STATISTICS IN MEDICINE, VOL. 16, 2267–2282 (1997)

SIZING CLINICAL TRIALS WITH VARIABLE ENDPOINT EVENT RATES

LEMUEL A. MOYÉ*

University of Texas School of Public Health, Houston, Texas, U.S.A.

SUMMARY

Although many researchers in cardiovascular clinical trials have disciplined themselves to execute sample size calculations in the design of their studies, these computations become difficult in the presence of control group endpoint event rate uncertainty. Recent experience in cardiovascular clinical trials suggests that, although one may know the control group event rate during the design phase of the trial, it can decrease during the trial's execution. Its resultant overestimation can lead to a power reduction with serious consequences for the trial's interpretation. Although the investigators may acknowledge the likelihood that the control group event rate will decrease during the time course of the trial, there is no formal means to adjust the design phase estimate. In this paper, I first formulate the sample size as a function of the control group event rate θ and then I place a proper probability distribution on θ, allowing for the uncertainty in this parameter's value during the course of the study. From this assumption, the sample size itself becomes a random variable, whose expectation and variance are computed. I explore the implications for sample size for various reasonable proper probability distributions on the control group event rate. © 1997 by John Wiley & Sons, Ltd.

Statist. Med., **16**, 2267–2282 (1997)
No. of Figures: 6 No. of Tables: 3 No. of References: 22

INTRODUCTION

The sample size computation is more than merely an exercise in arithmetic in prospectively designed, randomized controlled clinical trials. After much discussion and debate, agreement must be reached on the sidedness of the trial, the magnitude of the type I and type II errors, the intervention's efficacy, and the anticipated control group event rates. Only after one reaches consensus on these issues can the trial proceed. Thus, the sample size calculation and its attendant discussions are a centre of gravity in the design phase of the trial, bringing together and eventually aligning the investigators, statisticians, and trial sponsors on design issues.

There are many guidelines available for the selection of the test sidedness, the magnitude of the type I and type II errors, and the intervention, for example, Meinert.[1] Lachin[2] and Sahai[3] provide extensive aids for computing the trial sample size, a task made easier when parameter estimates are known. Sometimes, as in SHEP,[4,5] an NHLBI sponsored trial to examine the role of antihypertensive therapy in reducing the combined event rate of fatal and non-fatal strokes in the elderly U.S. population, both a pilot study and the available literature at the inception of the main

* Correspondence to: Lemuel A. Moyé, University of Texas School of Public Health, Ruell A. Stallones Building, 1200 Herman Pressler, Houston, Texas 77030, U.S.A.

CCC 0277–6715/97/202267–18$17.50

Received June 1996
Revised December 1996

trial suggested that the combined endpoint event rates were stable and easily estimated. However, in other cardiovascular clinical trials, the estimation of the control group event rate θ has been problematic. This has been particularly true in clinical trials that examine cardiovascular sequelae in post myocardial infarction populations.[6–8] In these trials, the primary endpoint event rate in the control group decreased over time due to improved care for the post myocardial infarction population, the population from which the sample was randomly selected. The unfortunate consequence of this trend was that the control group event rate at the time of the design of the trial when the sample size was computed was greater than that during the trial's execution, and led to a substantial overestimation of the trial's power.

Thus uncertainty in control group event rates is both crucial and difficult to predict. The methodology developed here posits that investigators are better served in these circumstances by not treating the control group event rate θ as a constant but as the outcome of an experiment, that is, θ is a random variable that has a probability distribution. This probability distribution should simultaneously allow more than one guess for θ, but also requires all of its probability to support the notion that the true value of θ is less than the values provided by the available guesses. Once one has specified the probability distribution, one can derive the expected sample size. This procedure translates information about the distribution of the control rate into an expected sample size.

THEORETICAL DEVELOPMENT

Consider a sample of size N subjects randomly drawn from the population of interest and randomized with equal probability to receive either control or treatment therapy. Let the endpoint of the experiment be dichotomous (such as total mortality). The goal of the trial is to compare the cumulative event rate in the control group θ_c with the cumulative event rate in the treatment group, θ_t. The null and alternative hypotheses are

$$H_0: \theta_c = \theta_t \text{ versus } H_1: \theta_c \neq 0_t$$

with a two-sided type I error rate α and power $1 - \beta$. The test statistic for the hypothesis is

$$\frac{p_c - p_t}{\sqrt{\left\{\frac{2}{N}[p_c(1 - p_c) + p_t(1 - p_t)]\right\}}}$$

referred to as the exact unconditional method by Sahai and Khuroid[3] and has a standard normal distribution. Here p_c is the observed cumulative control group rate and estimates θ_c and p_t is the observed cumulative treatment group rate and estimates θ_t. The minimum sample size to test this hypothesis in a two-sided fashion with type I error $= \alpha$ and power $= 1 - \beta$ is from Snedecor[4]

$$N = \frac{2[\theta_c(1 - \theta_c) + \theta_t(1 - \theta_t)][Z_{1-\alpha/2} - Z_\beta]^2}{[\theta_c - \theta_t]^2}. \quad (1)$$

Here Z_p is the pth percentile value of the standard normal distribution. Our goal is to consider the sample size N as a function of θ_c. Begin by letting e denote the efficacy of the trial, that is, the

relative reduction in the cumulative event rate associated with the intervention

$$e = \frac{\theta_c - \theta_t}{\theta_c}. \tag{2}$$

Let $\theta = \theta_c$ for notational convenience. Then we can write N as

$$N = \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(2 - e)\left[\frac{1}{\theta}\right] - \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(1 + (1 - e)^2). \tag{3}$$

