Department of Statistical Science
Southern Methodist University
Technical Report SMU-TR-302
March 2002

Optimal One-way Random Effects Designs for the Intraclass Correlation Based on Confidence Intervals

Brent D. Burch

Department of Mathematics and Statistics
Northern Arizona University
Flagstaff, Arizona 86011
(brent.burch@nau.edu)

Ian R. Harris

Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275
(iharris@mail.smu.edu)

Optimal One-way Random Effects Designs for the Intraclass Correlation Based on Confidence Intervals

Brent D. Burch

Department of Mathematics and Statistics
Northern Arizona University
Flagstaff, Arizona 86011
(brent.burch@nau.edu)

Ian R. Harris

Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275
(iharris@mail.smu.edu)

Confidence intervals for the intraclass correlation coefficient (ρ) are used to determine the optimal allocation of experimental material in one-way random effects models. Designs that produce narrow intervals are preferred as they allow more precise inference about the value of ρ. The authors investigate the number of classes and the number of observations per class required to minimize the expected length of confidence intervals. We obtain results using asymptotic theory and compare these results to those obtained using exact small-sample calculations. The results suggest that for fixed sample size and fixed number of classes (or groups), one should select a balanced design, or the design which is closest to balanced. If one is allowed to choose the number of groups, the best design depends on the unknown value of ρ. A good overall recommendation, which appears to work well for all values of ρ, is to choose a design having group sizes of about 4 each.

KEY WORDS: Expected length; Variance components; Optimal allocation.

1 Introduction

The one-way random effects model is an important design due to its wide range of applicability. For example (Vangel, 1992), in industrial applications where a product is manufactured in batches, this model serves as a tool to highlight how the batch variability influences the variability in the finished goods. As a second illustration (Gibbons and Bhaumik, 2001), interlaboratory studies are conducted to determine how the variability of measurements between laboratories relates to the variability of measurements within laboratories. The source of variation under study, be it batches in a manufacturing process or laboratories in a interlaboratory study, are random effects in the model if the intent of the investigator is to quantify how variability in a population effect the measurements.

In the above examples, the parameters of interest are the variance components associated with the random effects in the model. A particular function of the variance components, the intraclass correlation coefficient, is the correlation between measurements in the same class (or group) of the random effect. If σ_1^2 is the variance of the observations between groups and σ_2^2 is the variance of the observations within groups and all the effects combine linearly, then the intraclass correlation coefficient, denoted by $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$, is also the proportion of total variation in the measurements due to the source of variation under study.

In many instances the objective of the study is to estimate the intraclass correlation coefficient while being constrained by a limited number of experimental units. When resources are restricted, the investigator must judiciously select the number of groups and/or the number of measurements per group. The optimal allocation of resources in a one-way random effects model is the subject of this paper.

According to Scheffé (1956) and Searle, Casella, and McCulloch (1992), the first explicit use of one-way random effects models was made by Airy (1861). Without explicitly referring to the one-way random effects model, Bessel (1838) recommended 16 repeated observations in connection with the measurement of the parallax of the star 61 Cygni, and in 1839 recommended 7 repeated observations when measuring distances of craters on the moon through a telescope (Koziel, 1985). The optimal number of measurements per group was also discussed by Chauvenet (1863a, 1863b) using data from an astronomical sighting experiment. Chauvenet (1863a) recommended 5 to 7 repeated observations of right ascension of a star, and later compared the results obtained from the 7 observations to that of the 5 observations (1863b). He concluded that more than this would be wasteful, and not reduce "total error" very much. These early recommendations on designs seem to have been concerned with the error of the average measurement, $\sigma_1^2 + \sigma_2^2/b$, where b is the number of measurements per group. It appears that the astronomers used initial data to get estimates of the variance components and then decided how large b should be to make this error close to the unattainable σ_1^2 that would arise if there was no measurement error.

