Balancing the Number and Size of Sites: An Economic Approach to the Optimal Design of Cluster Samples

Connelly, Luke B.

The University of Queensland, Centre of National Research on Disability and Rehabilitation Medicine

2003

Balancing the Number and Size of Sites: An Economic Approach to the Optimal Design of Cluster Samples

(Short running title: Optimal Design of Cluster Samples)

Luke B. Connelly, BA(Econ), MEconSt, PhD
Associate Professor in Health Economics and Assistant Director
Centre of National Research on Disability and Rehabilitation Medicine (CONROD)
Herston Road
HERSTON Q 4006
AUSTRALIA

tel.: +61 7 3346 4838 (direct)
+61 7 3365 6560
fax.: +61 7 3346 4603
email: l.connelly@uq.edu.au

Reprints will not be available from the author.

ABSTRACT

The design of randomised controlled trials (RCTs) entails decisions that have economic, as well as statistical implications. In particular, the choice of an individual or cluster randomisation design may affect the cost of achieving the desired level of power, other things equal. Furthermore, if cluster randomisation is chosen, the researcher must decide how to balance the number of clusters, or "sites", and the size of each site. This paper investigates these interrelated statistical and economic issues. Its principal purpose is to elucidate the statistical and economic trade-offs to assist researchers to employ RCT designs that have desired economic, as well as statistical, properties.

Keywords: cluster sample, optimal design, economic analysis.

INTRODUCTION

In the design of randomised controlled trials (RCTs), a fundamentally important technical consideration is the statistical power (1-β) that will be produced, given the specified critical difference (d_c) of interest, and the chosen level of significance (α). These statistical issues, however, are not usually considered in isolation. Rather, as Selvin points out "...a choice of an acceptable...[level of] statistical power is usually made on nonstatistical grounds" [1, p.102). While clinical or other technical considerations are most likely to govern the selection of the effect size, economic considerations are likely to be central to the choice of power and significance levels.

This paper outlines a process whereby the resource constraints to which researchers are usually subject, may be considered during the sample design process. In particular, the paper illustrates how researchers can minimise the cost of producing a given level of statistical power, other things equal (i.e. given the specified null and alternative hypotheses, d_c and α levels); or, equivalently, how power can be maximised given a specified budget.

This optimisation problem has apparently not been explored, in any depth, in the existing literature on cluster randomisation. That is not to say that the matter of cost-effective sample design has been ignored in that literature; monographs and journal articles on cluster randomisation routinely refer to the matter of cost-effective sample design. A method for systematically determining the optimal combination of sites and site size has not, however, previously been elucidated. The focus of the current paper is on the statistical and economic trade-offs between the number and the size of sites chosen that arise in cluster designs. Its objective is to provide an appropriate framework for the construction of economically-efficient samples that have the desired statistical properties.

RCT DESIGN AS AN ECONOMIC PROBLEM

The problem of sample design may be conceived as an economic, as well as a statistical problem. The problem for the RCT designer is typically to develop a study design that not only satisfies the statistical requirements of the investigators, but also conforms to some budget constraint. As such, the task of sample design may be viewed as a classic economic production problem.

Economists consider the problem of production as one that is of a "dual" nature. That is to say that production problems can be viewed from two, analytically equivalent, perspectives, *viz.* as (i) that of minimising the cost of producing a specified level of output; or (ii) that of maximising the output produced with a specified budget. (For both (i) and (ii) it is understood that the quality of the output under consideration is to be constant when cost minima and output maxima are considered. For a more detailed statement of duality theory, see, e.g. [2].) When a production process satisfies the condition described by (i) and (ii), it said to be "technically efficient".

For RCT design, the economic problem may be cast in the following "dual": (i) maximise the statistical power of the test, given a specific resource constraint (and the required level of precision); or (ii) minimise the cost of producing a specific level of statistical power (and the required level of precision). Why might researchers be interested in finding an accurate solution to this optimisation problem? The reasons become clear when the consequences of economically-inefficient sample designs are considered. An economically-inefficient sample design, by definition, produces less power than a cheaper alternative to it, other things equal. Expressed differently, when an inefficient sample design is chosen, the power produced by it is lower than the power that **could** have been produced with the same budget. Thus, since an efficient sample design produces more "power per dollar" than an inefficient one, an optimal solution to the design problem can lower the cost of an RCT (for the same level of power), increase the power of the statistical test (for the same total cost), or do both.

Economic production theory can be applied to resolve the optimisation problem just described. In this paper, the conceptual basis of the optimisation solution is outlined and an illustration of its application is provided. The purpose of the paper is to illustrate how this solution can be used to balance the size and number of sites used in cluster designs in an optimal way.

THE EXISTING LITERATURE ON OPTIMAL DESIGN

The modern literature on optimal experimental design has largely been concerned with methods for computing optimal regression designs, and can be traced to a seminal paper by Elfving [3] and, subsequently, Keifer [4], (also see [4] for a useful review of the early literature.) In the late 1960s, Conlisk and Watts [5] produced an influential paper in which, amongst other things, considered the problem of "budget minimization" for "a pre-selected maximum admissable error" in the context of optimal regression experimental designs. Conlisk [6] subsequently extended this work by illustrating the impact of uncertainty about the functional form of a relationship on the optimal design problem.

