2}g_i(\xi_i)) - E_0(h_i(\xi_i)))}}. \tag{10}
\]

Note that the fact that \(g_1\) is increasing and the requirement (2) ensures that both numerator and denominator are positive so that the square root exists. It is also interesting to note that the asymptotic density \(\pi(\xi_1), \xi_2\) has a Bernoulli distribution with parameter \(\lambda = 1 + \lambda^{-1} \sum_{j=1}^{n_1+n_2} \lambda^{-j} / (n_1 + n_2) h_i(\xi_j)\). The total expected gain if treatment \(i\) is recommended for all further \(N - n_1 - n_2\) patients is then

\[
 n_1^* = \sqrt{\frac{N \int v_1(g_2(\xi_1))g_2'(g_2(\xi_2))\pi(\xi_1, g_2(\xi_2))d\xi_1}{2(E_0(\max_{i=1,2}g_i(\xi_i)) - E_0(h_i(\xi_i)))}}. \tag{11}
\]

When \(n_2 = \infty\), the prior distribution has mass at \(\xi_2 = y_{n_2}\) only, and so may be written as a univariate density \(\pi(\xi_1)\), so that the optimal value of \(n_1\) becomes

\[
 n_1^* = \sqrt{\frac{N v_1(g_1^{-1}(g_2(y_{n_2})))g_1'(g_1^{-1}(g_2(y_{n_2}))))\pi(g_1^{-1}(g_2(y_{n_2})))}{2(E_0(\max_{i=1,2}g_i(\xi_i)) - E_0(h_i(\xi_i)))}}. \tag{12}
\]

As \(n_2 \to \infty\), \(\partial E_Y(\max_{i=1,2} E_{y_i|Y}(g_i(\xi_i))|Y) / \partial n_2 \to 0\). Thus from an expression similar to (8) giving the derivative with respect to \(n_2\) and (2), the derivative is negative and \(n_2^* = 0\).

When \(g_1\) and \(g_2\) are identical, (10) becomes

\[
 n_1^* = \sqrt{\frac{N \int v_1(\xi_2)g_1'(\xi_2)\pi(\xi_2, \xi_2)d\xi_2}{2(E_0(\max_{i=1,2}g_i(\xi_i)) - E_0(h_i(\xi_i)))}}.
\]

and when \(g_i(\xi_i) = h_i(\xi_i) = \xi_i, i = 1, 2\), this becomes

\[
 n_1^* = \sqrt{\frac{N \int v_1(\xi_2)\pi(\xi_2, \xi_2)d\xi_2}{2(E_0(\max_{i=1,2}\xi_i)) - E_0(\xi_1))}}
\]

showing that this is a generalization of the expression obtained by Cheng et al. (2003) for the case in which \(Y_{ij}\) has a Bernoulli distribution with parameter \(\xi_i\) and \(v_i(\xi_i) = \xi_i(1 - \xi_i)\).

3.2 Geometric discounting case

Consider next the second setting introduced above; that of an infinite population with geometric discounting.

In a two-arm trial it is assumed that \(n_1\) and \(n_2\) are sufficiently large and randomisation to treatments 1 and 2 sufficiently balanced that the gain to patients receiving treatment \(i\) in the trial can be taken to be \(n_i \sum_{j=1}^{n_1+n_2} \lambda^{-j} / (n_1 + n_2) h_i(\xi_j)\). The total expected gain if treatment \(i\) is recommended for all further \(N - n_1 - n_2\) patients is then

\[
 \frac{n_1}{n_1 + n_2} \sum_{j=1}^{n_1+n_2} \lambda^{-j} E_{\tilde{Y}|Y}(h_i(\xi_1)|Y) + \frac{n_2}{n_1 + n_2} \sum_{j=1}^{n_1+n_2} \lambda^{-j} E_{\tilde{Y}|Y}(h_2(\xi_2)|Y) + \sum_{j=n_1+n_2+1}^{\infty} \lambda^{-j} E_{\tilde{Y}|Y}(g_i(\xi_j)|Y).
\]
The optimal action at the end of the trial is thus again to treat all future patients with the treatment with the largest value of $E_{\xi,Y}(g_i(\xi_i) \mid Y)$ and the expected gain assuming this action is taken is equal to

\[
\frac{n_1}{n_1 + n_2} \sum_{j=1}^{n_1 + n_2} \lambda^{j-1} E_{\xi,Y}(h_1(\xi_i) \mid Y) + \frac{n_2}{n_1 + n_2} \sum_{j=1}^{n_1 + n_2} \lambda^{j-1} E_{\xi,Y}(h_2(\xi_i) \mid Y) + \sum_{j=n_1 + n_2 + 1}^{\infty} \lambda^{j-1} \max_{i=1,2} E_{\xi,Y}(g_i(\xi_i) \mid Y).
\]