By considering $Z_{1-\alpha/2}$, Z_β, and e as constants, N is a linear function of the reciprocal of the cumulative control group endpoint event rate. For example, if the trial is to have 90 per cent power to detect an efficacy of 20 per cent with a two-sided type I error of 5 per cent, we may write N as

$$N = 944{\cdot}78\frac{1}{\theta_c} + 860{\cdot}80.$$

The estimation of $\theta_c = \theta$ is the focus here. If we treat θ not as a constant but as a random variable, with a probability distribution $f_\theta(\theta)$, we may find the expected value of θ^{-1}, and compute $\mu_N = E[N]$

$$E[N] = \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(2 - e)E\left[\frac{1}{\theta}\right] - \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(1 + (1 - e)^2) \tag{4}$$

and the variance of the sample size is $\sigma_N^2 = \text{var}[N]$

$$\text{var}[N] = \frac{4[Z_{1-\alpha/2} - Z_\beta]^4}{e^4}(2 - e)^2 \text{ var}\left[\frac{1}{\theta}\right]. \tag{5}$$

Form of the density for θ_c

The choice of the probability density function $f_\theta(\theta)$ must be proper (that is, integrate to one over the real line), and directed to the particular problem confronting the investigators, that is, overestimation of the cumulative placebo event rate θ. Suppose current investigator opinion suggests that the value of θ is a. Then consider the probability density function

$$f_\theta(\theta) = (k + 1)\frac{\theta^k}{a^{k+1}} I_{0 \leqslant \theta \leqslant a}$$

where I_x is the indicator function, equal to one when condition x is true, 0 otherwise. This density function $f_\theta(\theta)$ places probability over the $[0, a]$ interval where $0 \leqslant a \leqslant 1$. The constant a is an available guess for θ. For this probability density function, we consider k the uncertainty parameter. The larger the value of k, the more closely the probability for θ concentrates near the value of a. Thus, in circumstances where the value of θ is known with some assurance, large values of k are the most appropriate in describing the density of θ_c. Smaller values of k are appropriate in situations when the investigators suspect a larger drift downward for θ.

To take advantage of multiple candidate values for the maximum value of θ, we can incorporate the notion of several 'guesses' $a_1, a_2, a_3, \ldots, a_m$ being available for θ such that $0 \leqslant a_i \leqslant 1$, $i = 1, 2, 3, \ldots, m$. In this circumstance, each value of a_i is 'believed' with probability $p_i, i = 1, 2, 3, \ldots, m$ such that

$$\sum_{i=1}^{m} p_i = 1.$$

In this general case we may write the probability density function for θ_c as

$$f_\theta(\theta) = \sum_{i=1}^{m} p_i(k_i + 1) \frac{\theta^{k_i}}{a_1^{k_i+1}} I_{0 \leqslant \theta \leqslant a_1} \tag{6}$$

Using this density function we can compute the expected value of the sample size from (4) as

$$E[N] = \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(2 - e) \sum_{i=1}^{m} \left[\frac{p_i(k_i + 1)}{k_i a_i}\right] - \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(1 + (1 - e)^2) \tag{7}$$

and the var$[N]$ as

$$\text{var}[N] = \frac{4[Z_{1-\alpha/2} - Z_\beta]^4}{e^4}(2 - e)^2 \left[\sum_{i=1}^{m} \frac{p_i(k_i + 1)}{(k_i - 1)a_i^2} - \left(\sum_{i=1}^{m} \frac{p_i(k_i + 1)}{k_i a_i}\right)^2\right] \tag{8}$$

RESULTS

The purpose of this methodology is to develop a useful family of proper probability density functions for the population placebo cumulative event rate and to demonstrate the effect of that density choice on the sample size estimates. However, before any group of investigators can make a rational choice for which density to use, these investigators must be able to match the density parameter k to the degree of uncertainty they have for θ. Figure 1 shows the relationship between the value of k and the density function for θ. In this example, there is only one guess for θ, and that is the value $\theta = 0{\cdot}25$. In each case, all of the probability is located to the left of the guess, conforming to the assumption that θ is smaller at the trial's conclusion. For $k = 1$, the probability density function is a straight line, distributing probability for θ across a wide range of values of $\theta \leqslant 0{\cdot}25$. For $k = 5$, the bulk of the probability is more closely located to θ. For the highest value of k, the probability is more tightly clustered near θ. The larger the value of k, the more closely the probability clusters near the guess. Investigators who are confident of their guess with little chance of overestimating θ would choose a large value of k, while investigators who suspect that θ could decrease substantially during the course of the trial would choose a smaller value of k.

More than one source of information about the value of θ can lead to more than one guess. For example, there may be two different literature sources that suggest different values for θ, guess a_1 and guess a_2. The availability of more than one guess is considered in this model by having a value of $m > 1$. Figure 2 demonstrates the different patterns of $f_\theta(\theta)$ for $m = 2$. In part (*a*), two