Tippett (1931) concluded that for estimating the variance components in a one-way random effects model, an arrangement with many groups and few individuals per group is better than one with few groups and many individuals per group. Hammersley (1949) and Anderson and Crump (1967) minimize the variances of the ANOVA estimators of the individual variance components in order to obtain optimal allocation results. Anderson and Crump (1967) also considered estimating σ_1^2/σ_2^2, which we call ϕ. The authors noted that the optimal design is a function of the unknown parameter values. In particular, as the total sample size goes to infinity, the optimal value of b for estimating σ_1^2 and ϕ are $1 + 1/\phi$ and $2 + 1/\phi$, respectively. As a function of ρ, the optimal group sizes are $1/\rho$ and $1 + 1/\rho$, respectively. When estimating σ_1^2 or ϕ, one should use a few large groups if ρ is small but many small groups if ρ is large.

The references mentioned above do not directly address the selection of a design when estimating the intraclass correlation coefficient. Donner and Koval (1982) took up this problem and indicate that the balanced design is usually preferred, but do not cover the scenario where many competing balanced designs having the same number of total observations are under consideration. Shoukri and Ward (1984) and Walter, Eliasziw, and Donner (1998) consider optimal sampling designs for ρ based on hypothesis test requirements. Walter, Eliasziw, and Donner (1998) conclude that when $\rho \geq 0.4$ for balanced designs in which $\alpha = 0.05$ and $\beta = 0.20$, groups of size 2 or 3 will minimize the total number of observations. Lohr (1995) examines optimal designs for the one-way random effects model from a Bayesian perspective. Optimal group sizes used to estimate ρ in balanced designs range from just over 3 to less than 12, depending on the form of the prior distribution. Assuming a uniform prior for ρ, one can show that the optimal group size is $1 + \sqrt{10}$, a noninteger value. The Bayesian results given above correspond to those obtained using A-optimality, in which the average asymptotic variance of a point estimator of ρ is minimized.

Unlike previous literature on the subject, we consider confidence intervals for ρ as a measure of the quality of the design. For equal-tailed intervals having a fixed level of confidence, short intervals are desirable as they indicate with a high degree of certainty plausible values of ρ. We investigate the optimal allocation of resources in terms of number of groups and number of observations per group in order to minimize the expected length of confidence intervals for ρ. We calculate expected length using two methods. The first is an exact calculation based on a pivotal quantity for ρ. While this approach is valid for both small and large sample applications, it can be computationally intensive and theoretical results are absent. The second method, which is approximate, is based on the asymptotic normal approximation to the ANOVA estimator of ρ. We obtain theoretical results based on this second method and use empirical results based on the first to justify the theory. In some cases the allocation of resources depends on the true value of the parameter ρ. We suggest removing the dependence by considering the average expected length of confidence intervals or the minimax expected length of confidence intervals for ρ.

The paper is organized as follows. Section 2 provides background information and introduces the notation used in the one-way random effects model. Included in the discussion are the computational aspects associated with the expected length of confidence intervals for ρ. In Section 3, examples are given to help illustrate the use of the optimal allocation procedure described in this paper. Optimal designs are based on large and small-sample computations of the expected length of confidence intervals for ρ. The designs recommended in this paper do not depend on the value of ρ. Section 4 presents a discussion and summary.

2 The One-way Random Effects Model

Consider the one-way random effects model given by

$$Y_{ij} \quad = \quad \mu + A_i + e_{ij}, \tag{1}$$

where $i = 1, ..., a$, $j = 1, ..., b_i$, and $\sum_{i=1}^{a} b_i = n$. Y_{ij} is the j^{th} observation associated with the i^{th} class (or group) of factor A. The a groups of A in the model are assumed to be randomly selected from some large population of groups. Furthermore, a random sample of size b_i has been obtained from the i^{th} group. e_{ij} is often referred to as random error. It is assumed that $A_i \overset{iid}{\sim} N(0, \sigma_1^2)$, $e_{ij} \overset{iid}{\sim} N(0, \sigma_2^2)$, and that A_i and e_{ij} are mutually independent. In addition, $\sigma_1^2 \geq 0$ and $\sigma_2^2 > 0$. μ is a fixed but unknown quantity that represents the overall mean of Y_{ij}.