The prior predicted expected utility for a trial with $n_i$ patients receiving treatment $i$ is thus

\[
\frac{n_1}{n_1 + n_2} \sum_{j=1}^{n_1 + n_2} \lambda^{j-1} E_0(h_1(\xi_i)) + \frac{n_2}{n_1 + n_2} \sum_{j=1}^{n_1 + n_2} \lambda^{j-1} E_0(h_2(\xi_i)) + \sum_{j=n_1 + n_2 + 1}^{\infty} \lambda^{j-1} E_0 \left( E_Y \left( \max_{i=1,2} E_{\xi,Y}(g_i(\xi_i) \mid Y) \right) \right).
\]

(13)

Optimal sample sizes, $n_i^*$ and $n_2^*$, can be found directly using this expression and a numerical search in cases when this is computationally feasible.

As above, it is of interest to seek asymptotic approximations to $n_i^*$ and $n_2^*$, in this case as the geometric discounting parameter, $\lambda$, approaches 1 from below. We will again assume that $n_{0i}$ is finite and obtain first an approximation for $n_1^*$.

The derivatives of $\sum_{j=1}^{n_1 + n_2} \lambda^{j-1} = (1 - \lambda^{n_1+n_2})/(1 - \lambda)$ and $\sum_{j=n_1 + n_2 + 1}^{\infty} \lambda^{j-1} = \lambda^{n_1+n_2}/(1 - \lambda)$ with respect to $n_1$ are respectively $-\lambda^{n_1+n_2} \log \lambda/(1 - \lambda)$ and $\lambda^{n_1+n_2+1} \log \lambda/(1 - \lambda)$ which, by L’Hospital’s rule, tend to 1 and $-1$, respectively as $\lambda \rightarrow 1$, so that, since the limit as $\lambda \rightarrow 1$ of $\sum_{j=1}^{n_1 + n_2} \lambda^{j-1}$ is $n_1 + n_2$, the derivative of the term $n_i \sum_{j=1}^{n_1 + n_2} \lambda^{j-1}/(n_1 + n_2)$ with respect to $n_1$ tends to 1 if $i = 1$ and 0 if $i = 2$.

As $\lambda \rightarrow 1$, the derivative of (13) with respect to $n_1$ thus approaches

\[
h_1(\xi_i) - E_Y \left( \max_{i=1,2} E_{\xi,Y}(g_i(\xi_i) \mid Y) \right) + \left( \sum_{j=n_1 + n_2 + 1}^{\infty} \lambda^{j-1} \right) \frac{\partial E_Y(\max_{i=1,2} E_{\xi,Y}(g_i(\xi_i) \mid Y))}{\partial n_1}.
\]

The argument above gives an approximation to this derivative of

\[
h_1(\xi_i) - \max_{i=1,2} g_i(\xi_i) + \\
\sum_{j=n_1 + n_2 + 1}^{\infty} \frac{\lambda^{j-1}}{n_1} \int \frac{v_1 (g_1^{-1}(g_2(\xi_2)))}{2} (g_1^{-1}(g_2(\xi_2))) \pi (g_1^{-1}(g_2(\xi_2)), \xi_2) \, d\xi_2.
\]

which, since \( \sum_{j=n_1+n_2+1}^{\infty} \lambda^{j-1} = \sum_{j=1}^{\infty} \lambda^{j-1} - \sum_{j=1}^{n_1+n_2} \lambda^{j-1} \), with the former term, which is equal to \((1 - \lambda)^{-1}\), dominating, gives the approximation

\[
n_1^* = \sqrt{\frac{\int y_1 \left( g_1^{-1} \left( g_2(\xi_j) \right) \right) g_1' \left( g_1^{-1} \left( g_2(\xi_j) \right) \right) \pi \left( g_1^{-1} \left( g_2(\xi_j) \right) \right)}{2 \left( E_0 \left( \max_{j=1,2} g_1(\xi) \right) - E_0 \left( h_1(\xi) \right) \right)}}.
\]

