Estimating and maximising the expected utility of pilot trials using surrogate decision rules

Journal Title
XX(X):2–14

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Duncan T. Wilson[1]

Abstract

Keywords
Clinical trial, pilot trial, external pilot, statistical decision theory, optimal design, expected value of sample information, Monte Carlo methods

1 Introduction

One approach to choosing the sample size of a clinical trial is through Bayesian decision theory. Given an appropriate utility function, we can choose the same size which maximises the expected utility with respect to a prior distribution on all unknown parameters. These parameters will generally include a treatment effect and one or more so-called nuisance parameters. For example, the standard deviation of a normally distributed outcome. The more that is known about these parameters prior to designing the trial, the more able we will be to make good design decisions.

When there is substantial uncertainty regarding the parameters prior to the planned definitive trial, a small pilot trial can be run to obtain some relevant information. Many methods have been proposed to design these pilot trials and to guide how their data should be used to inform the definitive trial, when the goal is to estimate a nuisance parameter.[2 3 4 5 6 7 8 9 10]

Pilot trials have also been used to learn about the effectiveness of the intervention. This use of pilot trials is more controversial, due to the possible issues arising from the small sample size and potential bias in pilot estimates. In particular, conducting a test of effectiveness using pilot data has been advised against as such tests will generally have low power. However, an analysis

[1]Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK

Corresponding author:
Duncan T. Wilson, Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, LS2 9JT, UK
Email: d.t.wilson@leeds.ac.uk

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

of a Bayesian decision-theoretic approach to setting the parameters of such tests has suggested that they can be useful, and that not testing will rarely be optimal. [11 12 13 14 15 16 17]

Given the small sample size of pilot trials and the subsequent inability to make any firm conclusions or provide very precise parameter estimates under a frequentist model, a Bayesian approach to their design and analysis is attractive. Given appropriate prior distributions describing initial knowledge and uncertainty in any unknown parameters, the pilot data can be used to update these and provide posterior distributions, which can in turn be used to determine the optimal design of the definitive trial (including the decision of whether or not to run the trial at all) under a Bayesian statistical decision framework (which requires the specification of a utility function). The pilot trial itself can then be designed so as to maximise the expected utility of the pilot-definitive trial programme. This is also known as maximising the expected value of sample information. [18 19 20 21 22 23 24 25]

A practical consideration when taking this approach is the computational challenge it presents. Other than in very special cases involving conjugate priors, simulation-based methods are generally required. A common approach involves a two-level nested MC routine, where pilot data are simulated and for each data set, a set of posterior samples is generated using e.g. MCMC and this is then used to find the optimal terminal decision (in our case, the optimal definitive trial design). As this can take weeks to run, several researchers have developed fast approximations, generally based on using some kind of regression to approximate the expected utility, conditional on the pilot data sufficient statistics, of each decision. This approach may prove difficult if we have a large space of decisions (e.g. a large potential definitive trial sample size) or a continuous one (e.g. the alpha of the definitive trial). We could discretise the space, to get the speed but at the cost of accuracy. [26 27 28 29 30 31]

We propose a different approach, where we recognise that an MEU analysis of pilot data will map the data to corresponding optimal definitive trial designs, i.e. it will provide a decision rule. Moreover, this decision rule is optimal in the sense that it will have the largest expected utility. We will show that because we can estimate the expected utility of any decision rule, we can solve the problem by considering surrogate rules and search over these, knowing that the optima will coincide with MEU.

The remainder of this paper is structured as follows...

2 Problem and motivating example

Consider settings where an external pilot is to be conducted in advance of a definitive trial. Denote the per-arm sample size of these trials by n_1 and n_2 respectively. The primary analysis of the definitive trial will be a conventional hypothesis test comparing the control and intervention arms, using a typical type I error rate of $\alpha = 0.05$ (two-sided). We assume that the power of the definitive trial, conditional on θ, can be calculated. The pilot trial data is to be used to inform the choice of sample size for the definitive trial.

We work in a Bayesian-frequentist hybrid framework. Denoting the unknown parameters by θ, we specify a prior distribution $p(\theta)$ before the pilot trial is conducted. The pilot data, denoted x_1, is then used to obtain the posterior $p(\theta \mid x_1, n_1)$. The sample size of the definitive trial is then chosen so as to maximise expected utility with respect to this distribution.