Recall that $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$, which may be interpeted as the correlation between two observations within the same group or as the proportion of the variation in the Y_{ij}'s attributed to factor A. By definition, $0 \leq \rho < 1$. Since ρ is a function of variance components, and the objective is to select the design which provides the most information about ρ, we begin by examining the properties of a set of quadratic forms used to estimate the variance components. These quadratic forms, denoted by $Q_1, ..., Q_d$, form a set of minimal sufficient statistics associated with the reduced linear model void of the parameter μ. The number of quadratic forms and their corresponding distributions depend on the underlying model structure.

The properties of the quadratic forms may be obtained by diagonalizing the variance-covariance matrix of a linear transformation of the observations. Specifically, if $0 \leq \Delta_1 < ... < \Delta_d$ represent distinct eigenvalues of that part of the aforementioned variance-covariance matrix associated with σ_1^2, and each Δ_m has multiplicity r_m, then it can be shown that

$$Q_m \sim \frac{\sigma_2^2}{1-\rho}\Big(1+\rho(\Delta_m - 1)\Big)\chi^2(r_m), \tag{2}$$

where $m = 1, ..., d$. By construction, $Q_1, ..., Q_d$ are independent. A complete description of the distributional theory associated with the quadratic forms in a one-way random effects model is described by LaMotte (1976). Burch and Iyer (1997) discuss the theory used to construct the quadratic forms and associated eigenvalues in a more general setting.

The total variation in the observations, given by $\sum_{i=1}^{a}\sum_{j=1}^{b_i}\left(Y_{ij} - \overline{Y}_{..}\right)^2$ where $\overline{Y}_{..}$ is the overall sample mean, may be partitioned (see LaMotte, 1976) as

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b_i}\left(Y_{ij} - \overline{Y}_{..}\right)^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b_i}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + \sum_{i=1}^{a} b_i\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\ &= Q_1 + Q_2 + ... + Q_d. \end{aligned} \tag{3}$$

Of particular interest is the fact that $\Delta_1 = 0$ if at least one $b_i > 1$. The zero eigenvalue signifies that there is replication in the experiment (multiple observations per group) and thus an estimate for σ_2^2 is readily available. For the one-way random effects model, $Q_1 \sim \sigma_2^2\chi^2(r_1)$ with $r_1 = n - a$. In addition, $\sum_{i=1}^{a}\sum_{j=1}^{b_i}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 = Q_1$ and it follows that $\sum_{i=1}^{a} b_i\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 = Q_2 + ... + Q_d$, which is distributed as a linear combination of scaled chi-squared variates. The analysis of variance table for any one-way random effects model of the form (1) is given in Table 1.

Consider the example presented by Vangel (1992) in which tensile-strength measurements were made on five consecutive batches of composite material used to make aircraft components. Five measurements per batch were obtained for a total sample size of 25. The coded strength measurements for this balanced model are displayed in Table 2. The focus of attention is not on the actual measurements, but on the design itself. For this design, $d = 2$ where $\Delta_1 = 0$ and $\Delta_2 = 5$. In addition, $r_1 = 20$ and $r_2 = 4$. Using (2) we find that $Q_1 \sim \sigma_2^2\chi^2(20)$. The 20 degrees of freedom are generated from the five groups, each group contributing 4

Table 1: ANOVA Table

Source	df	Sum of Squares
Between Groups	$a-1$	$Q_2 + ... + Q_d$
Within Groups	$\sum_{i=1}^{a} b_i - a$	Q_1
Total	$\sum_{i=1}^{a} b_i - 1$	$\sum_{i=1}^{a} \sum_{j=1}^{b_i} (Y_{ij} - \overline{Y})^2$