The 1970s (particularly the late 1970s) witnessed a flurry of activity on problems of optimal design, including emphases on cost-control. Interest in the field no doubt owed something not only to the relaxation of mechanical constraints on computation, but also to numerous, large, inherently expensive, social experiments (e.g., the Graduated Work Incentive Experiment [7], the Rand Health Insurance Experiment (HIE) [8], and negative income tax experiments [6]) that were conducted during that period. A shift in emphasis from considerations of exclusively statistical dimensions of optimality (see, e.g. [4]), to those that involving economic considerations, is evident in the following statement by Conlisk (in which constant returns to sampling expenditures are implicitly assumed):

> Since the criterion *F* is a variance magnitude, this ratio is a relative efficiency measure in the usual variance sense. In dollar terms if one design has 0.5 efficiency relative to another, it means that a doubling of the budget would be needed to bring the one design up to the accuracy level of the other.
>
> [E]fficiencies can vary widely; hence efficiency questions in multi-million dollar...experiments...are quite important in dollar terms [6, p.649].

Economic considerations led Morris [9] to extend the Conlisk-Watts [5] approach to asymmetric sample applications with a finite population. His so-called "Finite Selection Model" (FSM) was developed to assign families enrolled in the Rand HIE to insurance plans "according to the general design goals of optimality (i.e., precision), balance and robustness" [9, p.44] and involves a sequential (approximate) optimisation process.

At the same time, Aigner [10, 11] considered optimal design issues in the context of electricity load control and pricing experiments, as well as providing a useful overview on matters of optimal design [12]. Aigner and Balestra [13] have also extended, in the context of error components models, the analysis of optimal experimental design to include considerations of inter-temporal controls (an idea proposed originally proposed by Hausman [14]). More recently, Aigner and Schönfeld [15] have shown how optimal design can be approached in the complex circumstances associated with direct-metering multiple appliances, in a sample of households, to measure electricity end-uses.

More recently still, several papers have been concerned with matters of design and efficiency. However, recent papers that have been concerned with efficiency (e.g., those by Cohen and Yu [16], Cohen and Machlin [17], Howes and Lanjouw [18], Van Praag, Kloek and De Leeuw [19]) have focused explicitly only on the statistical dimensions of efficiency. That economic issues have, once again "taken a back seat" in the literature on sample design is not at all surprising - much of what can be said about optimal sample design has, in the context of experiments designed to enable parameter estimation, apparently already been said. However, consider the distinction drawn by Aigner [12. p.7] in the following statement about sample design:

> The main principle that guides "good" experimental design is straightforward enough: Given **either parameter estimation or hypothesis testing** as the

ultimate purpose to which the data are to be put, one attempts to utilize a given number of observations in order to **maximize the precision of estimation or the power of a test.**

This distinction between parameter estimation and hypothesis testing is important. The existing literature on optimal design is dominated by a concern with the precision of parameter estimation, rather than with hypothesis testing *per se*. Furthermore, with one exception, those studies cited above that are expressly concerned with both the economic and statistical dimensions of efficiency are either (i) concerned exclusively with simple random or stratified random samples; or (ii) assume away any effects of clustering, albeit implicitly, in their treatment of design issues. Aigner and Schönfeld [15] are the only authors to give a conceptual treatment consider of an effect that might be considered a "clustering" effect, *viz.* "the correlation between the disturbances when the same households are used for purposes of direct metering". See also the empirical work of Fiebig, Bartels and Aigner [20], in which this issue is explicitly acknowledged. Notwithstanding, [15] and [20] are chiefly concerned with matters of parameter estimation and precision.

Thus, this paper fills a somewhat surprising, and important, gap in the existing literature on optimal design. It provides a conceptual account of an approach to optimal design that can be applied in the context of cluster sample experiments that are designed for hypothesis testing. The analysis not only fills a conceptual void, but also addresses an issue that is of practical economic importance to sample designers.

SAMPLING, AS PRODUCTION

The relationships between inputs and the quantity (and quality) of output they produce is fundamental to the economic analysis of any production process. An understanding of such relationships demands a clear definition of the relevant inputs and outputs. In this regard, it is important to distinguish between the concepts of interest and the relevant measures of them. The choice of an appropriate measure of output is not always straightforward. For example, consider the production activities of hospitals. A range of conceptions of hospital output exists (e.g., changes to health status, treatment of a case), along with a variety of output measures (e.g. number of occupied bed-days, number of admissions, number of live discharges, number of treated cases). A detailed statement on the correspondence between such concepts and their measures, in relation to hospitals, is provided in [21].

Conceptually, the output of an RCT may be viewed as "information". For the purposes of this paper, though, statistical power is arguably a useful proxy measure of this output: the reliability of the information produced by an RCT is directly related to its power. Furthermore, it is more useful for analytical purposes to treat power, rather than d_c, as the objective function, since the relevant effect size is typically determined exclusively on non-economic (e.g., clinical) grounds.

It is useful to conceive of the inputs in the production process as the number of subjects or, in the case of cluster designs, the "number of clusters" (or "number of sites") and the "number of subjects per cluster" (or the "size per site"). These inputs could be disaggregated further into, e.g. labour, capital and land. While this primary-level disaggregation is not initially useful for the purpose of this paper, the problem of primary input selection is considered later on. The primary focus of the paper is, however, on trade-offs between the number and size of sites employed in cluster designs.

Attention is now directed to a discussion of the specific characteristics of the production relationship between these statistical inputs and the (power) output of RCTs.

INPUT SUBSTITUTION AND PRODUCTION

In sample design, the researcher often is faced with a choice between several approaches that will fit the purpose of the study. First, the study designer can often choose between simple random sample and cluster designs. Then, if cluster sampling is considered, a variety of statistically-equivalent combinations of site size and the number of sites usually exists. In order to understand the role economics can play in the choice of an optimal combination of site numbers and size, the technical characteristics of the production relationship must be appreciated. Indeed, the substitutability of sampling inputs is a fundamental consideration in the process of selecting the optimal input mix in any production process. Thus, the statistical properties of cluster samples are central to the economic analysis.