Given this procedure, the only decision to be made is the choice of pilot sample size n_1. We do this by maximising the expected utility of the full pilot-then-definitive trial programme, with respect to the prior $p(\theta)$. Aside from the operational issues of specifying an appropriate prior and utility function, the main methodological issue is calculating the expected utility of a programme with pilot sample size n_1.

2.1 Motivating example - REACH

REACH was two-arm parallel group cluster randomised external pilot, where each cluster was a care home. REACH aimed to recruit $n_1 = 6$ care homes into each arm. The number of residents from each care home subsequently recited into the trial was random, but with some prior idea in terms of the average number and the variation in it between care homes.

The definitive trial was to be designed to detect an effect of $\mu = 0.2$, using the standard design effect formula for cluster randomised trials with uncertain cluster sizes:

$$n_2 = n[1 + (\bar{m} + cv^2 - 1)\rho].$$

A more direct way to state the definitive trial sample size calculation notes that the intended analysis is a z-test of cluster means. We can therefore focus on the variance of cluster means as the nuisance parameter of interest, and use a standard sample size formula. That is, rather than decompose the variance into its various components, we can estimate it directly from the pilot data.

We specify the prior using the decomposition. Specifically, we put an inverse normal gamma joint prior on the mean and variance of the cluster sizes. We then use an inverse gamma prior for the within-cluster variance σ_w^2, and a beta prior for the intracluster correlation ρ. Together, these imply a prior for the nuisance parameter of interest, the cluster mean variance σ_c^2. This is not a conjugate prior for the corresponding sampling model (where the estimate follows a scaled chi-squared distribution), and so we would need to use some numerical sampler when analysing the pilot data and computing the posterior.

Given this prior on $\theta = (\mu, \sigma_c^2)$, we set up a utility function as in WP3.1.

3 Methods

Utility is a function of the true parameter value, the sample size used at each stage, and the data obtained at each stage (see WP3.1) written $u(\theta, n_1, x_1, n_2, x_2)$. After observing pilot data x_1 trial we will choose the definitive trial sample size n_2 so as to maximise expected utility with respect to the posterior $p(\theta \mid x_1, n_1)$:

$$\max_{n_2} \mathbb{E}_{\theta|x_1,n_1}[u(\theta, n_1, x_1, n_2, x_2)].$$

The expected utility of the full programme with a given pilot sample size n_1 is then

$$\mathbb{E}_{x_1|n_1}\left(\max_{n_2} \mathbb{E}_{\theta|x_1,n_1}[u(\theta, n_1, x_1, n_2, x_2)]\right). \tag{1}$$

We then want to choose n_1 to maximise this value. Writing the problem in terms of the corresponding integrals,

$$\max_{n_1} \int \left(\max_{n2} \left[\int \mathbb{E}_{x_2|\theta}[u(\theta, n_1, x_1, n_2, x_2)] p(\theta \mid x_1, n_1) d\theta \right] \right) p(x_1 \mid n_1) dx_1.$$

Given the utility function described in Section 2, the expectation term can be calculated analytically. We consider the most general case where the posterior distribution $p(\theta \mid x_a, n_1)$ is not known in closed form, resulting in a Monte Carlo approximation of the inner integral being required. This implies that the inner maximisation cannot be solved analytically, requiring numerical methods instead. The marginal distribution $p(x_1 \mid n_1)$ may often be available, we do not assume it has a sufficiently convenient form to allow for a quadrature type method for the outer integral. The outer maximisation will again require numerical methods.

3.1 Two-level nested Monte Carlo

We can solve the problem by estimating the expected utility (1) for a range of possible n_1 choices, possibly in the context of a simple optimisation algorithm such as the bisection method. For any given n_1, expected utility can be estimated using a two-level nested Monte Carlo routine as described in Algorithm 1.

Algorithm 1 Two-levl nested Monte Carlo

1: **for** $1 = 1, 2, \ldots N$ **do**
2: Sample $\theta^{(i)} \sim p(\theta)$
3: Sample pilot data $x^{(i)} \sim p(x \mid \theta^{(i)}, n_1)$
4: Draw M samples from the posterior, $\theta^{(j)} \sim p(\theta \mid x^{(i)}, n_1)$
5: Find $n_2^{(i)} = \arg\max_{n_2} \frac{1}{M} \sum_{j=1}^{M} u(\theta^{(j)}, n_1, x_1^{(i)}, n_2)$
6: **end for**
7: Calculate $\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{x_2|\theta^{(i)}, n_1, x_1^{(i)}, n_2^{(i)}}[u(\theta^{(i)}, n_1, x_1^{(i)}, n_2^{(i)})]$

The algorithm involves first sampling N draws $\theta^{(i)}, i = 1, \ldots, N$ from the prior distribution $p(\theta)$, and for each of these sampling a corresponding pilot data $x^{(i)} \sim p(x \mid \theta^{(i)}, n_1)$. Then, for each data set we use a numerical Bayesian sampler (e.g. Stan) to obtain a set of samples from the posterior distribution $p(\theta \mid x^{(i)}, n_1)$. We then use these to determine the optimal sample size of the definitive trial. Finally, we average over the N samples in the outer loop to obtain a Monte Carlo estimate of the expected utility of the full programme.