Table 2: Balanced Design of 25 Observations

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
379	357	390	376	376
363	367	382	381	359
401	402	407	402	396
402	387	392	395	394
415	405	396	390	395

degrees of freedom so that $5(4) = 20$. The second eigenvalue is equal to the number of replicates per group and the degrees of freedom associated with the second quadratic form is the number of groups minus 1. In general, for a balanced one-way random effects model having a groups and $b = n/a$ replicates per group, $d = 2$ with $\Delta_1 = 0$, $r_1 = a(b-1)$, $\Delta_2 = b$, and $r_2 = a - 1$. In this case the sum of squares column in Table 1 contains only two quadratic forms with $Q_2 = b\sum_{i=1}^{a} (\overline{Y}_{i.} - \overline{Y}_{..})^2$.

Table 3 displays a configuration for an unbalanced model using a sample of size 25. In

Table 3: Unbalanced Design of 25 Observations

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6	Batch 7	Batch 8	Batch 9
Y_{11}	Y_{21}	Y_{31}	Y_{41}	Y_{51}	Y_{61}	Y_{71}	Y_{81}	Y_{91}
Y_{12}	Y_{22}	Y_{32}	Y_{42}	Y_{52}	Y_{62}	Y_{72}	Y_{82}	Y_{92}
–	–	Y_{33}	Y_{43}	Y_{53}	Y_{63}	Y_{73}	Y_{83}	Y_{93}

this case, $d = 4$ with $\Delta_1 = 0$, $r_1 = 16$, $\Delta_2 = 2$, $r_2 = 1$, $\Delta_3 = 2.16$, $r_3 = 1$, $\Delta_4 = 3$, and $r_4 = 6$. The values of Δ_2 and r_2 are due to the two replications per group in the first two groups. The values of Δ_4 and r_4 are due to the three replications per group in the last seven groups. The noninteger eigenvalue is a result of the unbalanced nature of the design.

For unbalanced designs having two distinct group sizes, there are at most three nonzero eigenvalues. In general, suppose there are c groups each of size x and $a-c$ groups each of size y, where $x < y$ and $cx + (a-c)y = n$. It can be shown (see the Appendix for details) that $\Delta_1 = 0$, $r_1 = n - a$, $\Delta_2 = x$, $r_2 = c - 1$, $\Delta_3 = axy/n$, $r_3 = 1$, $\Delta_4 = y$, and $r_4 = a - c - 1$.

As the total sample size is the same for the designs displayed in Tables 2 and 3, it would be interesting to determine which setup provides more information about ρ. This question, as well as similar questions, will be addressed later in this article.

2.1 Exact Confidence Intervals for ρ

Equation (2) indicates that the distributions of the quadratic forms involve the parameter of interest, ρ, and a nuisance parameter, σ_2^2. A ratio of linear combinations of the quadratic forms produces a pivotal quantity for ρ since the nuisance parameter is eliminated in the resulting distribution. Pivotal quantities that can be inverted to obtain confidence intervals for ρ and their associated distributions are

$$\frac{\sum_{m=k+1}^{d} \frac{Q_m}{1+\rho(\Delta_m - 1)} \Big/ \sum_{m=k+1}^{d} r_m}{\sum_{m=1}^{k} \frac{Q_m}{1+\rho(\Delta_m - 1)} \Big/ \sum_{m=1}^{k} r_m} \sim F\left(\sum_{m=k+1}^{d} r_m, \sum_{m=1}^{k} r_m\right) \tag{4}$$

where k ranges from 1 to $d-1$. Note that the quadratic forms in the numerator of (4) correspond to the larger eigenvalues and the quadratic forms in the denominator of (4) correspond to the smaller eigenvalues. Only by grouping the Q's in order of their eigenvalues can one ensure the pivotal quantity is a monotonic function of ρ. In this manner the pivotal quantity in (4) for a given value of k produces a single confidence intervals as opposed to a union of disjoint intervals. See Burch and Iyer (1997) for more details on the relationship between pivotal quantities and confidence intervals for ρ.