(14)

Writing \( N^* = \sum_{j=1}^{\infty} \lambda^{j-1} = (1 - \lambda)^{-1} \), (14) can be written as

\[
n_1^* = \sqrt{\frac{\int y_1 \left( g_1^{-1} \left( g_2(\xi_j) \right) \right) g_1' \left( g_1^{-1} \left( g_2(\xi_j) \right) \right) \pi \left( g_1^{-1} \left( g_2(\xi_j) \right) \right)}{2 \left( E_0 \left( \max_{j=1,2} g_1(\xi) \right) - E_0 \left( h_1(\xi) \right) \right)}}.
\]

directly analogous to (10) with \( N^* \) replacing \( N \).

For \( n_{02} \) finite, \( n_2^* \) is again given by symmetry by an expression analogous to (11). For \( n_{02} = \infty \), \( n_2^* = 0 \) and \( n_1^* \) is given by an expression analogous to (12) with \( N \) replaced by \( N^* \).

4 Examples

4.1 Single arm trials with Bernoulli data

We consider first the case of Bernoulli data. In this case the distribution of \( y \) and \( y_j \) is known with value \( y_{02} \), that is with \( n_{02} = \infty \), so that \( n_2^* = 0 \) and a single-arm trial is optimal.

Given observation of data \( Y_{ij} = y_{1j}, j = 1, \ldots, n_1, \xi_1 \) has a Beta \((a_1 + \sum_{j=1}^{n_1} y_{1j}, b_1 + n_1 - \sum_{j=1}^{n_1} y_{1j})\) posterior distribution, and the prior predictive distribution of \( \sum_{j=1}^{n_1} Y_{ij} \) is Betabinomial \((n_1, a_1, b_1)\).

We will assume that the gain from patients receiving treatment 1 will be determined by whether or not the treatment is successful, so that \( g_i(\xi_j) = h_i(\xi_j) = \xi_j, i = 1, 2 \). From (4) with \( n_2 = 0 \), the prior expected gain is thus \( n_1 E_0(\xi) + (N - n_1) E_0(\max_{j=1,2} E_2(\xi_j) | Y) \).

Following one of the examples considered by Cheng et al. (2003), we take \( a_1 = 1 \) and \( b_1 = 1 \), the known value of \( \xi_2 \) to be 0.5 and \( N = 100 \). Figure 1 shows the prior expected gain for a range of values of \( n_1 \), here plotted on a logarithmic scale. As given by Cheng et al. (2003), the optimal value of \( n_1 \) is equal to 9, which is marked with a plus sign. The approximation to the prior expected gain given by (7) is also shown on the figure as a dashed line, showing that in this case the approximate and true values are close even for small \( n_1 \). The approximately optimal value of \( n_1 \) given by (12) is 10. This is close to the value of \( n_1 \) maximizing the approximation given by (7), though in general the additional approximation leading to (12) means that this need not be the case.

Figure 2 gives values of \( n_1^* \) from (12) along with exact optima for a range of values of \( N \), with both \( n_1^* \) and \( N \) plotted on logarithmic scales so that the square root relationship between \( n_1^* \) and \( N \) given by (12) corresponds to a straight line with slope 1/2. The approximation given by (12) is close to the true value, and approaches it as \( N \) increases as would be expected.

The effect of varying the prior distribution for \( \xi_1 \) was also investigated. Web Figure 1 in Web Appendix B shows the optimal sample size \( n_1^* \) with \( N = 100 \) and \( \xi_2 = 0.5 \) for a range of \( n_{01} = a_1 + b_1 \) and \( y_{01} = a_1 / (a_1 + b_1) \) values. When the prior mean value of \( \xi_1 \), that is \( y_{01} \), is equal or greater than the fixed value of \( \xi_2 \), the optimal sample size increases with prior weight \( n_{01} \). This is reasonable since as \( n_{01} \) increases we are increasingly confident that patients are not harmed by being in the trial and a larger sample size gives more information for the final decision. When \( y_{01} \) is less than \( \xi_2 \), the optimal sample size increases with \( n_{01} \) for small \( n_{01} \), so that more information can be collected for the final
Figure 1  Prior expected gain (solid line) and approximate prior expected gain (dashed line) for a range of \( n_1 \) for the first single-arm Bernoulli example with \( a_1 = b_1 = 1, \xi_2 = 0.5, \) and \( N = 100. \) The optimal and approximately optimal values of \( n_1 \) are marked by + and ⋄, respectively.