Algorithm 1 will produce an unbiased estimate of the true expected utility, with a known standard error that will reduce at a rate of $1/\sqrt{(}N)$, under the assumption that the sampling and optimisation of steps 4 and 5 are sufficiently robust. It is computationally intensive, particularly because for each of the N outer loop iterations we are required to run a numerical sampler and then solve an optimisation problem.

3.2 Surrogate decision rules

Algorithm 1 involves finding an optimal definitive trial sample size for a number of simulated pilot data sets (see steps 4 and 5), and then using these n_2's to estimate expected utility. Here we consider an alternative approach where steps 4 and 5 are replaced by a pre-specified decision rule $d(x_1) : \mathcal{X}_1 \to N$. Note that

$$\max_{n_2} \mathbb{E}_{\theta|x_1,n_1}[u(\theta, n_1, x_1, n_2, x_2)] \geq \mathbb{E}_{\theta|x_1,n_1}[u(\theta, n_1, x_1, d(x_1), x_2)]$$

for all d. So, we can recast the problem as one of finding the optimal decision rule d, knowing that the global optima coincides with equation (1). We can re-write the expected utility of a pilot trial with sample size n_1 and decision rule d as

$$\begin{aligned}\mathbb{E}_{x_1|n_1}\left[\mathbb{E}_{\theta|x_1,n_1}[u(\theta, n_1, x_1, d(x_1), x_2)]\right] &= \mathbb{E}_{x_1,\theta|n_1}\left[\mathbb{E}_{x_2|\theta,x_1,n_1}[u(\theta, n_1, x_1, d(x_1), x_2)]\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{x_2|\theta^{(i)},n_1,x_1^{(i)}}[u(\theta^{(i)}, n_1, x_1^{(i)}, d(x_1^{(i)}), x_2)], \quad (2)\end{aligned}$$

where $\theta^{(i)}, x^{(i)} \sim p(\theta, x)$. As noted in Section 2, we are considering problems where we have a power function for the definitive trial and as such the expectation in th summation can be calculated quickly and exactly. This avoids the computationally intensive nested Monte Carlo of algorithm 1, at the expense of requiring equation (2) to be calculated repeatedly in the context of a search over a space of decision rules.

3.2.1 Splines We consider decision rules defined over a space of cubic spline functions. Considering first the case where the pilot data can be summarised by a scalar sufficient statistic, we set the spline knots at even quantiles of the samples $x_1^{(1)}, \ldots, x_1^{(N)}$ (with the number of knots to be chosen). Given these pre-specified knots, we can obtain the the set of $k + 4$ cubic splines which serve as a basis for the space of all cubic splines defined over the same knots. Denoting these so-called b-splines by $\phi_i(x_1), k = 1, \ldots, k + 4$, a decision rule is then

$$d(x_1) = \left\lfloor \sum_{i=1}^{k+4} a_i \phi_i(x_1) \right\rfloor,$$

where we apply the floor operator $\lfloor . \rfloor$ to convert the continuous spline function into the required integer output. Searching over the space of decision rules then amounts to searching over the space of vectors $\mathbf{a} = (a_0, a_1, a_2, \ldots, a_{k+3})$. We can solve this problem using an appropriate optimisation algorithm, such as Subplex [32] as implemented in R via the package 'nloptr' [33].

For sufficient statistics of dimension ≥ 2, we can apply the same approach using tensor product splines. For example, we can deal with the two statistics x_1 and y_1 by first constructing the b-splines for each one dimensional space, $\phi_i(x_1), k = 1, \ldots, k + 4$ and $\psi_i(x_1), k = 1, \ldots, k + 4$ respectively, and then using decision rules of the form

$$d(x_1, y_1) = \left\lfloor \sum_{i=1}^{k+4} \sum_{j=1}^{k+4} a_{i,j} \phi_i(x_1) \psi_j(y_1) \right\rfloor.$$

Although this approach can be extended in theory to any number of dimensions, in practice we find we are limited to two or three at most due to the explosion in number of coefficients, and therefore in the dimensions of the optimisation problem's solution space, resulting from the nested summations. For example, using b-splines defined over $k = 6$ knots in three dimensions requires $(k + 4)^2 = 1000$ coefficients. When solving problems in two dimensions, we use the large-scale optimisation algorithms implemented in the R package 'BB'[34].