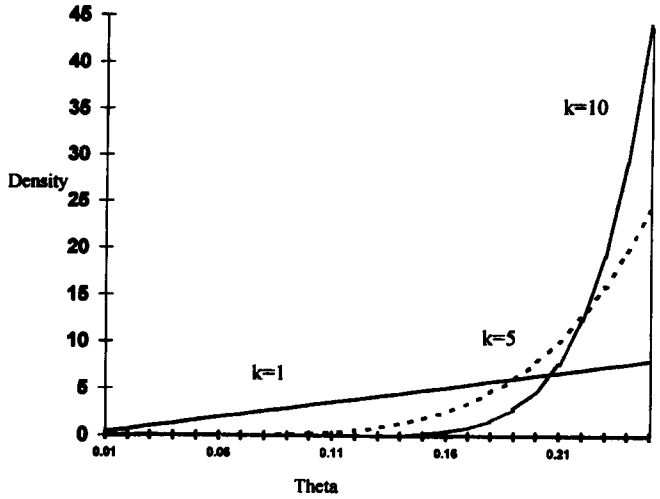

Figure 1. Density for theta as a function of k

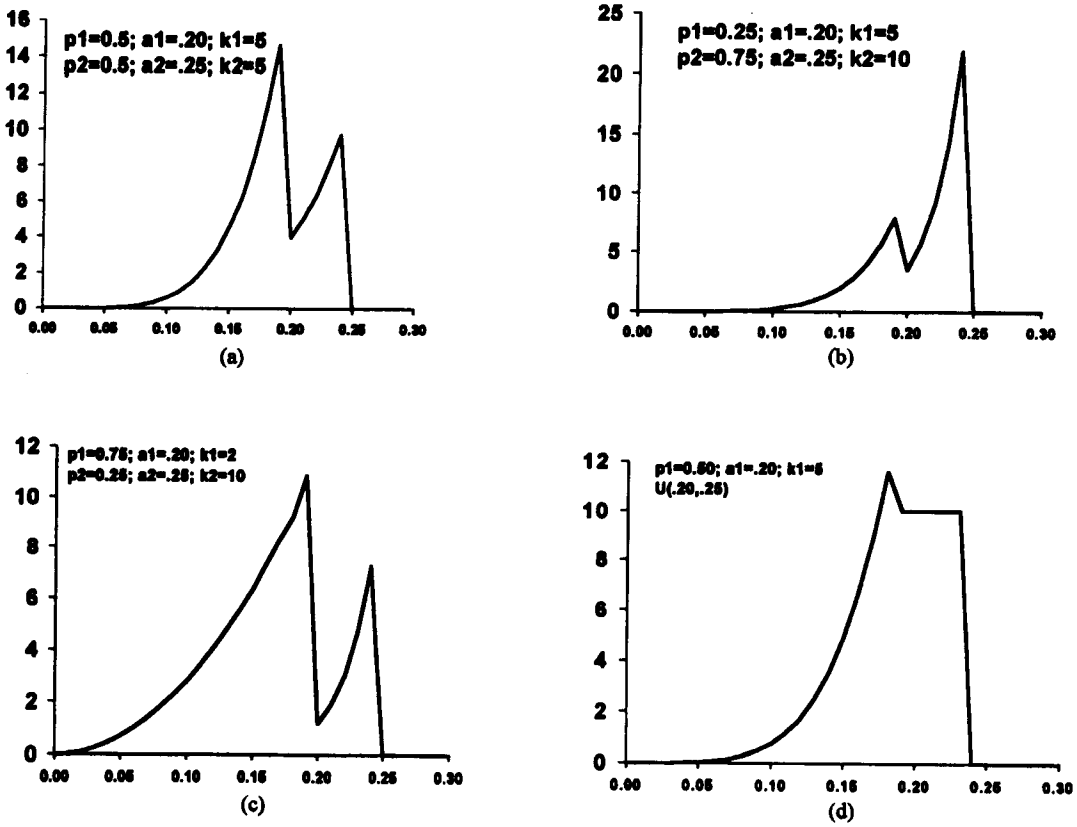

Figure 2. Examples of mixture densities for theta

Table I. Expected trial size and its standard error as a function of $k(a = 0{\cdot}20$, efficacy = 20 per cent, two-sided $\alpha = 0{\cdot}05$; power = 90 per cent)

	k										
	2	3	4	5	6	7	8	9	10	11	12
μ_N	6225	5438	5044	4808	4650	4538	4454	4388	4336	4293	4257
σ_N	2273	1237	847	643	518	432	372	326	290	261	238

guesses for θ are available, $\theta = 0{\cdot}20$ and $\theta = 0{\cdot}25$. The investigators consider each as equally likely, therefore $p_1 = p_2 = 0{\cdot}50$. However, regardless of the guess, chosen, θ is likely to be a value less than that guess during the course of the trial; we measure the degree to which it is overestimated by the parameters k_1 and k_2, where k_1 measures the degree of potential underestimation of guess a_1 and k_2 measures the degree of potential overestimation based on guess a_2. In this example, we believe each guess is an overestimate to the same degree, so we let $k_1 = k_2 = 5$. The resulting density places local maximum values of θ at the two guesses, but allows for the fact that each may represent an overestimate. The global maximum of the density occurs near $\theta = 0{\cdot}20$ since overestimating θ for the guess of 0·25 increases the probability that θ will be near 0·20. Figure 2(*b*) and (*c*) provide additional probability density functions for different choices of p_1, p_2, k_1 and k_2. It is important to note that this family of densities distinguishes the guesses in two ways. The first is the likelihood that the guess a_i is the best available. This is determined by the parameter p_i. The second is the potential for overestimation, as determined by k_i.

Figure 2(*d*) is a density of a different but related form. It presumes that any value of θ between the two guesses a_1 and a_2 is equally likely but that they are each overestimates of its true value. Thus θ has a power distribution on $[0, a_1]$ with probability p_1 and with probability p_2 follows a uniform distribution between a_1 and a_2.