For balanced one-way random effects models, $d = 2$ and the value of k is self evident. That is, $k = 1$ and there is one quadratic form in the numerator and one quadratic form in the denominator of (4). For the unbalanced model, the investigator must select k. The natural choice is $k = 1$ since the resulting pivotal quantity partitions $(Q_1, ..., Q_d)$ into Q_1 and $Q_2, ..., Q_d$, which is consistent with the division of the total variation into the "between groups" and "within groups" sources found in the analysis of variance table in Table 1. The confidence interval built from the pivotal quantity having $k = 1$ also corresponds to the Wald (1940) interval. Furthermore, selecting $k = 1$ as opposed to another value of k produces desirable large-sample properties for point estimators of ρ as discussed by Burch and Harris (2001).

Let $F_{\alpha/2}$ and $F_{1-\alpha/2}$ be the $\alpha/2$ and $1-\alpha/2$ percentiles of the F distribution having numerator and denominator degrees of freedom equal to $\sum_{m=2}^{d} r_i = a - 1$ and $r_1 = n - a$, respectively. A 100(1-α)% confidence interval for ρ is given by $(L(\mathbf{Q}), U(\mathbf{Q}))$ where

$$P\Big[F_{\alpha/2} \le \frac{\sum_{m=2}^{d} \frac{Q_m}{1+\rho(\Delta_m - 1)} / \sum_{m=2}^{d} r_m}{\frac{Q_1}{1-\rho}/r_1} \le F_{1-\alpha/2}\Big] = P\Big[L(\mathbf{Q}) \le \rho \le U(\mathbf{Q})\Big] \tag{5}$$

and $\mathbf{Q} = (Q_1, ..., Q_d)$.

When $d = 2$, as is the case for the balanced model (or unbalanced models with $a = 2$) the endpoints of the confidence interval for ρ are available in closed-form. They are

$$L(\mathbf{Q}) = \frac{a(b-1)Q_2 - (a-1)F_{1-\alpha/2}Q_1}{a(b-1)Q_2 + (a-1)(b-1)F_{1-\alpha/2}Q_1} \tag{6}$$

$$U(\mathbf{Q}) = \frac{a(b-1)Q_2 - (a-1)F_{\alpha/2}Q_1}{a(b-1)Q_2 + (a-1)(b-1)F_{\alpha/2}Q_1}. \tag{7}$$

When $d > 2$, however, the endpoints must be obtained via numerical methods. A relatively straight-forward procedure derived by Pratt (1961) can be used to compute the expected length of the confidence interval. From Pratt (1961),

$$E_\rho\Big[U(\mathbf{Q}) - L(\mathbf{Q})\Big] = \int_{\rho^* \ne \rho} P_\rho\Big[L(\mathbf{Q}) \le \rho^* \le U(\mathbf{Q})\Big] d\rho^* \tag{8}$$

where $P_\rho\left[L(\mathbf{Q}) \le \rho^* \le U(\mathbf{Q})\right]$ is the probability of covering false values of the parameter when ρ denotes the true value of the unknown parameter. When $\rho^* = \rho$, it follows that $P_\rho\left[L(\mathbf{Q}) \le \rho \le U(\mathbf{Q})\right] = 1 - \alpha$. Equation (8) illustrates the fact that the expected length of a confidence interval is equal to the integral over false values of the probability of false coverage.

Combining (5) and (8) and noting the parameter space of ρ is $[0, 1)$, we obtain

$$E_\rho\Big[U(\mathbf{Q}) - L(\mathbf{Q})\Big] = \int_0^1 P_\rho\Big[F_{\alpha/2} \le \frac{\sum_{m=2}^{d} \frac{Q_m}{1+\rho^*(\Delta_m - 1)} / \sum_{m=2}^{d} r_m}{\frac{Q_1}{1-\rho^*}/r_1} \le F_{1-\alpha/2}\Big] d\rho^*. \tag{9}$$

The expected lengths of confidence intervals for ρ are essentially integrals of cumulative distribution functions of linear combinations of independent χ^2 random variables. One may note that the computed expected lengths depend on the true value of the parameter. Furthermore, since cumulative distribution functions take the form of integrals, the expected lengths can be obtained by computing the double integrals implicitly given in (9). See Burch and Iyer (1997) for details.