3.2.2 Optimising n_1 Although we could use the above to estimate the expected utility for a range of n_1 and choose the maximiser as our pilot sample size, an alternative is to incorporate n_1 into the optimisation problem **??** as a solution variable. Thus, we now search over the joint space of decision rules and pilot sample size to find that with maximum expected utility. The addition of an extra solution variable does not have a large impact on the difficulty of solving the problem; and we would expect it to lead to efficiencies over the alternative by allowing information regarding the form of decision rule to be shared, in the sense that we would expect similar optimal decision rules for similar n_1s.

In practice, we execute this approach by first simulating a sequence of pilot data sufficient statistics for each sampled parameter $\theta^{(i)}$, giving a set of samples $\mathbf{x}^{(m)}, m = 0, 1, \ldots, n_1^{max}$. For each $\mathbf{x}^{(m)}$ we must set up the b-spline basis functions. When optimising we treat n_1 as a continuous variable, and estimate the expected utility for any non-integer n_1 as the linear combination of the expected utility

A further advantage of the surrogate decision rule approach over the nested MC method is that optimisation over the pilot sample size, n_1, can be incorporated directly into the existing optimisation problem. This leads to efficiencies since it effectively allows some sharing of information in the sense that a optimal decision rule for some n_1 will likely look similar to the optimal rule for some other n_1', depending on how close they are.

3.3 Nonparametric regression

An alternative approach has been suggested in the context of calculating the expected value of information for a definitive trial[27]. Applying to our problem, the method involves considering each decision which could be made after the trial (in our case, each definitive trial sample size) and using a nonparametric regression model to predict expected utility based on the pilot data x. Specifically, we begin as before by generating a set of joint samples from $p(\theta, x)$, and then calculate the associated vector of utilities $u_k^{(i)}$ for each definitive sample size $k = 0, 1, \ldots, n_{max}$. Then, for each k, we regress $u_k^{(i)}$ on $x^{(i)}$. The mean regression line corresponds to the expected utility of decision k, conditional on observing x. Given this set of regression models, we then proceed to estimate the conditional expected utility for each k for each $x^{(i)}$ and choose the k with the largest of these. Thus, we have mapped our samples $x^{(i)}$ to decisions, and can estimate expected utility as before. In order to choose the optimal pilot sample size n_1, we repeat the full process in a manner similar to that described in Section 3.1.

It is suggested that cubic splines should be used in the regressions, via the 'gam' function in the R package 'mgcv'. These can be fitted to large data sets relatively easily, but the large number of decsisons here (as opposed to the two or three motivating the method originally) may lead to a considerable time requirement. Moreover, the standard errors of the esimtated expected

utility will be larger using this method than either of the two previously described, meaning that a larger number of samples will be needed to attain the same precision.

3.4 Standard errors

Need to quantify the uncertainty in our estimates of EVSI.

For the surrogate approach, this is easy - conditional on the rule, we just have MC error from the outer loop. So if we use the rule(s), we have a good estimate. If we plan on doing an MCMC analysis and subsequent optimisation, the rule will be biased to the extent allowed by our optimisation. We could potentially measure this using some nested MC runs.

For the nested MC approach this is easy enough too - we need to assume that the MCMC routine is reliable enough (in particular, M is large enough) to get the right choice of n_2. Then all uncertainty comes from the outer loop, as above. Note that we might still get a biased answer - only unbiased if the repeated MCMC runs are exactly as we will use in the pilot analysis (same model, M, and random seed). In reality we may do some iteration as we explore the model diagnostics and adjust e.g. the parametrisation or the sampling routine to get a better sample.

3.5 Implementation

For efficiency, we evaluate each of the b-splines at each of the $x_1^{(i)}$ sampled points and storing these in a matrix B. the vector of the definitive sample sizes corresponding to some rule a is then just Ba.

When using MC samples for the integration, we use sample path optimisation - increasing the number of samples, and using the previous optima as the starting point in the search.

Add a penalty term to ensure decision rule gives positive n_2 - means that we can't calculate the gradient.