Using this family of densities, we can apply formula (4) to compute the expected value and formula (5) to compute the variance of the trial size. Table I showing μ_N and σ_N based on $m = 1$ as a function of k reveals that the sample size decreases as k increases. Since large values of k reflect that there is not much overestimation of θ, much of the probability is focused around $\theta = 0{\cdot}20$, and the sample size is smaller. The smaller the value of k, the greater becomes the problem of overestimation, increasing the likelihood of smaller event rates, increasing the expected value of the sample size. This same relationship occurs between the standard deviation of the sample size σ_N and k. The more certain the investigators are of θ, the smaller is the problem of overestimation, the less variance there is with θ, the smaller σ_N. For smaller values of k, circumstances where there may be an extreme overestimation of θ, σ_N is large. Figure 3 depicts this relationship between the expected sample size and k. Here μ_N is plotted as well as $\mu_N \pm 1{\cdot}96\sigma_N$. What is particularly striking is the confidence interval range for the sample size. The confidence interval is wide for small values of k, that is, for circumstance where the likelihood of an overestimation of θ is large. The more confident the investigators are with their estimate of θ, the less likely they are to be plagued with an overestimate, the smaller their confidence interval for μ_N.

For each of the parts in Figure 2, I provide the expected value of the sample size μ_N and its standard deviation σ_N (Table II).

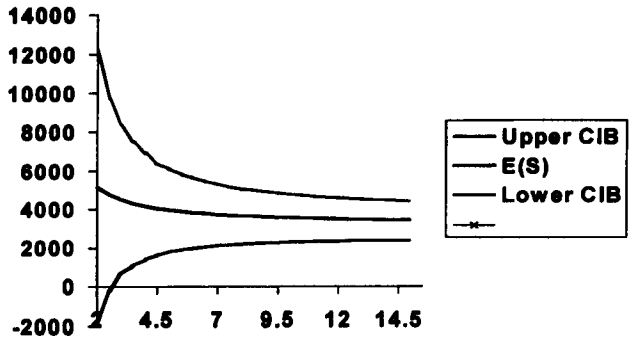

Figure 3. Sample size and 95 per cent CI as a function of k

Table II. Expected trial size and its standard error for four different mixture scenarios (efficacy = 20 per cent, two-sided $\theta = 0{\cdot}05$; power = 90 per cent)

	Part (*a*)	Part (*b*)	Part (*c*)	Part (*d*)
μ_N	4241	3674	5493	4082
σ_N	662	525	2094	1111

DISCUSSION

As clinical trial experience has matured, a number of strategies have arisen to compute the sample size and power in experimental circumstances. Relevant summaries are provided by Lachin[2] and more recently by Sahai[3] and Donner.[10] Sample size computations have been extended to more than two groups by Davy,[11] and strategies for multiple endpoints have been explored by Follman.[12] George and Desu[13] have taken a different perspective, that is , suggesting that the investigators might control the length of the clinical trial, and Hornberger[14] provides help for the investigator who wishes to design a cost effective clinical trial by allowing cost benefit considerations to influence the sample size. However, sample size estimation in post infarction clinical trials has been problematic because of the anticipated change in the cumulative placebo rate over time. The observation of declining value of p_c, (that is, declining event rates) in post myocardial infarction cardiovascular clinical trials settings relates to the major advances in the treatment of coronary artery disease made since the 1970s continuing to and through the current time. These rapidly discovered and rapidly implemented treatment modalities reduce the event rate of the morbidity/mortality measure in experiments currently underway. For example, since the inception of SAVE in January 1987, the uses of β-blockade, aspirin, percutaneous, transluminal coronary angioplasty (PTCA), thrombolytic therapy, ACE inhibitor therapy has each become increasingly common. Thus the institution of a clinical trial to evaluate the use of a new intervention in this dynamic clinical setting must reduce the incidence of an endpoint (for example, the occurrence of an acute myocardial infarction) which is already under reduction with new recently proven therapies. This will continue as an issue for the foreseeable future, as new

cholesterol reducing therapies are become available for physicians to decrease the myocardial infarction rate even further. In these circumstances, investigators assume constant control group event rates at their own peril. The unfortunate consequence of this trend has been that the control group event rate during the design of the trial (at the time of the sample size computations) is larger than that during the trial's execution. Studies that ignore this phenomenon experience lower than expected primary endpoint event rates, and may be underpowered at the trial's end.

However, the impact can be difficult to predict since the choice of the correct value of θ when we know θ to be changing is inherently subjective and somewhat of an art. Unfortunately, investigators are not assured but are discomforted by this state of affairs. When actually confronted with estimation of θ in the face of uncertainty, clinical investigators desire a formal way to remove the uncertainty or to express it. Unfortunately, it is not easily removed. Regression analysis, which would attempt to model the change in θ over time, is of little help since there is no consensus on the form of the model. The decline in θ over time could be linear, quadratic, or higher order. There may be a lag between the introduction of the intervention that will produce the reduced rate and its impact in the population. The occurrence of side-effects, costs, and patient/physician resistance to increasingly complicated treatment programmes can blunt the change. Specification of a regression model on which an assumption for θ depends presumes underlying confidence in the model itself, a confidence that clinical trial workers often do not have. Investigators therefore turn to their own interpretation of the results of previous trials with guidance by their intuition. CARE[7] (Cholesterol and Recurrent Events) was a prospective, double-blind clinical trial that recently completed a five year follow-up on 4159 patients, that evaluated the role of HMG CoA reductase inhibitor therapy in reducing post infarction coronary artery disease. In this trial, the endpoint was fatal and non-fatal coronary artery disease. The primary event rate in this trial was smaller than anticipated due to the declining myocardial infarction rates during the course of the trial. Although the trial was able to reject the null hypothesis due to an observed intervention efficacy that was greater than expected, the smaller than expected event rate that occurred during the course of the trial was particularly vexsome.