Figure 2  Optimal (solid lines with points) and approximately optimal (dashed lines) values for \( n_1 \) for a range of \( N \) values for the first single-arm Bernoulli example with \( a_1 = b_1 = 1 \) and \( \xi_2 = 0.5. \)

decision and decreases for larger \( n_{01} \) when there is strong prior belief that treatment 2 is superior to treatment 1. It is interesting to note that when the prior weight, \( n_{01}, \) is kept fixed, the optimal sample size increases as the prior mean \( y_{01} \) of treatment 1 increases. This is in contrast to frequentist sample size that would decrease as the assumed success rate for treatment 1 is increased (and is larger than the success rate of treatment 2).

We next consider an example based on Stallard (1998), who also considered a single-arm phase II trial with a Bernoulli outcome. In this case the gain function was chosen to reflect the financial costs and rewards associated with the conduct of the trial assuming that, if successful, it would be
followed by a further trial with a frequentist design and analysis. Assuming the probability of success for treatment 2, here taken to be the current standard treatment, to be known, costs and rewards were taken relative to continuing to give all patients the current treatment. Thus if this treatment is recommended, the gain to patients outside the trial is taken to be zero, that is \( g_2 = 0 \). The gain per patient outside the trial if the experimental treatment is recommended was taken to be of the form \( g_1(\xi_1) = l(1 - \Phi(z_{a/2} - (z_{a/2} + z_\beta)\theta/\theta_1)) - m \) where \( z_a = \Phi^{-1}(1 - \alpha), \theta = \text{logit}(\xi_1) - \text{logit}(\xi_0) \), \( \theta_1 = \text{logit}(\xi_0 + \delta_0) - \text{logit}(\xi_0) \), and \( \text{logit}(\xi) = \log(\xi/(1 - \xi)) \) for some \( l, \alpha, \beta \), and \( \delta_0 \). This form reflects a fixed cost of \( m \) per patient with a gain of \( l \) per patient if the treatment is shown to be effective in the subsequent trial where that trial has frequentist (one-sided) type I error rate of \( \alpha/2 \) and power \( 1 - \beta \) to detect a log-odds ratio of \( \theta_1 \). Stallard assumed linear discounting for \( j \) less than some constant, \( n_0 \), with geometric discounting for \( j > n_0 \) whereas we will assume geometric discounting for all \( j \). Patients in the trial were taken to have constant (discounted) cost, that is \( h_1(\xi_1) = -k \), for some \( k \). Since we know \( n_2 = 0 \), it is not necessary to specify \( h_2 \). The parameter \( \xi_1 \) was again taken to have a Beta(\( a_1, b_1 \)) prior distribution. As the gain function \( g_2(\xi_2) \) does not depend on \( \xi_2 \), it is not necessary to specify a value for the point-prior for this parameter in this case.

Since \( g_2(\xi_2) = 0 \) for all \( \xi_2 \), the expression (12) becomes

\[
n_1^\ast = \sqrt{-N g_1'(g_1^{-1}(0))v_1(g_1^{-1}(0)) \pi(g_1^{-1}(0)) \over 2(E_0(h_1(\xi_1)) - E_0(\max(g_1(\xi_1), 0)))}
\]

where \( \pi(\xi_1) \) is the (univariate) prior density for \( \xi_1 \). The value of \( E_0(\max(g_1(\xi_1), 0)) \) can be evaluated using numerical integration and, as \( h_1 = -k \), we have \( E_0(h_1(\xi_1)) = -k \).