4 Illustrative applications

For all examples we will use the same basic utility set-up. Following WP3.1, we use an exponential utility (with risk parameter ρ) based on an additive value function with three attributes: the total sample size, the change in average primary endpoint, and the cost of treatment.

For each example we will compare with the nested method, which we may be able to shortcut in some way. We want to compare performance for different outer samples N. For the surrogate method, we may want to see how performance changes if we use $N = 10^6$ for the final expectation, but use only a subset for the building of the rule. Equivalently, ask whether we can get the extra precision of a high N but without the bias of a low rule-building N.

For all examples, compare methods in terms of estimating the utility of a programme for a fixed n_1. After this, show how we can also optimise n_1 when using the surrogate approach. Don't need to implement the bisection search using the nested MC or GAM approach - can just give an indication of runtime based on complexity bounds for a simple bisection search.

4.1 Normal endpoint with known variance (OK-Diabetes)

To make the ideas clear, apply the new method to a simple hypothetical example which we can solve (almost) exactly using method 1 - a normal prior on treatment effect, known variance.

Use the three methods to optimise n_1. Use a bisection search for the last two methods. Report the suggested sample size with the associated estimated utility and the standard error in it. For the optimal n_1, illustrate the method more fully by showing the decision rules given by each method, and showing how the spline function is formed.

Do this for $10^4, 10^5$ and 10^6 samples.

4.2 Binary endpoint

Parametrised with a beta prior on the control arm probability, and a normal prior on the log odds ratio. Utility is based on the absolute difference between control and intervention. Sufficient statistics are counts in each arm.

Although we do require MCMC here since the implied prior on the intervention rate is not conjugate, the computational burden is much less and scales well with the number of outer samples. This is because the possible outcomes of the pilot are finite due to the binary outcome - n_1^2 at most. And generally much less - we can simulate the N pilot data first and then just extract the unique outcomes. For example, taking $n_1 = 100$ and $N = 10^6$ led to only 2250 unique outcomes. So, we can do MCMC and optimisation for these, and then use these repeatedly over the full N sample.

Still of interest to compare the two methods, as the surrogate rule approach will still be faster. And it means we can do a gold standard comparison without it taking weeks.

[35] consider a similar problem but with a single arm pilot and a fixed phase III sample size (so the decision is just stop/go). Same with[36].

4.3 Normal endpoint with clustering (REACH)

Setting: a cRCT with fixed m but unknown total variance and unknown ICC; priors on these as following Speigelhalter 2000, giving a non-conjugate prior on the variance of cluster means; normal prior on treatment effect. For the definitive trial sample sizing we are interested only in the cluster mean variance and effect size, so a 2D summary statistic.

Break this down into two examples. First, we consider learning only about the cluster mean variance. This corresponds to assuming the effect at stages 1 and 2 are independent, which we can think of as an extreme version of the 'pilots are biased' argument. In this case we have a 1D summary stat.

Next, look at the 2D problem where we measure and learn both the variance and the effect size.

For both sub-problems, we will need to use the full nested MC approach as the comparison, which will take days/weeks to run.

4.4 Variance components for a cRCT

Consider designing a cRCT, where we need to choose both the number of clusters k and the size of each cluster m. We us the pilot data to learn about the variance components σ_w^2 and

σ_b^2, but not about the effect size. So we have two dimensions to our sufficient statistic, and two dimensions to the optimal design.

We will need to use the full nested MC approach for comparison.

To use the nonparametric regression approach, we need to build 2D gam models for the range of ks and ms; so maybe of the order 1000 models if using full granularity.

5 Results

6 Extensions

6.1 Optimising definitive trial α

Should just be a computational issue, as it will double the solution space when searching for decision rules as we now need two rules, one for n and one for alpha.

6.2 Internal pilots

Main thing is to control the overall type I error. Can't do this analytically, but for any decision rule we can use a sample under the null to estimate alpha and use this as a constraint in our optimisation. Will need to optimise over the main trial critical value here, so follows on from the above - but note that we don't have to find an optimal decision rule for alpha. Rather, we just need to search over the definitive nominal alpha, and penalise if the actual alpha exceeds the threshold.

We can go beyond a single interim analysis. Suppose we have k interim analyses, at each stage deciding the sample size of the next stage. At the final stage, the chosen sample size will be recruited and the full sample will be used in a frequentist test, and we can control the actual type I error as above. Thus, we need to find k decision rules rather than just one.