To avoid power overestimation, some investigators at the trial's beginning will attempt a guess at a reduction factor to modify the estimate of the control event rate at the trial's end. This was the case in SAVE[8,9] (Survival and Ventricular Enlargement), a prospective, double-blind, randomized placebo controlled clinical trial that examined the impact of captopril therapy on post myocardial infarction survival. The endpoint of this trial was total mortality. Believing that the control group event rate at the end of the trial would be two-thirds of its original value at the trial's inception, the investigators reduced their estimate of the rate by one-third in the sample size estimate. The difficulty with this approach is that there is very little reliable information to guide the guess. This intuitive procedure was successful because it was both conservative and did not lead to a large sample size that made the experiment prohibitively expensive. However the 'two-thirds' rule in SAVE led to an assumption of a cumulative event rate of 20 per cent, and resulted in a sample size that was somewhat larger than optimal for the greater event rate actually observed at the trial's end.

Clinical trialists have designed pilot trials to justify sample size assumptions of cumulative control group event rates. In SHEP[5,6], an NHLBI sponsored clinical trial to evaluate the effect of the treatment of isolated systolic hypertension in the elderly to prevent fatal and non-fatal strokes, the cumulative placebo event rate was estimated from a pilot study. This strategy served

Table III. Relationship between degree of overestimation and the control event rate p_c ($p_1 = 1{\cdot}00$, $a_1 = 0{\cdot}25$, two-sided $\alpha = 0{\cdot}05$, power = 90 per cent)

	k										
	1	2	3	4	5	6	7	8	9	10	11
θ	0·125	0·167	0·188	0·200	0·208	0·214	0·219	0·222	0·225	0·227	0·229

the trial well, but, as suggested by Browne,[15] use of pilot study data to estimate event rates does not provide complete protection against overestimation of population parameters.

The methodology developed here posits that the control group event rate θ is not constant but is itself the outcome of a sampling experiment, that is, θ is a random variable that has a probability distribution. The probability distribution developed here is sufficiently robust to incorporate simultaneously discrete information (that is, multiple guesses) for θ and reflect the intuition that the event rate chosen likely overestimates the control group event rates observed during the trial. Once we have specified the probability distribution, we have identified the expected sample size. This procedure translates information about the distribution of the control rate into an expected sample size that explicitly takes into account that information. Such a result can provide new guidance in determining clinical trial sample sizes.

However, to utilize this, the investigator must have a procedure to estimate the parameter k of the power density given in formula (6). To generate further intuition for the value of k, Table III provides the results of an inversion. Comparing formula (3) in which θ is a constant to formula (4) where θ is random, and, by letting $m = 1$, compute

$$\theta = \frac{p_1(k_1 + 1)}{k_1 a_1}.$$

This provides the value for θ (assuming θ is a constant) that yields the sample size that stems from the assumption that θ is random with a power density provided by guess a_1 with uncertainty parameter k_1 (Table III). For example, with one guess for θ of 0·25, assumption of a value of $k = 10$ results in the same expected sample size that we obtain with the assumption with no uncertainty that $\theta = 0{\cdot}227$. Assuming a value of $k = 5$ (much more uncertainty) yields the equivalent sample size of 0.208. This relationship allows the investigators to calibrate their degree of uncertainty with possible values of θ. Of course, the difficulty is that the investigators do not know to choose 0·185, they only know an approximate value of θ represented by the guesses, and that these values will likely be shown to be overestimates of θ's true value.

The result of this effort develops the notion that the sample size N is not a constant, based on a constant value of θ, but is itself is a random variable, a function of the random variable θ. The sample size thus has a probability density function with a mean μ_N and a variance σ_N. Although investigators are unaccustomed to incorporation of a sample size variance in their considerations, the concept is a natural consequence of the assumption that θ is not fixed but is a random variable and the sample size is a function of that random variable. By randomly selecting the study sample from the population at large, the investigators are randomly selecting θ. Since θ varies among randomized samples, the required minimum sample size varies as well. In fact, we can obtain the

probability density function of the sample size N from equations (4) and (6) as

$$f_N(n) = \sum_{i=1}^{m} p_i(k_i + 1)\left(\frac{c_1}{a_i}\right)^{k_i+1}\left(\frac{1}{n + c_2}\right)^{k_i+2} I_{(c_1/a_1)-c_2<n<\infty}$$

where

$$c_1 = \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(2 - e) \text{ and } c_2 = \frac{2[Z_{1-\alpha/2} - Z_\beta]^2}{e^2}(1 + (1 - e)^2).$$

Of course, the investigators do not have the luxury of choosing many samples from their population of interest. They choose one sample, based on their assumptions concerning θ. By doing this, the sample size becomes fixed. Given that the sample size is fixed, we need to compute the power of the trial. In general, we consider θ to be a constant. However, in this setting, θ is not a constant but a random variable, varying according to a probability distribution. Thus, we cannot compute the power, but can derive the expected value of the power. Once we have chosen the sample size, the paradigm shifts from one of exploring the relationship between sample size and θ to one of the relationship of power and θ. We can readily explore this relationship as well. Power is the probability that test statistic falls in the critical region under the alternative hypothesis when efficacy e is non-zero. For fixed θ, this is

$$P\left[N(0,1) \geqslant Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{\left\{\frac{2}{N}[(2 - e) - \theta(1 - (1 - e)^2)]\right\}}}\right].$$