Although the form of utility function proposed by Stallard (1998) was not exactly that proposed here, based on values given, we took \( k = \exp(-0.00173) = 0.99827, l = 12.79, m = 0.346, \alpha = 0.05, \beta = 0.1, \xi_0 = 0.2, \delta_0 = 0.15, a_1 = 0.845 \) and \( b_1 = 9.155 \). Figure 3 shows the prior distribution for \( \xi_1 \) along with the form of \( g_1(\xi_1) \) in this case. Figure 4 shows the prior expected gain calculated exactly using the betanomial prior predictive distribution for \( \sum_{j=1}^{n_1} Y_{1j} \) for a range of values of \( n_1 \) values, plotted on a logarithmic scale, along with the approximation given by (7). It can be seen that in this case the approximation (7) to the expected gain is rather poor, particularly for smaller \( n_1 \). The value of \( n_1^\ast \) obtained using (15) in this case is 102. This value and the associated prior expected gain is marked on the figure by a plus sign. Note that \( n_1^\ast \) does not maximise the approximate gain given by (7) that is shown by the dashed line on the plot. In this case \( n_1 \) the value of \( n_1^\ast \) is quite far from the value of \( n_1 \) maximizing the approximate gain, though it is closer to the true optimal value of 95, which marked on the figure by a plus sign. Figure 5 gives values of \( n_1^\ast \) from (15) along with exact optima for a range of values of \( N_1 \), again with both plotted on a logarithmic scale. The approximation given by (15) is again close to the true value, and approaches it as \( N_1 \) increases as would be expected.

The effect of varying the prior distribution for \( \xi_1 \) was again investigated, and again illustrated in Web Appendix B. Web Figure 2 shows the optimal sample size \( n_1^\ast \) with \( N = 5000 \) for a range of \( n_{01} = a_1 + b_1 \) and \( y_{01} = a_1/(a_1 + b_1) \) values. In this case as \( y_{01} \) increases from 0.04 to 0.0845 the optimal sample size is increased in a similar way to that noted for the example above. In this case as \( y_{01} \) increases further, however, the optimal sample size is reduced. Here, since \( h_1 = -k \), there is a cost associated with experimentation so that when there is strong prior belief that treatment 1 is superior to treatment 2, rather than giving many patients treatment 1 in a trial, a smaller trial is optimal.

At the end of the trial, treatment 1 will be recommended if \( E_{\bar{Y}_{1Y}}(g_1(\xi_1) \mid Y) > 0 \). For large \( n_1 \), that is approximately if \( E_{\bar{Y}_{1Y}}(\xi_1 \mid Y) > g_1^{-1}(0) \), which, using expression (1), is true when \( n_1(\bar{Y}_1 > g_1^{-1}(0)(n_{01} + n_1) - n_{01}y_{01}) \). Considering recommendation of treatment 1 to correspond to rejection of the null hypothesis that \( \xi_1 = g_1^{-1}(0) = 0.201 \), frequentist error rates attained for the optimal designs obtained can be derived.

Figure 3  Prior distribution for $\xi_1$ (upper panel) and gain function $g_1(\xi_1)$ giving gain from treating each future patient with treatment 1 (lower panel). The gain function $g_2(\xi_2) = 0$ is shown as a dashed line on the right hand panel for comparison (see text for details).

With a prior distribution with $a_1 = 0.845$ and $b_1 = 9.155$, taking $n_1 = 102$ gives a type I error rate of 0.39. The form of $g_1(\xi_1)$ shown in Fig. 3 suggests that we might require a test with high power when $\xi = 0.35$, since this value of $\xi_1$ is associated with a high gain value. The power of the optimal test in this case is 0.999.

Type I and type II error rate values for the optimal designs and prior distributions considered in Web Fig. 2 are shown in Web Fig. 3. As the prior weight becomes small, the optimal decision at the end of the trial is to select treatment 1 whenever it has observed mean exceeding $g_1^{-1}(0)$, so the type I error rate approaches 0.5. As the prior weight increases, since in this case the optimal value of $n_1$ is relatively small, prior information comes to dominate the final decision and the type I error rate approaches zero or one, with the type II error approaching one or zero, depending on whether the prior mean is less than or greater than $g_1^{-1}(0)$.
4.2 A two-arm trial with Poisson data