What about a more flexible sequential design? The standard approach finds efficacy and futility bounds at each sampling stage. We could do standard forward sampling here to find these bounds, for given stages. But, we could allow not just stopping, but choice of next stage sample size. This would need some kind of ineterim analysis penalty. Could link this to time - doing an interim will stop recruitment for the length of the endpoint. Would then need to tradeoff effect obtained by time we obtain it, so for our utility we would need linear value in time and preference independence - seems reasonable. And this would correspond with a fixed additive penalty for each analysis. In the simplest case of known variance, we would then want to find a smooth surface over sample mean and sample size giving the next stage sample size, together with two functions representing efficacy and futility stopping boundaries (mapping sample size to sample mean in each case).

7 Discussion

Assuming that we have a power function for the definitive trial - could relax, more computations.

Form of utility very flexible.

Our method involves, at each iteration, calculating the utility of a definitive trial from a set of θ, n_2 pairs. The θs are the same each time, but the n_2's vary. For us, this is not a problem as the

utility calculation is simple; but for complex health economic models it may be computationally demanding. The HE literature often talks about generating one PSA sample and not requiring any more - our method has this property, but may limit it since here we have a large set of decisions, not just 2 or three. Perhaps a key distinction between this and the HE context - utilities generally quite simple? But need to show that in this simple setting, the HE methods are not (uniformly) better than the proposal. Note that [29] notes that [27] will tend to be better for low dimensional summary stats, so we can assume this is the competitor for our problem.

...

Acknowledgements

Acknowledgements.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

Funding

This work was supported by the Medical Research Council [grant number MR/N015444/1].

References

1. Browne RH. On the use of a pilot sample for sample size determination. *Statistics in Medicine* 1995; 14(17): 1933–1940. DOI:10.1002/sim.4780141709. URL `http://dx.doi.org/10.1002/sim.4780141709`.
2. Teare M, Dimairo M, Shephard N et al. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials* 2014; 15(1): 264. DOI: 10.1186/1745-6215-15-264. URL `http://www.trialsjournal.com/content/15/1/264`.
3. Julious SA and Owen RJ. Sample size calculations for clinical studies allowing for uncertainty about the variance. *Pharmaceutical Statistics* 2006; 5(1): 29–37. DOI:10.1002/pst.197. URL `http://dx.doi.org/10.1002/pst.197`.
4. Whitehead AL, Julious SA, Cooper CL et al. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research* 2015; DOI:10.1177/0962280215588241. URL `http://smm.sagepub.com/content/early/2015/06/27/0962280215588241.abstract`. `http://smm.sagepub.com/content/early/2015/06/27/0962280215588241.full.pdf+html`.
5. Eldridge SM, Costelloe CE, Kahan BC et al. How big should the pilot study for my cluster randomised trial be? *Statistical Methods in Medical Research* 2015; DOI:10.1177/0962280215588242. URL `http://smm.sagepub.com/content/early/2015/06/12/0962280215588242.abstract`. `http://smm.sagepub.com/content/early/2015/06/12/0962280215588242.full.pdf+html`.
6. Lake S, Kammann E, Klar N et al. Sample size re-estimation in cluster randomization trials. *Statistics in Medicine* 2002; 21(10): 1337–1350. DOI:10.1002/sim.1121. URL `http://dx.doi.org/10.1002/sim.1121`.