We need to find the expected value of this expression with respect to θ. When $f_\Theta(\theta)$ is as in (6), a computation (see Appendix) shows the expected power $E[P]$ is

$$E[P] = \sum_{i=1}^{m} p_i\left[1 - \Phi_z(h(a)) - \left(\frac{b_1}{a_i}\right)^2 \int_{h(a)}^{h(0)}\left[\frac{Z_{1-\alpha/2} - z}{e^2 + b_2(Z_{1-\alpha/2} - z)^2}\right]^{k_i+1} \varphi_z(z)\,dz\right] \quad (9)$$

where $b_1 = 2((2 - e))/N$, $b_2 = 2(1 - (1 - e)^2)/N$, $\phi_z(z)$ is the density of a standard normal random variable, $\Phi_z(z)$ is the cumulative distribution function of a standard normal random variable and

$$h(\theta) = Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{(b_1 - b_2\theta)}}.$$

This computation provides a way to compute the effect of k on the expected power for a fixed sample size. Consider a clinical trial scenario, when for a two-sided type I error of 0·05, power of 90 per cent, efficacy of 20 per cent, we compute a sample size of 3863 using (1) and $\theta = 0{\cdot}20$. If the assumption of constant θ is wrong, and in fact the value of θ is random with a density function $f_\Theta(\theta)$, as in equation (6), then we can use relationship (9) to explore the impact of the parameter k on power. For example, for $k = 1$, which reflects substantial decrease in θ over the course of the trial, the expected value of the power is 24 per cent. For $k = 5$, the $E[P] = 0{\cdot}61$. Figure 4 depicts the relationship between the posterior power of the trial and k.

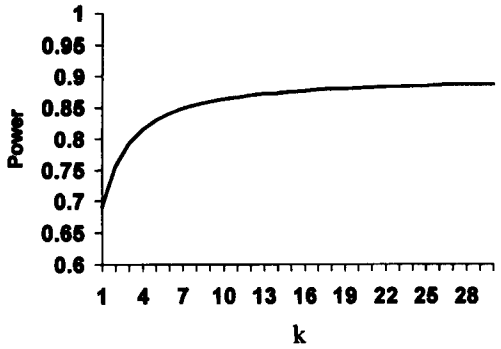

Figure 4. Expected power as a function of k ($N = 3863$, type I error is two-sided 0·05)

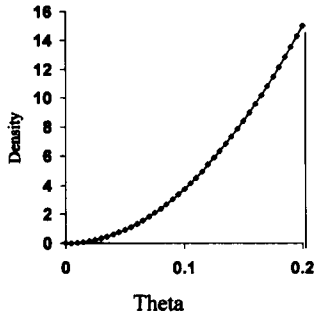

Figure 5. Density function for theta from CARE

The example from CARE is illustrative. In that trial, the investigators, using the best information available from the literature, anticipated a five year cumulative primary event rate (fatal and non-fatal coronary artery disease) of 0·20, which resulted in a sample size of 4000. At the conclusion of recruitment, 4159 patients had been randomized to CARE. At the conclusion of the five year follow-up, a cumulative placebo event rate was 0·132. Thus, there was a substantial overestimation of θ. It is of interest to examine the density function that most closely corresponds to the observed value of θ in CARE, that is, to find the value of k that results in an expected value of $\theta = 0{\cdot}132$. Using equation (3) the value of θ at the conclusion of CARE is equivalent to a k of 1·2. Figure 5 denotes the density of θ from CARE.

There are alternative algorithms one can use to compute the sample size of a clinical trial when information is unavailable for the important parameters. Simulation is a useful tool in this regard. For example, Feng[16] provides guidelines for sample sizes when communities are the unit of randomization, but there is insufficient information available for the intraclass correlation, a necessary parameter for sample size. One could simulate the sample size for a wide range of θ. Alternatively, one could conduct a sensitivity analysis using a range of values of θ and then select a relatively conservative solution. These procedures have the advantage of examining a wide

range of alternatives for θ. However, they do not incorporate the wide range of values into a single summary measure. The key is to choose an algorithm that guides the selection. An alternative procedure is to choose a sample size that provides good power for a range of values, similar to a minimax rule. For example, choose n so as to be 90 per cent sure of having, say, 85 per cent power for a reasonable range of θ, that is

$$\min_{\theta \in [b,1]} P_\theta(\text{reject } H_0 \text{ when } H_1 \text{ is true}) \geqslant 0{\cdot}85$$

where

$$P(\theta \in [b,1]) = 0{\cdot}90.$$

In general, the choice of the probability density for θ, that is, the value of k should be directly reflective of the investigators' perceived degree of uncertainty. To facilitate this process, we can collapse the degree of uncertainty to one of three levels:

(i) *minimal uncertainty* corresponds to $k = 15$;
(ii) *moderate uncertainty* corresponds to $k = 5$;
(iii) *extreme uncertainty* corresponds to $k = 1$.

These choices ease the decision process yet capture the essence of the methodology. In CARE, and perhaps in the design of contemporary cardiovascular clinical trials, extreme uncertainty prevails.

Although the purpose here is to respond specifically to the issue of overestimation of θ, we can generalize to a case where the probability distribution provides non-zero probability for values of θ greater than the current best guess. However, an underestimation of θ, all other parameter estimates being accurate, leads to more power than originally designed.