The third example is based on an example given by Berry et al. (1994), who describe a trial of an HIB vaccine in Navajo children aged 2–18 months. The number of HIB cases is assumed to follow a Poisson distribution. Rather than expressing $n_1$, $n_2$, and $N$ in terms of child-months, we will assume that all children are followed up for the entire 16-month period, and refer to the number of children.
of children in the trial and population. The observed number of cases per child \( j \) in group \( i \), will be denoted by \( Y_{ij} \), \( j = 1, \ldots, n_i \), \( i = 1, 2 \). The distribution of \( Y_{ij} \) can be parameterised such that \( \xi_i \) and \( \xi_2 \) are the expected numbers of cases per child for treatments 1 (the new vaccine) and 2 (the placebo), respectively. Thus \( p(y_{ij} \mid \xi_i) = \xi_i^{y_{ij}} \exp(-\xi_i)/y_{ij}! \), \( j = 1, \ldots, n_i \), \( i = 1, 2 \), so that \( y_{ij} \) has mean \( \xi_i \), with \( \xi_i \) following independent prior gamma(\( \alpha_i \), \( \beta_i \)) distributions, that is with density \( \pi(\xi_i, \xi_2) = \prod_{i=1}^2 \xi_i^{y_{ij}-1} \exp(-\xi_i) \beta_i^{\alpha_i}/\Gamma(\alpha_i) \) for some \( \alpha_i, \beta_i, i = 1, 2 \). Note that \( \xi_i \) has prior mean \( \alpha_i/\beta_i \), and prior variance \( \alpha_i/\beta_i^2 \). The posterior distribution of \( \xi_i \) given \( Y \) is a gamma \((\alpha_i + \sum_{j=1}^{n_i} y_{ij}, \beta_i + n_i)\) distribution and the prior predictive distribution of \( \sum_{j=1}^{n_i} y_{ij} \) is NegBin \((\alpha_i, (1 + n_i/\beta_i)^{-1})\).

Berry et al. (1994) include in their gain function a term that depends on the observed data that reflects the probability of obtaining regulatory approval for the vaccine. Here, we assume the gain from a child receiving treatment \( i \) depends on \( \xi_i \) alone and, since \( \xi_i \) gives the rate of HIB cases, which we would like to minimize, we take gain functions \( h_i(\xi_i) = g_i(\xi_i) = -\xi_i \), \( i = 1, 2 \). The case of gain functions that depend on the observed data is considered briefly in the discussion section below.

The optimal values may be approximated using expressions (10) and (11). In this case \( n_i^* \) is

\[
\sqrt{\frac{N \int \xi \pi(\xi, \xi) d\xi}{2(E_0(\min_{i=1,2} (-\xi_i)) - E_0(-\xi_1))}} = \sqrt{\frac{N \int \xi \pi(\xi, \xi) d\xi}{2(E_0(\xi_1) - E_0(\min_{i=1,2} \xi_i))}},
\]

the integral in the numerator being equal to

\[
\int \xi^{-1-n_i} \exp(-\xi) \beta_1^{\alpha_1} \xi_2^{\alpha_2-1} \exp(-\xi) \beta_2^{\alpha_2} \xi_1^{\alpha_1} \Gamma(\alpha_1) \Gamma(\alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2) \beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1 + \alpha_2) (\beta_1 + \beta_2)^{\alpha_1 + \alpha_2}}.
\]

Following Berry et al. (1994), we take \((\alpha_1, \beta_1) = (1, 200)\) and \((\alpha_2, \beta_2) = (5, 667)\), the latter corresponding to the placebo (note that the \( \beta_i \) values given by Berry et al. are 16 times those used here as they take \( \xi_i \) to give the rate of HIB cases per child-month). Berry et al. report that approximately 5400 Navajo are born each year so that minimization of HIB cases over a 20-year period would correspond to \( N = 108,000 \). Figure 6 shows a contour plot giving the prior expected gain for this \( N \) for a range of \( n_1 \) and \( n_2 \) values (plotted on logarithmic scales) together with the approximation given by (7) (dashed lines). It can be seen that even for small sample sizes, (7) gives a close approximation to the true prior expected gain. The optimal design has \( n_1^* = 3162 \) and \( n_2^* = 1585 \), and is marked by the plus sign. The approximately optimal design given by (10) and (11) has \( n_1^* = 3524 \) and \( n_2^* = 2089 \), and is marked by a circle. The prior expected gain, in this case corresponding to minus one times the prior expected gain, is \( -416.9 \) using the optimal design and \(-417.4\) using the approximately optimal design.

Figure 7 shows the values of \( n_1^* \) and \( n_2^* \) along with the approximations from (10) and (11) (dashed lines) for a range of \( N \) values, again with both plotted on logarithmic scales. It can again be seen how the approximations become increasingly accurate as \( N \) increases.