Cluster samples are generally subject to larger standard errors than simple random samples (for a given sample size n): subjects in a given cluster tend to be more alike than are subjects drawn, at random, from a population [22]. This source of power loss has important statistical implications for the calculation of the sample size that is required to produce a given level of power. One measure of the magnitude of this loss of precision is the so-called "design effect" (*DE*), which is measured as

$$DE = \frac{Var_{cl}}{Var_{srs}} \quad (1),$$

where Var_{cl} is the variance of the estimate based on a cluster sample and Var_{srs} is the variance of the estimate calculated on a simple random sample of individuals. When $Var_{cl} > Var_{srs}$, a loss of power is associated with the cluster, relative to the simple random sample. The response is typically to increase *n*, at the design stage, in order to compensate for the power loss associated with the cluster design. A detailed empirical study of these effects was provided by [23] recently, in this *Journal*.

Notwithstanding the larger *n* that is generally required when a cluster design is chosen, the economies achieved by sampling in clusters may more than offset the cost of increasing *n*. In such cases, the cluster design is more cost-effective than the simple random sample. It is worth emphasising that the possibility for resource conservation arises due to the technical (or statistical) substitutability of these sample types. Substitutability, however, extends beyond that of substituting one sample method for another.

When cluster designs are considered, it is the substitutability of the site numbers and site size that becomes important. Since the design effect is directly related to the size of the clusters sampled, site size and the number of sites are, themselves, **statistical substitutes**. To select the economically-efficient combination from those available, information about the statistical relationship between sites and site size must be considered alongside information about the costs of increasing the number, or the size, of study sites. The analysis below demonstrates how these data can be wedded to select the economically-efficient combination(s) of sites and site size.

The statistical example used by Kerry and Bland [24] is employed here to illustrate the central principles. These authors were concerned with a two-sample test of means, and their purpose was to illustrate the statistical features of cluster samples. (The choice of a test of means is, for the purposes of this paper, somewhat arbitrary. The economic principles illustrated here are equally applicable to sample design for tests of proportions.) The authors were interested in a behavioural intervention, performed in general practices, that was designed to lower blood cholesterol concentrations. The minimum clinical difference of interest was a change of 0.1 mmol/l (i.e., d_c=0.1), to be tested at the five per cent level of significance, with power of 90 per cent. The required *n* for this study (for each group, in this two-sample test of means) was thus

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 s^2}{d^2} = \frac{21 s^2}{d^2} \qquad (2),$$

where s^2 is the variance of the outcome measure.

For randomised cluster samples, the relationship between statistical power and sample size is complicated by the need to account for two types of dispersion, *viz.* inter- and intra-cluster variance. Thus, for example, the total number of clusters, *c*, required to achieve 90% power, at 5% significance is calculated as

$$c = \frac{21(s_c^2 + s_w^2 / m)}{d^2} \qquad (3);$$

where c is the number of clusters, s_c^2 and s_w^2 are the inter- and intra-cluster variances, respectively, of the outcome measure, and m is the number of subjects per site.

Equations (2) and (3) describe statistical production relationships: that is, they describe the relationship between inputs (clusters, subjects per cluster, and subjects) and statistical power, or significance for specified effect or "difference". The economic production function itself may be written, more generally, as follows:

$$Power = f(c, m) \quad (4).$$

SITE NUMBERS, SITE SIZE AND THE PRODUCTION OF POWER

The optimisation problem of interest in this paper is, strictly speaking, an integer programming problem. However (paraphrasing [5, p.151]) "practically speaking, little will be lost in practice by treating [c and m] as continuous in solving [the optimisation problem] and then rounding off." Invoking this conception of the problem along with a simplifying assumption employed throughout this study, i.e. that $m_i = c/n$ for all i, consider the relationship between the statistical output, *Power*, and the inputs c and m. It is useful to employ the economic concept of the "marginal product" (*MP*) of each input in this context. The *MP* of an input is the additional output that is produced by adding an extra unit of that input, while holding all other inputs constant. For c and m, these are

$$M(Power) = MP_c = \frac{\partial(1-\beta)}{\partial c} = f_c \quad (5a)$$

and

$$M\,(Power) = MP_m = \frac{\partial(1-\beta)}{\partial m} = f_m \tag{5b),}$$

where MP_c is the marginal physical product of "clusters" and MP_m is the marginal product of "number of subjects per cluster". The second-order partial derivatives of (4) may be written as follows:

$$\frac{\partial MP_c}{\partial c} = \frac{\partial^2(1-\beta)}{\partial c^2} = f_{cc} < 0 \tag{6a}$$

and

$$\frac{\partial MP_m}{\partial m} = \frac{\partial^2(1-\beta)}{\partial m^2} = f_{mm} < 0 \tag{6b).}$$

The characteristics described in (6a) and (6b) indicate that the statistical production process is subject to "diminishing marginal returns" with respect to both "clusters" and "subjects per cluster". In other words, increasing c or m while holding all other inputs constant increases output at a decreasing rate. It is worthwhile to note that diminishing marginal returns are not imposed by the solution but are a natural property of this and, indeed, many other production processes.