Comparing the expected lengths of confidence intervals for ρ from different one-way random effects designs will serve as a way to ascertain which designs yield the most precise inference about ρ. Designs which result in short intervals indicate more efficient use of the experimental material. Since expected length depends on the value of the parameter, it may be the case that one design is not uniformly better than another design. The dependence on ρ may be eliminated by computing the average or the minimax expected length of the confidence intervals.

2.2 "Approximate" Confidence Intervals for ρ

"Approximate" confidence intervals for ρ are based on the asymptotic distribution of the point estimator of ρ. For simplicity, we use an ANOVA estimator of ρ,

$$\widehat{\rho} \;=\; \frac{r_1 \sum_{m=2}^{d} Q_m - \sum_{m=2}^{d} r_m Q_1}{(\overline{\Delta}-1)\sum_{m=2}^{d} r_m Q_1 + r_1 \sum_{m=2}^{d} Q_m} \tag{10}$$

where

$$\overline{\Delta} \;=\; \frac{\sum_{m=2}^{d} r_m \Delta_m}{\sum_{m=2}^{d} r_m}. \tag{11}$$

Equations (10) and (11) may be written in terms of sample size and group sizes by recognizing that $r_1 = n - a$, $\sum_{m=2}^{d} r_m = a - 1$, and

$$\sum_{m=2}^{d} r_m \Delta_m \;=\; n - \frac{\sum_{i=1}^{a} b_i^2}{n}. \tag{12}$$

See Donna and Koval (1982) for further discussion of the ANOVA estimator. The asymptotic properties of the ANOVA estimator of ρ can be determined using regularity conditions. Burch and Harris (2001) show that

$$\widehat{\rho} \;\overset{asymp}{\sim}\; N\Big(\rho, V(\widehat{\rho})\Big) \tag{13}$$

where $V(\widehat{\rho})$ is the asymptotic variance of $\widehat{\rho}$ given by

$$V(\widehat{\rho}) = \frac{2(1-\rho)^2}{(n-a)(a-1)} \frac{(A\rho^2 + B\rho + C)}{\overline{\Delta}^2} \tag{14}$$

where

$$A = (n-a)Var(\Delta^*) + (n-1)(\overline{\Delta}-1)^2 \tag{15}$$

$$B = 2(n-1)(\overline{\Delta}-1) \tag{16}$$

$$C = n-1 \tag{17}$$

and $Var(\Delta^*) = \sum_{m=2}^{d} r_m(\Delta_m - \overline{\Delta})^2/(a-1)$ is the variance in $\Delta_2, ..., \Delta_d$ (as if the Δ's were random variables themselves). The asymptotic variance of $\widehat{\rho}$ given in (14) is equivalent to that given by Donner and Koval (1982) since

$$\sum_{m=2}^{d} r_m \Delta_m^2 = \frac{\left(\sum_{i=1}^{a} b_i^2\right)^2}{n^2} - 2\frac{\sum_{i=1}^{a} b_i^3}{n} + \sum_{i=1}^{a} b_i^2. \tag{18}$$

For a balanced design, the asymptotic variance of $\widehat{\rho}$ reduces to

$$V(\widehat{\rho}) = \frac{2(n-1)(1-\rho)^2}{(n-a)(a-1)} \frac{\left(1+\rho(b-1)\right)^2}{b^2}. \tag{19}$$