The effect of varying the prior distributions for \( \xi_1 \) and \( \xi_2 \) was again investigated. Web Fig. 4 in Web Appendix B shows the optimal sample sizes, \( n_1^* \) and \( n_2^* \), for a range of prior means and prior weights, here equal for the two priors, when \( N = 108,000 \). When the prior means are equal, the optimal sample size increases with prior weight. For unequal prior means, more patients are assigned to the arm considered \textit{a priori} to be superior, in this case corresponding to a lower prior mean since \( h_i(\xi_i) = g_i(\xi_i) = -\xi_i \), with the number assigned to the inferior arm increasing with prior weight when this is small and decreasing for larger prior weight values. Web Fig. 5 shows the effect of changing the prior weight for \( \xi_1 \) alone when the prior mean is equal to, greater than or less than that for \( \xi_2 \). In this case increasing the prior weight leads to an increase in the optimal sample size for both arms, with the arm with the lower prior mean having a smaller optimal sample size.
**Figure 6** Contour plot of prior expected gain (solid lines) and approximate prior expected gain (dashed lines) for $N = 108,000$ for a range of $n_1$ and $n_2$ values assuming gamma (1, 200) and gamma (5, 667) prior distributions. The optimal and approximately optimal values of $n_1$ and $n_2$ are marked by $+$ and $\circ$, respectively.

**Figure 7** Optimal (solid lines with points) and approximately optimal (dashed lines) values for $n_1$ (upper lines) and $n_2$ (lower lines) for a range of $N$ values assuming gamma (1, 200) and gamma (5, 667) prior distributions.

The examples above compare the large sample approximation for the optimal sample size for arm $i$, $n_i^*$, given by expression (10), with that obtained by exact numerical optimization in two examples, enabling assessment of the approximation for smaller values of $N$. The derivation of (10) relies on large sample approximations in two ways; first the distribution of the posterior expected utility $E(g_i(\xi_i) \mid \mathbf{Y})$
is approximated by its asymptotic normal form in (5) using the central limit theorem and the delta method, and second, the derivative of the expected gain, given by (8) is approximated by (9). The first approximation is exact when \( Y_{ij} \) are normally distributed and \( g_i(\xi_i) \) are linear. The first and second examples suggest that both approximations are sufficiently accurate for Bernoulli data even for quite small \( N \) when \( g_i(\xi_i) \) are linear, but less so for nonlinear \( g_i(\xi_i) \), when the first approximation may be poor, as noted by Bernado and Smith (2000). For nonnormal data the accuracy of the first approximation also improves as the prior weight \( N_0 \) increases, though, as illustrated in the third example above, this also leads to smaller \( n^*_i \), so that the overall accuracy of the asymptotic approximation to the optimal sample size might be poorer.

### 5 Discussion

The work reported above leads to expressions for the optimal sample size in a clinical trial to compare two treatments or to compare a single experimental treatment with a historical control with a trial and to compare a single experimental treatment with a historical control. The optimal sample size of one parameter exponential family form, with mean \( \xi_i \), assumed to have a conjugate prior distribution. The forms of the utility functions \( h_i(\xi_i) \) and \( g_i(\xi_i) \), giving the gain to patients inside and outside the trial who receive treatment \( i \). Although the expressions obtained could be directly used to design a trial, it is also of more general interest to see how the optimal sample size depends on the size of the population under investigation. We have shown that if the population is assumed to be of some known size, \( N \), for any \( h_i \) and \( g_i \), satisfying sufficient regularity conditions for expectations to exist, differentiable with \( g_i \) strictly increasing, and satisfying the condition given by (2); that is \( E_0(h_i(\xi_i)) \leq E_0(\max_{i=1,2}(g_i(\xi_i))) \), the optimal sample size is \( O(N^{1/2}) \) as \( N \to \infty \). If it is assumed that there is an infinite population with geometric discounting with discounting factor \( \lambda \), under the same conditions the optimal sample size is \( O(N^{\gamma/2}) \) as \( N^* \to \infty \) where \( N^* = (1-\lambda)^{-1} \). This extends previous work by Cheng et al. (2003).