The marginal products of inputs are important in production inasmuch as these provide information about the impact, on output, of a change in one input quantity. More usefully, though, this information can be used to determine which combinations of two inputs can be used to produce any given level of output. Mathematically,

$$Power = f\,(c, m) = 1 - \beta \tag{7}$$

Equation (7) is an isoquant for a given level of power: it represents all combinations of the inputs c and m that produce the $1-\beta$ level of statistical power (/significance).

(Literally, "isoquant", means "same quantity".) Along any given isoquant, the technical possibilities for substitution are derived by setting the total differential of the production function (4) equal to zero, and hence we may write:

$$\frac{MP_c}{MP_m} = MRTS_{cm} = \left.\frac{-dm}{dc}\right|_{1-\beta} \tag{8}$$

where $MRTS_{cm}$ is, literally, "the marginal rate of technical substitution of clusters for subjects per cluster". Expressed more intuitively, this is the rate at which an increase in the number of clusters can be met with a decrease in the number of subjects per cluster (or *vice-versa*), without affecting the power of the design. It is equivalent to the ratio of the marginal product of "clusters" to the marginal product of "subjects per cluster" because this ratio also indicates the quantity of c that must be foregone for a given change in m in order to hold power constant. Finally, that the greater (smaller) is the marginal product of m, relative to c, the greater (smaller) the quantity of c that must be foregone to offset the output change that results from increasing m. For this reason (i.e. that of diminishing marginal returns), the $MRTS_{cm}$ varies along the length of the isoquant. Indeed, the slope of the isoquant is simply the $-MRTS_{cm}$. (Note that, according to this discussion, equation (2) can be viewed the isoquant for the 90 per cent power and five per cent significance levels for the study referred to in [24].)

The practical importance of the discussion highlights an important characteristic, not only of the production of statistical power in cluster samples, but of most production processes. Namely, a variety of combinations of inputs can usually be used to produce a desired level of output; but the productivity, and hence substitutability, of each input depends on its relative quantity.

This rather simple observation can be used to economic advantage: although different combinations of clusters and subjects can be used to produce an identical statistical result, these combinations are unlikely to have identical costs. Subsequently, it will be shown that equating the $MRTS_{cm}$ to the sampling costs associated with subjects and clusters yields the optimal economic choice from the technical possibilities available. First, we employ an example from the literature to illustrate the relationships introduced so far.

AN ILLUSTRATION WITH DATA

To cement concepts, we now consider an example from the literature in the context of production theory. Kerry and Bland [24], in their note on cluster randomisation describe a trial to study the effect of a behavioural intervention, undertaken in general practice (GP), to lower serum cholesterol values. In this study, the intervention group was provided with an intensive, behavioural, dietary intervention by practice nurses, while the control group received usual (GP) care. The outcome measure in this study was to be the mean cholesterol value for patients in each group, one year later. The minimum clinical difference of interest in the study was a change of 0.1 mmol/l (i.e., d_c=0.1) to serum cholesterol levels.

Table 1 presents data that were calculated using the difference (d_c=0.1), variance (s_m^2=1.28 s_c^2=0.0046), and power (1-β=0.90) and significance ($\alpha/2$=0.05) levels adopted in [24]. Column (i) presents the number of patients per cluster (or practice) (m) and is subject to an upper bound of 500 – which represented the maximum size of the general practices in the population of interest. The chosen values of m are somewhat arbitrary, but are useful for the purposes of illustration. Column (ii) presents the standard deviation, and accounts for both inter- and intra-cluster deviations, where relevant. Column (iii) contains the number of clusters (c) that is required to achieve the desired power level **given** the number of patients per practice chosen in column (i). (The data in column (iii) were calculated by applying equation (3), above, to the data in columns (i) and (ii).) The data in column (iv) are twice the product of the data in columns (i) and (iii), i.e. column (iv) gives the total number of subjects ($2n$) required for this two-sample study. Column (v) provides the calculated design effect, (1); column (vi) contains Arabic labels for each (c,m) combination, and column (vii) provides the $MRTS_{cm}$, calculated, according to equation (8), for $-dm = 1$.

[TABLE 1 ABOUT HERE]

Each of the combinations in Table 1 produces (approximately, due to rounding) the same level of statistical power and significance. That is, Table 1 provides isoquant data and shows that there is input (c and m) substitutability in the production of statistical power. The data in column (vii) also show that the inputs are imperfect substitutes, i.e. $MRTS_{cm} \neq 1$ for all c and m combinations. Rather, the $MRTS_{cm}$ is an increasing function of m because, all else equal, the MP_m falls as m increases. A comparison of the data in columns (v) and (vii) also reveals that there is a direct relationship between the *DE* and $MRTS_{cm}$.

A more complete depiction of the isoquant described in Table 1 is provided, geometrically, by the curve 1-β_{10} in Figure 1. The combinations labelled *A* to *I* in Table 1 appear on the isoquant depicted in this figure. Note, though, that all combinations of *c* and *m* that lie on 1-β_{10} also produce the same level of statistical power, *viz.* 90 per cent. Statistically, these combinations are perfect substitutes. Recall that the slope of the isoquant at any point = -$MRTS_{cm}$. It is also noteworthy that a "special case" of cluster randomisation, in which $m=1$ and $c=2682$, constitutes (one outcome of) a simple random sample.

[FIGURE 1 ABOUT HERE]

It should also be noted that the isoquant depicted in Figure 1 is one of an infinite population of isoquants. Specifically, an isoquant may be drawn for any desired level of statistical power and the lower (higher) is the level of statistical power, the closer (farther) the isoquant will be from the origin. It is also noteworthy that each of these isoquants will be convex to the origin (because the marginal products of *c* and *m* change according to their relative quantities) and will not cross one another (since, for a specified d_c, any combination of *c* and *m* produces a unique level of statistical power). A more detailed statement of these general properties of isoquants may be found in [25], for example.