Since we place a probability distribution on θ and we take an expectation based on it, the temptation is great to call this a Bayesian approach to sample size estimation. However, there are important criteria that a Bayes procedure must meet and that are lacking here. A truly Bayesian procedure has not just a prior distribution on θ, but also computes a formal Bayes' analysis, that is, one identifies a posterior distribution, and through the incorporation of a loss function, computes a statistic which is the Bayes' procedure. Spiegalhalter *et al.*[17] have reviewed the increasing use of Bayesian procedures in clinical trials. Arjas and Liu[18] have reviewed the use of Bayesian procedures in clinical trial analysis in the special case of missing covariates. Whitehead[19] has provided a Bayesian procedure in dose response experiments. Thall,[20] Berry[21] and Rosner[22] have provided useful contributions to the use of Bayesian procedures in interim monitoring of these experiments.

APPENDIX

We must find the expected value of the power of a clinical trial when that power function is in terms of θ, the cumulative control group event rate which is itself a random variable that follows the mixture probability density function

$$f_\theta(\theta) = \sum_{i=1}^{m} p_i(k_i + 1)\frac{\theta^{k_i}}{a_1^{k_i+1}} I_{0 \leqslant \theta \leqslant a_1}$$

where

$$\sum_{i=1}^{m} p_i = 1.$$

Power is the probability that the test statistic of the primary hypothesis of interest falls in the critical region under the alternative hypothesis. Let e denote the efficacy of the trial under the alternative hypothesis, α the two-sided type I error and $N/2$ the subjects randomly allocated to each of two groups. Then

$$\text{Power} = P\left[\text{N}(0, 1) \geqslant Z_{1-\alpha/2} - \frac{\theta - (1-e)\theta}{\sqrt{\left\{\frac{N}{2}(\theta(1-\theta) + (1-e)\theta[1-(1-e)\theta])\right\}}}\right]$$

which we can rewrite as

$$\text{Power} = P\left[\text{N}(0, 1) \geqslant Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{\left\{\frac{2}{N}[(2-e) - \theta(1+(1-e)^2)]\right\}}}\right].$$

The expected value of the power $E_{\Theta}[P] = E[P]$ follows as

$$E[P] = E_{\Theta}\left[1 - \Phi_z\left[Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{\left\{\frac{2}{N}[(2-e) - \theta(1+(1-e)^2)]\right\}}}\right]\right]$$

where $\Phi_z(z)$ is the cumulative distribution function for a standard normal random variable. Given $f_{\Theta}(\theta)$ is a mixture of density functions, each of which represents the probability density function of θ with probability p_i, we may write the expectation as

$$E[P] = \sum_{i=1}^{m} p_i E_{\Theta_i}\left[1 - \Phi_z\left[Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{\left\{\frac{2}{N}[(2-e) - \theta(1+(1-e)^2)]\right\}}}\right]\right]$$

and we must identify the expectation for each component density in $f_{\Theta}(\theta)$. Begin by writing

$$b_1 = \frac{2}{N}(2-e): \quad b_2 = \frac{2}{N}(1+(1-e)^2)$$

and the component expectation we must now find is

$$E_{\Theta_i}\left[1 - \Phi_z\left[Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{(b_1 - b_2\theta)}}\right]\right]$$

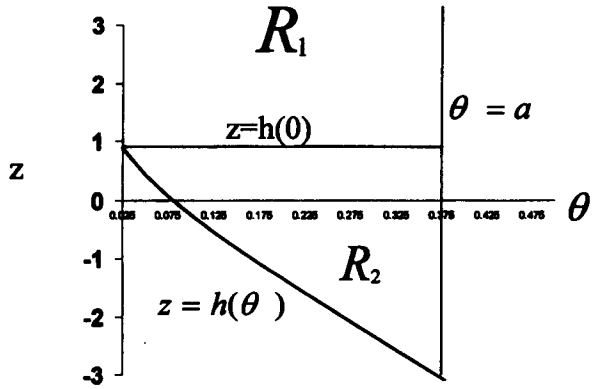

Figure 6. Region of integration

where

$$f_{\Theta}(\theta) = \frac{(k+1)\theta^k}{a^{k+1}} I_{0 \leqslant \theta \leqslant a}.$$

Figure 6 reveals that we need to evaluate this double integral over the union of the regions R_1 and R_2 Proceed by integrating first with respect to θ, then with respect to z over each of these regions.

$$\text{Power} = \iint_{R_1} f_\theta(\theta)\varphi_z(z)\,d\theta\,dz + \iint_{R_2} f_\theta(\theta)\varphi_z(z)\,d\theta\,dz.$$

The evaluation of the double integral is quickly seen to be $\alpha/2$ since the integral with respect to θ over this region is 1 and the double integral reduces to the probability that a standard normal random variable is greater than the $1 - \alpha/2$ percentile value.

The double integral over the region R_2 is the integral over the region bounded by the curve $z = h(\theta)$, $z = h(0)$ *and* $\theta = a$ where

$$h(\theta) = Z_{1-\alpha/2} - \frac{e\sqrt{\theta}}{\sqrt{(b_1 - b_2\theta)}}.$$

By writing $\theta = h^{-1}(z)$ where

$$h^{-1}(z) = \frac{b_1[Z_{1-\alpha/2} - z]^2}{e^2 + b_2[Z_{1-\alpha/2} - z]^2}$$

the integral over R_2 becomes

$$\iint_{R_1} f_\theta(\theta)\varphi_z(z)\,d\theta\,dz = \int_{h(a)}^{h(0)} \left[\int_{h^{-1}(z)}^{a} f_\theta(\theta)\,d\theta \right]_z \varphi(z)\,dz.$$

By evaluating the inner integral first, this iterated integral reduces to

$$\int_{h(a)}^{h(0)} \left[1 - \frac{[h^{-1}(z)]^{k+1}}{a^{k+1}}\right] \varphi_z(z)\,dz.$$

With simplification and adding to the value of the double integral over R_1 we find

$$\text{Power} = 1 - \Phi_z(h(a)) - \left[\frac{b_1}{a}\right]^{k+1} \int_{h(a)}^{h(0)} h^{-1}(z)^{k+1}\,\varphi_z(z)\,dz$$

and, for the mixture of probabilities, the power is the weighted sum of the powers from the densities in the mixtures.