Although we have considered general functions \( h_i(\xi_i) \) and \( g_i(\xi_i) \), giving the gain to patients inside and outside the trial who receive treatment \( i \). We have assumed that these are functions of \( \xi_i \) only. This is a common assumption and it seems reasonable that the benefit to a patient from taking a given treatment will depend only on the properties of that treatment (see, e.g. Lindley, 1997, who cites the seminal work by Raiffa and Schlaifer, 1961). Noting, however, that the gain functions \( g_i \) correspond to gain from future patients if the trial indicates that treatment \( i \) is superior, some authors have proposed gain functions for patients treated following the trial that depend, in addition to \( \xi_i \), on the observed trial data, \( Y \). In particular, gain functions have been proposed that reflect the fact that use of a novel treatment following a trial may depend on regulatory decisions that in turn depends on whether trial results are sufficiently compelling (see, e.g. Posch and Bauer, 2013). In both of the more realistic examples described above, the gain functions given by Stullard (1998) and Berry et al. (1994) depended on \( Y \), and we have simplified the gain functions when discussing these examples above.

The forms of the utility functions \( h_i \) and \( g_i \), \( i = 1,2 \) given above were motivated by consideration of the gain to each patient from participation in the trial or from being treated with treatment \( i \) following the trial, suggesting that the trial sample size is optimised from the patient’s perspective. The general form of the expected gain given by (3), however, can express any gain so long as this can be specified on a per-patient basis. The results obtained could thus also apply to financial gains from a commercial perspective or to societal gains from development of a novel therapy. In the latter cases it may be more appropriate for \( h_i \) and \( g_i \) to have a more complex form or to depend on trial data as discussed above.

The condition (2) ensures that the gain per patient in the trial does not exceed that per patient outside the trial if patients were to receive optimal treatment. If this does not hold the optimal design will be to continue with trial forever, giving all patients the treatment for which the prior expected gain \( E_{0i}(h_i(\xi_i)) \) is the largest. This restriction on \( h_i \) and \( g_i \) seems reasonable if \( h_i \) reflects not only the
benefit to patients in the trial receiving treatment \( i \), but also the cost of the trial, either in financial terms for the trial sponsor or funder or in terms of commitment by the patient, both of which may be considerable.

It is interesting to compare the optimal sample sizes obtained above with sample sizes typical for clinical trials. In particular, it might be of interest to consider the size of population for which conventional sample sizes would correspond to that of the optimal design. A method for frequentist sample size calculations for a single arm trial with a Bernoulli response is given by Fleming (1982), who shows that the sample size required for a trial with (one-sided) type I error rate \( \alpha \) and power \( 1 - \beta \) to detect an improvement to a success probability of \( p_1 \) from a control success probability of \( p_0 \) as \( p_1 \to p_0 \) is \( 
\frac{\sqrt{p_0(1-p_0)} \Phi^{-1}(1-\alpha) + \sqrt{p_1(1-p_1)} \Phi^{-1}(1-\beta)}{(p_1 - p_0)^2} \). As discussed above, the form of \( g_1(\xi_1) \) given in the second single arm Bernoulli data example above and shown in Fig. 3 suggests that an appropriate value for \( p_1 \) might be about 0.35, since this value of \( \xi_1 \) is associated with a high gain value. For \( \alpha = 0.025 \) and \( \beta = 0.9 \), this would give a sample size of 111. Using the gain function described above, this would be optimal for a population of size, or expected population size in the case of geometric discounting, of about 3000. The prior distribution used in the example above, and also shown in Figure 3 is such that a value of \( \xi_1 \) as large as 0.35 is highly unlikely, suggesting that a smaller value could be used for \( p_1 \) in the frequentist sample size calculation. The 95th percentile of the prior distribution is 0.256. To give power of 0.9 to detect a treatment effect corresponding to this value of \( p_1 \) would require a sample size of 704. This would be optimal for an expected population of size of a little over 100,000. It is important to note that even when the sample sizes are similar, the optimal designs obtained above may be very different from those obtained using the usual frequentist approach as, following the trial, a treatment is recommended depending on the posterior expected gains rather than on the basis of type I error rate control. As seen above, depending on the prior distribution, this can lead to type I error rates considerably higher than those conventionally used in large-scale confirmatory studies. In this regard, the designs obtained are more similar to those sometimes used for early-phase clinical trials or pilot studies (Schoenfeld, 1980; Stallard, 2012). Further comparison of frequentist and decision-theoretic approaches is an area where further research would be of interest.

Acknowledgement  The authors are grateful to two anonymous referees for their helpful comments. The work was conducted as part of the InSPiRe (Innovative methodology for small populations research) project funded by the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-602144.

Conflict of interest
The authors have declared no conflict of interest.

References