ECONOMICALLY-EFFICIENT SAMPLE DESIGN

As Figure 1 illustrates, the sample design possibilities are not limited to those shown in Table 1. Indeed, at least 500 different designs are available to the researcher in this example, including 499 different combinations of c and m, and the simple random sample. How is the least cost design to be chosen, from those available?

To solve the optimisation puzzle, we frame the problem as that of a trade-off between the "number of subjects (n)" and the "number of clusters (c)" and write:

$$Power = f(c, n) \qquad (9).$$

Recall, too, that the technical trade-offs that result from setting the total differential of the production function to zero were described as follows:

$$\frac{MP_c}{MP_n} = MRTS_{cn} = \left.\frac{-dn}{dc}\right|_{1-\beta} \qquad (8).$$

Table 2 provides the $MRTS_{cn}$ for the combinations that were described in Table 1. Figure 2 depicts the isoquant from which the combinations in Table 2 are taken.

[TABLE 2 ABOUT HERE]

The relationship between sample design and cost can now be considered by introducing a total cost function or "budget constraint":

$$TC = P_c c + P_n n \qquad (10),$$

in which P_c is the price per cluster and P_n is the price per subject. Thus, the total cost of the sampling exercise depends on both the number of clusters and the number of subjects chosen, along with their unit prices. In practical terms, P_c may be thought of as the fixed cost of establishing each separate cluster for the trial. It could include, for example, the cost of corresponding with and training of personnel at cluster locations. P_n may be conceived as the additional cost that is incurred each time an additional subject is included in the trial. For simplicity, and without loss of generality, it is useful to assume initially that P_c and P_n are constants.

The first order conditions for the cost minimisation problem, given $1-\beta_i = f(c,n) = 1-\beta_{10}$, can be derived by setting up the Lagrangean:

$$L = P_c c + P_n n + \lambda[f(c,n) - (1-\beta_{10})] \quad (11).$$

For a proper relative minimum, it is necessary that

$$\frac{\partial L}{\partial c} = P_c - \lambda \frac{\partial f}{\partial c} = 0 \quad (12a)$$

$$\frac{\partial L}{\partial n} = P_n - \lambda \frac{\partial f}{\partial n} = 0 \quad (12b)$$

$$\frac{\partial L}{\partial \lambda} = f(c,n) - (1-\beta_{10}) = 0 \quad (12c);$$

Or, dividing (12a) and (12b), that

$$\frac{P_n}{P_c} = \frac{\partial f / \partial n}{\partial f / \partial c} = MRTS_{cn} \quad (12d)$$

Equation (14d) produces the following, general result: the cost-minimising design can be found by equating the marginal rate of technical substitution of the two inputs with the ratio of their prices.

Consider, for example, the (hypothetical) case in which P_c=\$31.25 and P_n=\$10. P_n/P_c =3.125, and the design that has an $MRTS_{cn}$ closest to this price ratio is that for which n=3,000 and c=30 (and $MRTS_{cn}$=3.127). Thus, the optimal number of clusters is 30 and, since $m=n/c$, the optimal size of each cluster =3,000/30=100. (See Table 2.) Given the input prices, the minimum total cost per sample is TC=\$31.25(30)+\$10(3,000)=\$30,437. (Since the study referred to here involves a two-sample test of means, the minimum budget required would thus be \$60,874.)

This solution is depicted in Figure 2 by the point of tangency, *B*, between the isocost (literally "same cost") line, *ZW*, and the isoquant, $1\text{-}\beta_{10}$. The isocost *ZW* depicts all combinations of *c* and *n* that cost \$30,437. It has a slope of $-P_n/P_c$., i.e. the slope is determined by the relative prices of the inputs. Notably, no other combination of *c* and *n* that is depicted on *ZW* produces as much statistical power as *B*. Rather, all other combinations on *ZW* produce $1\text{-}\beta_i < 1\text{-}\beta_{10}$. For a more detailed discussion of the general properties of isocost curves see, e.g. [25].

The cost-minimisation problem just explored is, of course, also amenable to analysis as an output-maximisation problem, given a specified budget constraint (such as *ZW)*. Once again, the solution to the problem is provided by (14d), *viz.* given a specific budget, the power of the sample is maximised by selecting the combination of *c* and *n* so that the ratio of their prices is equal to the $MRTS_{cn}$.

EMPIRICAL COMPLICATIONS: FIXED COSTS, ECONOMIES OF SCALE AND UNCERTAINTY

It was assumed, above, that P_c and P_n are constants. By implication, this means that the prices of inputs are independent of the quantities employed. This assumption, while perhaps not unrealistic, may not be appropriate for some studies. For example, the unit cost of n may not be constant: for an increasing m, the unit cost per subject could decline (the case of scale economies) and/or increase (the case of scale diseconomies). Alternatively, or additionally, the price per cluster P_c might not be a constant, but could increase or decrease over particular values of c.

Scale effects of this kind can give rise to non-linearities in the isocost curve and, as a result, the cost-minimisation/output-maximisation problem may not have a unique solution. That is, several statistically-equivalent sample designs may also be equally economically-efficient. In practice, this complication may be dealt with by calculating the price ratios produced by the expected unit costs for c and m as their magnitudes change, to identify the optimal design(s).