ACKNOWLEDGEMENTS

The author wishes to thank the two referees for their constructive suggestions, particularly for the use of a minimax-like application presented in the discussion. Special thanks also go to the editor for helpful suggestions for the manuscript's style.

REFERENCES

1. Meinert, C. L. *Clinical Trials Design, Conduct and Analysis*, Oxford University Press, 1986.
2. Lachin, J. M. 'Introduction to sample size determinations and power analyses for clinical trials', *Controlled Clinical Trials*, **2**, 93–114 (1981).
3. Sahai, H. and Khuroid, A. 'Formulae and tables for determination of sample size and power in clinical trials for testing differences in proportions for the two sample design', *Statistics in Medicine*, **15**, 1–21 (1996).
4. Snedecor, G. W. and Cochran, W. G. *Statistical Methods*, 4th ed, Iowa University Press, 1980.
5. The SHEP Cooperative Research Group. 'Rationale and design of a randomized clinical trial on prevention of stroke in isolated systolic hypertension', *Journal of Clinical Epidemiology*, Vol **41**, (12), 1197–1208 (1988).
6. The SHEP Cooperative Research Group. 'Prevention of stroke by antihypertensive drug therapy in older persons with isolated systolic hypertension: final results of the systolic hypertension in the elderly program (SHEP)', *Journal of the American Medical Association*, **265**, (24), 3255–3264 (1991).
7. Sacks, F. M., Pfeffer, M. A., Moyé, L. A., Brown, L. E., Hamm, P., Cole, T. G., Hawkins, C. M., Braunwald, E. 'The cholesterol and recurrent events trial (*CARE Clinical Trial*): rationale, design and baseline characteristics of secondary prevention trial of lowering serum cholesterol and myocardial infarction', *American Journal of Cardiology*, **68**, (15), 1436–1446 (1991).
8. Moyé, L. A. for the SAVE Cooperative Group. 'Rationale and design of a trial to assess patient survival and ventricular enlargement after myocardial infarction', *Journal of American Cardiology*, **68**, 710–790 (1991).
9. Pfeffer, M. A., Braunwald, E., Moyé, L. A., Basta, L., Brown, E. J., Cuddy, T. E., Davis, B. R., Geltman, E. M., Goldman, S., Flkaker, G. T. C., Klein, M., Lamas, G. A., Packer, M., Rouleau, J., Rouleau, J. L., Rutherford, J., Wertheimer, J. K., Hawkins, C. M. 'Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction–results of the Survival and Ventricular Enlargement Trial', *New England Journal of Medicine*, **327**, (10), 669–677 (1992).
10 Donner, A. 'Approach to sample size estimation in the design of clinical trials – a review', *Statistics in Medicine*, **3**, 199–214 (1984).
11. Davy, S. J. and Graham, O. T. 'Sample size estimation for comparing two or more treatment groups in clinical trials', *Statistics in Medicine*, **10**, 3–43 (1991).
12. Follmann, D. 'Multivariate tests for multiple endpoints in clinical trials', *Statistics in Medicine*, **14**, 1163–1175 (1995).
13. George, S. L. and Desu, M. M. 'Planning the size and duration of a clinical trial studying the time to some critical event', *Journal of Chronic Disease*, **27**, 15–24 (1974).

14. Hornberger, J. C., Brown, B. W. and Halpern, J. 'Designing a cost effective clinical trial', *Statistics in Medicine*, **14**, 2249–2259 (1995).
15. Browne, R. H. 'On the use of a pilot study for sample size determinations', *Statistics in Medicine*, **14**, 1933–1940 (1995).
16. Feng, Z. and Grizzle, J. E. 'Correlated binomial variates: properties of estimation of intraclass correlation and its effect on sample size calculation, *Statistics in Medicine*, **11**, (12), 1607–1614 (1992).
17. Spiegahalter, D. J., Freedman, L. S. and Parmar, M. K. B. 'Applying Bayesian ideas in drug development and clinical trials', *Statistics in Medicine*, **12**, 1501–1511 (1993).
18. Arjas, E. and Liu, L. 'Nonparametric Bayesian approach to hazard regression: a case study with a large number of missing covariate values', *Statistics in Medicine*, **15**, 1757–1770 (1996).
19. Whitehead, J. and Brunier, H. 'Bayesian discussion procedure for dose determining experiments', *Statistics in Medicine*, **14**, 885–893 (1995).
20. Thall, P. F., Simon, R. M. and Estey, E. H. 'Bayesian sequential monitoring designs for a single arm clinical trials with multiple outcomes', *Statistics in Medicine*, **14**, 357–379 (1995).
21. Berry, D. A. 'Interim analysis in clinical trials: classical vs Bayesian approach', *Statistics in Medicine*, **4**, 521–526 (1995).
22. Rosner, G. L. and Berry, D. A. 'A Bayesian group sequential design for a multiple arm randomized clinical trial', *Statistics in Medicine*, **14**, 381–394 (1995).