7. Grayling MJ, Mander AP and Wason JMS. Blinded and unblinded sample size reestimation procedures for stepped-wedge cluster randomized trials. *Biometrical Journal* 2018; 60(5): 903–916. DOI:10.1002/bimj.201700125. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201700125`. `https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201700125`.
8. Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: A review. *Journal of Biopharmaceutical Statistics* 2005; 15(4): 559–574. DOI:10.1081/bip-200062852. URL `http://dx.doi.org/10.1081/BIP-200062852`.
9. S S and M M. Re-estimating sample size in cluster randomised trials with active recruitment within clusters. *Statistics in Medicine* 2014; 33(19): 3253–3268. DOI:10.1002/sim.6172. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6172`. `https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.6172`.
10. Wittes J and Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine* 1990; 9(1-2): 65–72. DOI:10.1002/sim.4780090113. URL `http://dx.doi.org/10.1002/sim.4780090113`.
11. Cocks K and Torgerson DJ. Sample size calculations for pilot randomized trials: a confidence intervalÂ approach. *Journal of Clinical Epidemiology* 2013; 66(2): 197 – 201. DOI:http://dx.doi.org/10.1016/j.jclinepi.2012.09.002. URL `http://www.sciencedirect.com/science/article/pii/S0895435612002740`.
12. Lee JJ and Chu CT. Bayesian clinical trials in action. *Statistics in Medicine* 2012; 31(25): 2955–2972. DOI:10.1002/sim.5404. URL `http://dx.doi.org/10.1002/sim.5404`.
13. Sim J. Should treatment effects be estimated in pilot and feasibility studies? *Pilot and Feasibility Studies* 2019; 5(1). DOI:10.1186/s40814-019-0493-7.
14. Lancaster GA, Dodd S and Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice* 2004; 10(2): 307–312. DOI:10.1111/j..2002.384.doc.x. URL `http://dx.doi.org/10.1111/j..2002.384.doc.x`.
15. Arain M, Campbell M, Cooper C et al. What is a pilot or feasibility study? a review of current practice and editorial policy. *BMC Medical Research Methodology* 2010; 10(1): 67. DOI:10.1186/1471-2288-10-67. URL `http://www.biomedcentral.com/1471-2288/10/67`.
16. Thabane L, Ma J, Chu R et al. A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology* 2010; 10(1): 1. DOI:10.1186/1471-2288-10-1. URL `http://www.biomedcentral.com/1471-2288/10/1`.
17. Stallard N. Optimal sample sizes for phase II clinical trials and pilot studies. *Statistics in Medicine* 2012; 31(11-12): 1031–1042. DOI:10.1002/sim.4357. URL `http://dx.doi.org/10.1002/sim.4357`.
18. Lindley DV. A statistical paradox. *Biometrika* 1957; 44(1/2): 187. DOI:10.2307/2333251. URL `http://dx.doi.org/10.2307/2333251`.
19. Gittins J and Pezeshk H. How large should a clinical trial be? *Journal of the Royal Statistical Society: Series D (The Statistician)* 2000; 49(2): 177–187. DOI:10.1111/1467-9884.00228. URL `http://dx.doi.org/10.1111/1467-9884.00228`.
20. Gittins J and Pezeshk H. A behavioral bayes method for determining the size of a clinical trial. *Drug Information Journal* 2000; 34(2): 355–363. DOI:10.1177/009286150003400204. URL `http://dij.sagepub.com/content/34/2/355.abstract`. `http://dij.sagepub.com/content/34/2/355.full.pdf+html`.

21. Gittins JC and Pezeshk H. A decision theoretic approach to sample size determination in clinical trials. *Journal of Biopharmaceutical Statistics* 2002; 12(4): 535–551. DOI:10.1081/BIP-120016234. URL http://dx.doi.org/10.1081/BIP-120016234. PMID: 12477073, http://dx.doi.org/10.1081/BIP-120016234.
22. Halpern J, Brown BW and Hornberger J. The sample size for a clinical trial: A bayesian decision theoretic approach. *Statistics in Medicine* 2001; 20(6): 841–858. DOI:10.1002/sim.703. URL http://dx.doi.org/10.1002/sim.703.
23. Willan AR and Pinto EM. The value of information and optimal clinical trial design. *Statistics in Medicine* 2005; 24(12): 1791–1806. DOI:10.1002/sim.2069. URL http://dx.doi.org/10.1002/sim.2069.
24. Turner RM, Toby Prevost A and Thompson SG. Allowing for imprecision of the intracluster correlation coefficient in the design of cluster randomized trials. *Statistics in Medicine* 2004; 23(8): 1195–1214. DOI:10.1002/sim.1721. URL http://dx.doi.org/10.1002/sim.1721.
25. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; 20(3): 435–452. DOI:10.1002/1097-0258(20010215)20:3⟨435::AID-SIM804⟩3.0.CO;2-E. URL http://dx.doi.org/10.1002/1097-0258(20010215)20:3<435::AID-SIM804>3.0.CO;2-E.
26. Strong M, Oakley JE and Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample. *Medical Decision Making* 2014; 34(3): 311–326. DOI:10.1177/0272989x13505910. URL https://doi.org/10.1177%2F0272989x13505910.
27. Strong M, Oakley JE, Brennan A et al. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: a fast, nonparametric regression-based method. *Medical Decision Making* 2015; 35(5): 570–583. DOI:10.1177/0272989x15575286.
28. Menzies NA. An efficient estimator for the expected value of sample information. *Medical Decision Making* 2015; 36(3): 308–320. DOI:10.1177/0272989x15583495.
29. Heath A, Manolopoulou I and Baio G. Efficient monte carlo estimation of the expected value of sample information using moment matching. *Medical Decision Making* 2017; 38(2): 163–173. DOI:10.1177/0272989x17738515.
30. Jalal H and Alarid-Escudero F. A gaussian approximation approach for value of information analysis. *Medical Decision Making* 2017; 38(2): 174–188. DOI:10.1177/0272989x17715627.
31. Heath A, Manolopoulou I and Baio G. Estimating the expected value of sample information across different sample sizes using moment matching and nonlinear regression. *Medical Decision Making* 2019; 39(4): 346–358. DOI:10.1177/0272989x19837983.
32. Rowan T. *Functional Stability Analysis of Numerical Algorithms*. PhD Thesis, 1990.
33. Ypma J, Borchers HW and Eddelbuettel D. The nlopt nonlinear-optimization package, 2018. URL https://cran.r-project.org/web/packages/nloptr/nloptr.pdf.
34. Varadhan R and Gilbert P. BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software* 2009; 32(4): 1–26. URL http://www.jstatsoft.org/v32/i04/.
35. Brunier HC and Whitehead J. Sample sizes for phase II clinical trials derived from bayesian decision theory. *Statistics in Medicine* 1994; 13(23-24): 2493–2502. DOI:10.1002/sim.4780132312. URL http://dx.doi.org/10.1002/sim.4780132312.