Another implicit assumption that was invoked, above, is that Equation (10) is an exhaustive account of all relevant inputs and their prices. This assumption may also be too simplistic, in practice. For example, some plant, infrastructure and administrative costs may be unrelated to the quantities of c and m that are chosen. In such cases, the budget constraint may be written as

$$TC = FC + P_c c + P_n n \qquad (13),$$

where FC is the fixed cost of sampling and is independent of quantities of c and m chosen.

The existence of fixed costs does not, however, complicate the problem of optimal sample design: these costs are independent of the chosen input combination, so the cost-minimising (or output-maximising) design is determined in the manner outlined above, and summarised in (12d).

The analysis presented here does not deal explicitly with the optimal selection of primary inputs (e.g., labour, land, capital). However, it should be noted that, if the combination of these inputs for the production of "clusters" and "number of subjects per cluster" is sub-optimal, the chosen quantities of c and m are also likely to be sub-optimal. That is, condition (12d) is a necessary, but not a sufficient condition for the design of an efficient cluster-randomised study. Rather, to guarantee technical efficiency, the ratios of prices and the rates of technical substitution between all primary inputs must also be equal. This is, however, a general requirement for productive efficiency and accessible accounts may be found in most texts in microeconomics, including [25].

A further matter that may concern practitioners is the problem of uncertainty regarding the true magnitudes of the intra- and inter-cluster variances, since these are central to the calculations performed above. The resulting problem of power miscalculations is not, of course, only economic. However, a typical economic responses to this kind of uncertainty might be either to (i) estimate the costs and benefits of underestimating these variances; or (ii) test the sensitivity of the study's power and costliness to the expected range in which these are considered likely to fall. Generally, the costs of overestimating statistical power are considerable in trials.

Thus, the result of a systematic economic approach to this particular issue is unlikely to yield any real additional benefits over the conservative approach to power calculations adopted by convention.

Finally, it should be noted that the complications introduced by multi-stage cluster sampling, stratification, and so on, have not been addressed explicitly here. For a recent discussion of the statistical implications of the simultaneous occurrence of these phenomena, see [18].

PRACTICAL IMPLICATIONS AND CONCLUSION

The practical importance of resource conservation hardly requires emphasis in the context of clinical trials. As practitioners well know, the resources available to conduct trials are typically limited and are therefore usually subject to considerable scrutiny from one or more of the players in the game, i.e. funders, investigators, facilitating organisations, subjects. For this reason, the costs and consequence of alternative scenarios, such as those that involve different numbers of clusters and cluster sizes are commonly weighed in practice.

The contribution of this paper is to place such considerations in a framework that has a strong theoretical pedigree in economics. While the elucidation of this framework has been detailed, it is worth emphasis that the general result does provide practitioners with a practical and relatively non-technical way of simplifying the search for an optimal trial design. Via a straightforward application of the optimisation rule – presented as (12d), above – the search for an optimal combination of clusters and cluster sizes can be simplified considerably. Furthermore, an application of this systematic approach to cluster design may reassure funders that not only will a given trial design produce important clinical information, but also do so at least cost. This may be a particularly important exercise in large and expensive trials.

At the same time, it is worth recognising that the systematic approach described in this paper will not always produce large efficiency gains over the types of aggregate costs comparisons that the architects of clinical trials are apt to consider anyway. If the number of alternative designs is strictly limited by, for example, institutional factors – and the limit case is that in which is where there is only *one* possible design – economics could have little, or nothing, to offer the trial architect. In less restrictive circumstances, though, the economic approach outlined here may be useful.

Finally, it is worth noting that a number of issues, which have not been addressed in this paper, provide opportunities for extensions to the work. The joint economic and statistical impacts of problems of attrition, multi-stage sampling, sample stratification, and so on, might usefully be considered in extensions of this work.

REFERENCES

1. Selvin, S, *Statistical Analysis of Epidemiological Data*. Oxford: Oxford University Press; 1996.
2. Deaton A and Muellbauer J, *Economics and Consumer Behavior*. Cambridge: Cambridge University Press; 1980.
3. Elfving G, Optimum allocation in linear regression theory, *Ann Math Statist* 1952; 23: 255-62.
4. Kiefer J, Optimum experimental designs, *J R Statist Soc* B 1959; 21: 272-304.
5. Conlisk J and Watts H, A model for optimizing experimental designs for estimating response surfaces, *Proceedings of the American Statistical Association* 1969, Social Statistics Section; 150-56.
6. Conlisk J, Choice of response functional form in designing subsidy experiments, *Econometrica* 1973; 41: 643-56.
7. American Economic Association: Current status of income maintenance experiments, session, *Papers and Proceedings* 1971; 61: 15-42.
8. Newhouse J P and the Insurance Experiment Group, *Free for All? Lessons from the Rand Health Insurance Experiment.* Cambridge (Mass.): Harvard University Press, 1993.
9. Morris C, A finite selection model for experimental design of the health insurance study, *J Econom* 1979; 11: 43-61.
10. Aigner DJ, Bayesian analysis of optimal sample size and a best decision rule for experiments in direct load control, *J Econom* 1979; 9: 209-21.

11. Aigner DJ, Sample design for electricity pricing experiments, *J Econom* 1979; 11: 195-205.

12. Aigner DJ, A brief introduction to the methodology of optimal experimental design, *J Econom* 1979; 11: 7-26.

13. Aigner DJ and Balestra P, Optimal experimental design for error components models, *Econometrica* 1988; 56: 955-71.

14. Hausman JA, The effects of time in economic experiments. In Hildenbrand W (ed) *Advances in Econometrics* (pp.79-108). Cambridge: Cambridge University Press, 1982.