36. Stallard N. Sample size determination for phase II clinical trials based on bayesian decision theory. *Biometrics* 1998; 54(1): 279. DOI:10.2307/2534014.

Appendix

Derivatives

Take each element of the sum as a function of coefficient a, with fixed $x_1 = x_1^{(i)}$. Then,

$$\begin{aligned}
f(a) &= g(a)u_1(a) + (1 - g(a))u_2(a) \\
f'(a) &= g'(a)u_1(a) + g(a)u_1'(a) + u_2'(a) - g'(a)u_2(a) + g(a)u_2'(a) \\
&= g'(a)(u_1(a) - u_2(a)) + g(a)(u_1'(a) - u_2'(a)) + u_2'(a) \\
&= g'(n(a))n'(a)(u_1(a) - u_2(a)) + g(a)(u_1'(n(a))n'(a) - u_2'(n(a))n'(a)) + u_2'(n(a))n'(a) \\
&= n'(a)[g'(n(a))(u_1(a) - u_2(a)) + g(a)(u_1'(n(a)) - u_2'(n(a))) + u_2'(n(a))]
\end{aligned}$$

In our case,

$$\begin{aligned}
g(a) &= 1 - \Phi\left(z - \frac{\mu}{\sqrt{2\sigma^2/n}}\right) \\
g'(n(a)) &= -\phi\left(z - \frac{\mu}{\sqrt{2\sigma^2/n}}\right)\frac{-\mu}{2\sqrt{2n\sigma^2}},
\end{aligned}$$

$$\begin{aligned}
u_1(a) &= 1 - e^{-\rho(k_d\mu + k_n(n_p+n)} \\
u_1'(n(a)) &= \rho k_n e^{\rho(k_d\mu + k_n(n_p+n)},
\end{aligned}$$

$$\begin{aligned}
u_2(a) &= 1 - e^{-\rho(k_n(n_p+n)+k_c} \\
u_2'(n(a)) &= \rho k_n e^{\rho(k_n(n_p+n)+k_c},
\end{aligned}$$

and

$$\begin{aligned}
n(a_j) &= \sum_{i=1}^{k+4} a_i\phi_i(x_1) \\
n'(a_j) &= \phi_j(x_1)
\end{aligned}$$

(drop x_1 to leave just ϕ_j?)

With respect to coefficient b, defining the decision rule for z, we have

$$\begin{aligned}
g'(a) &= z'(a)[g'(z(a))(u_1(a) - u_2(a)) + g(a)(u_1'(z(a)) - u_2'(z(a))) + u_2'(z(a))] \\
g'(z(a)) &= -\phi(z(a) - \frac{\mu}{\sqrt{2\sigma^2/n}},
\end{aligned}$$

$$u_1'(z(a)) = u_2'(z(a)) = 0,$$

and

$$z(a_j) = \sum_{i=k+5}^{2(k+4)} a_j \phi_j(x_1)$$
$$z'(a_j)) = \phi_j(x_1)$$

So the vector the partial derivative is

$$h'(a_j) = ...for j \leq k+4$$
$$= ...for j \geq k+5$$