15. Aigner DJ and Schönfeld P, Experimental design for direct metering of residential electricity end-uses. In Gabszewicz JJ, Richard J-F and Wolsey LA (eds) *Economic Decision-Making: Games, Econometrics and Optimisation*. Amsterdam: Elsevier Science, 1990.

16. Cohen SB and Yu WW, The impact of alternative sample allocation schemes on the precision of survey estimates derived from the national medical expenditure panel survey, *J Econ Soc Meas* 2000; 26: 111-128.

17. Cohen SB and Machlin SR, Survey attrition considerations in the medical expenditure panel survey, *J Econ Soc Meas* 2000; 26: 83-98.

18. Howes S and Lanjouw P, Does sample design matter for poverty rate comparisons?, *Rev Income Wealth* 1998; 44: 99-109.

19. van Praag B, Kloek T and De Leeuw J, Large-sample properties of method of moment estimators under different data-generating processes, *J Econom* 1988; 37: 157-69.
20. Fiebig, DG, Bartels R and Aigner DJ, A random coefficient approach to the estimation of residential end-use load profiles, *J Econom* 1991; 50: 297-327.
21. Butler JRG, *Hospital Cost Analysis*. Dordrecht: Kluwer; 1995.
22. Carlin JB and Hocking J, Design of cross-sectional surveys using cluster sampling: An overview with Australian case studies, *Aust N Z J Public Health*, 1999; 23:546-551.
23. Smeeth L and Ng E S-W, Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC Trial of the Assessment and Management of Older People in the Community, *Control Clin Trials* 2002; 23: 409-21.
24. Kerry SM and Bland JM, Sample size in cluster randomisation. *Br Med J* 1998; 316:549.
25. Nicholson W, *Microeconomic Theory: Basic Principles and Extensions*. Fort Worth: Dryden Press; 1998.

TABLE 1
TRADE-OFFS BETWEEN THE NUMBER AND SIZE OF CLUSTERS

No. of Subjects per Cluster (*m*) (i)	**Standard Deviation (S.D.)** (ii)	**No. of Clusters, Given *m* (*c*)** (iii)	**Total no. of Subjects (2*n*)** (iv)	**Design Effect (*DE*)** (v)	**Label** (vi)	***MRTS*$_{cm}$** (vii)
Individual randomisation	0.0046	-	5,364	1.000	-	-
10	0.3641	279	5,580	1.040	A	0.033
30	0.2174	100	6,000	1.119	B	0.324
50	0.1738	64	6,400	1.193	C	0.911
100	0.1319	37	7,400	1.380	D	3.683
150	0.1146	28	8,400	1.566	E	8.315
200	0.1049	24	9,600	1.790	F	14.807
300	0.0942	19	11,400	2.125	G	33.371
400	0.0883	17	13,600	2.532	H	59.375
500	0.0846	16	16,000	2.983	I	92.820

Notes:
(a) The calculations of *c*, *m* and *n*, contained in this table, are based on d_c=0.1, α=0.05, β=0.10, and $s_m^2 = 1.28$ $s_c^2 = 0.0046$.
(b) Data on the number of clusters, indicated in column (iii), have been rounded to the next higher integer. Data in columns (iv) and (v) were calculated from the rounded data.
(c) Data in columns (ii), (v) and (vii) were been rounded to three decimal places.
(d) $MRTS_{cm}$ is the marginal rate of technical substitution of 'number of clusters' for 'number of subjects per cluster' and is calculated for unit changes in *m* at $\beta = 0.10$ (cf. Equations (8) and (9)).

FIGURE 1
TRADE-OFFS BETWEEN THE NUMBER OF CLUSTERS (*c*) AND THE SIZE OF CLUSTERS (*m*)

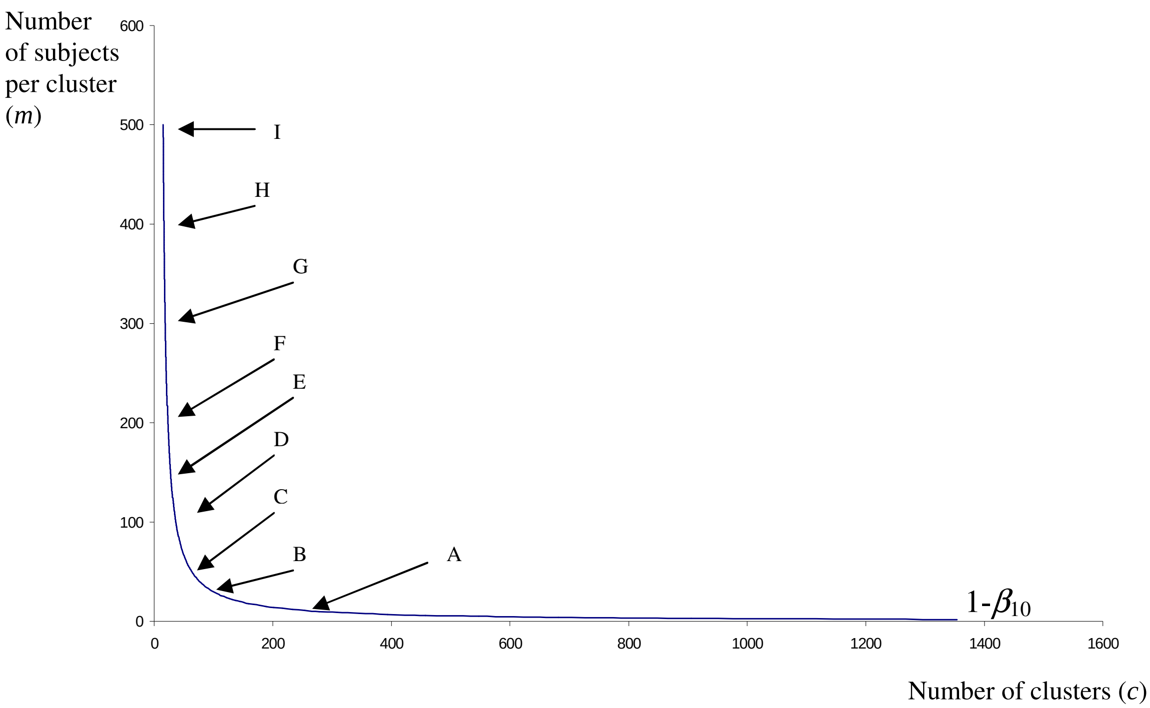


TABLE 2
TRADE-OFFS BETWEEN THE NUMBER OF CLUSTERS AND THE NUMBER OF SUBJECTS

No. of Subjects per Cluster (*m*) (i)	**No. of Clusters, Given *m* (*c*)** (ii)	**No. of Subjects (*n*)** (iii)	**Label** (iv)	***MRTS$_{cn}$*** (v)
10	279	2,790	A	0.323
30	100	3,000	B	3.127
50	64	3,200	C	8.805
100	37	3,700	D	35.578
150	28	4,200	E	80.320
200	24	4,800	F	143.031
300	19	5,700	G	322.359
400	17	6,800	H	573.563
500	16	8,000	I	896.641

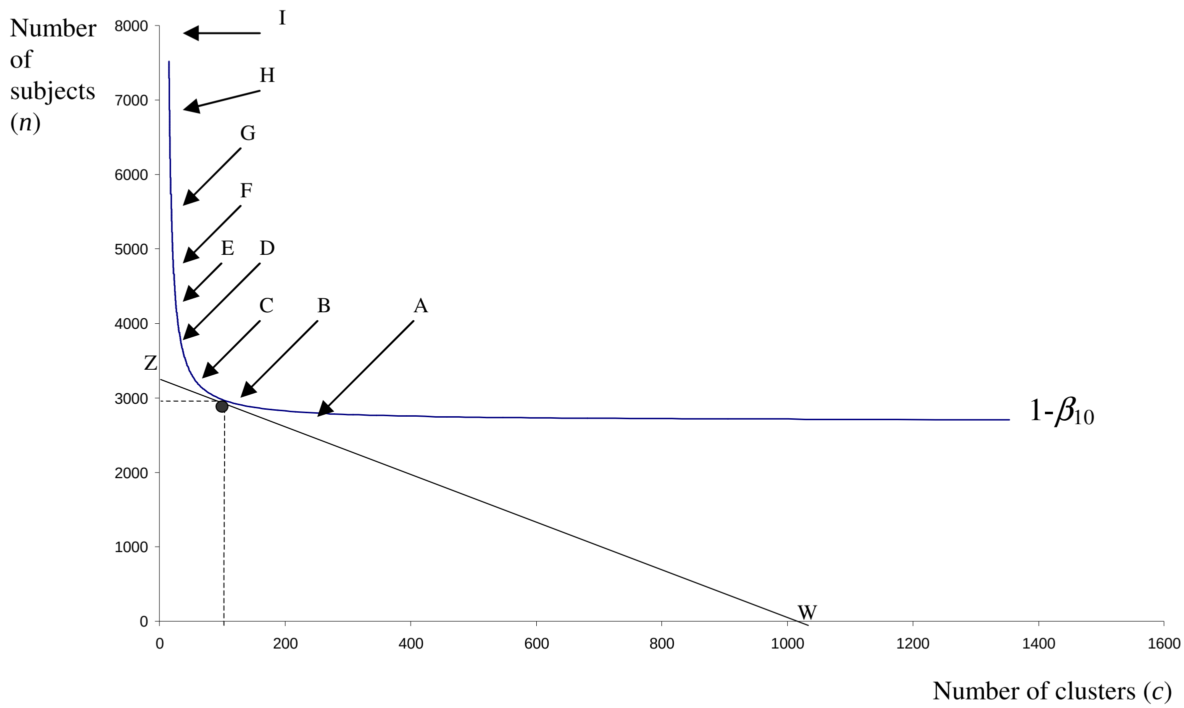
Notes: (a) The calculations of *c*, *m* and *n*, contained in this table, are based on d_c=0.1, α=0.05, β=0.10, and $s_m^2 = 1.28$ $s_c^2 = 0.0046$.

(b) Data on the number of clusters, indicated in column (ii), have been rounded to the next higher integer. Data in columns (iv) and (v) were calculated from the rounded data.

(c) Data in column (v) have been rounded to three decimal places.

(d) $MRTS_{cn}$ is the marginal rate of technical substitution of 'number of clusters' for 'number of subjects' and is calculated for unit changes in *n* at $\beta = 0.10$.

FIGURE 2
TRADE-OFFS BETWEEN THE NUMBER OF CLUSTERS (*c*) AND THE NUMBER OF SUBJECTS (*n*)



ACKNOWLEDGEMENTS

I wish to acknowledge the helpful comments of Dr H.Shelton Brown III, on an earlier draft of this paper. I also acknowledge helpful conversations with Dr Darrel Doessel and Professor Denzil Fiebig. I also wish to thank two anonymous referees whose comments led to substantive improvements to this paper. Needless to say, any errors that appear in this paper are my responsibility, alone.