At this point it is worth clarifying what is meant by "asymptotic" when talking about one-way random effects designs. For a balanced design with groups of size b the idea is simple; asymptotic results are generated by looking at the number of groups going to infinity ($a \to \infty$), with all group sizes fixed at b. Any particular design, say 5 groups of 4, is thought of as an element in a sequence of designs (in this case all designs with groups of size 4), where successive members of the sequence have progressively more observations. Note that both the number of groups and total number of observations go to infinity in this sequence. For an unbalanced design the challenge is to conceptually construct a sequence of designs which again has number of groups and number of observations going to infinity, but also contains the unbalanced design, and extends it in a reasonable way. There are various ways to do this, but we consider repetitions of the design. Thus, the design $(2, 3, 4)$ is considered as the first design in the sequence of designs $\{(2, 3, 4), (2, 2, 3, 3, 4, 4), (2, 2, 2, 3, 3, 3, 4, 4, 4), ...\}$. When we refer to the asymptotic variance associated with the design $(2, 3, 4)$, we mean the leading term in the limit of the variances of these designs. Whether or not this asymptotic formula will give a variance which is actually close to the true variance for a particular design is an open question, but the statistician's usual hope is that it will, at least for a large enough design.

A 100(1-α)% asymptotic confidence interval for ρ is given by $\widehat{\rho} \pm Z_{\alpha/2}\sqrt{\widehat{V}(\widehat{\rho})}$ where $\widehat{V}(\widehat{\rho})$ indicates that ρ in (14) is replaced by $\widehat{\rho}$ and $Z_{\alpha/2}$ is associated with the $\alpha/2$ percentile of the standard normal distribution. The expected length of the asymptotic confidence interval for ρ is $2Z_{\alpha/2}\sqrt{V(\widehat{\rho})}$. Comparing expected lengths associated with different designs using this approximate method is easier than comparisons using the exact method discussed in Section 2.1.

It can be shown in some cases that $V(\widehat{\rho})$ for a specific design is uniformly less than $V(\widehat{\rho})$ for a competing design. It follows that the expected length of the asymptotic confidence interval for ρ associated with the first design is uniformly less than the expected interval length associated with the alternative design. Note, however, that the asymptotic results rely on a normally distributed estimator whose distribution is not constrained to the unit interval. While truncation issues are not addressed in the approximate method, they are dealt with in the exact method since the probabilities in (5) are unaffected by truncation. Truncation is particularly common when ρ is small, so we expect poor agreement between the exact and asymptotic methods for small ρ, even for very large sample sizes.

3 Results on Best Designs

3.1 Fixed n (sample size) and a (number of groups)

Suppose that the total sample size as well the the number of groups are fixed. As shown in the theorem below, large-sample results verify that for fixed n and a, the balanced design yields more information about ρ than does any unbalanced design. See the Appendix for the proof of Theorem 1.

Theorem 1. In one-way random effects models where the total number of observations as well as the number of groups are fixed, the balanced design (if one exists) outperforms any unbalanced design when estimating ρ using the asymptotic variance criterion.

Theorem 1 assumes the existence of a balanced design, but this will often not be the case. Intuition suggests that the best design will be the one closest to balanced, that is, if $b = n/a$ is non-integer then the best design should consist of groups of size $[b]$ and $[b] + 1$ where $[b]$ represents the integer part of b. We will call this the straddling design. In a closely related problem concerning the asymptotic variance of the ANOVA estimator of σ_1^2/σ_2^2, Anderson and Crump (1967) showed that this is indeed the best design. The following theorem (Theorem 2) extends Anderson and Crump 's result to the intraclass correlation. The outline of the proof of Theorem 2 given in the Appendix is a modification of Anderson and Crump's result.

Theorem 2. In one-way random effects models where b is noninteger and total number of observations as well as the number of groups is fixed, the best design when estimating ρ using the asymptotic variance criterion is the straddling design.

It is difficult to obtain theoretical results concerning best designs using exact methods as the expected lengths of confidence intervals for ρ are computed using (9). However, we can examine results in small-sample scenarios to see if they agree with the asymptotic results. Consider the example from Vangel (1992) where $n = 25$ and $a = 5$. Table 4 lists a selection of competing designs that satisfy the sampling constraints. The designs in Table 4 are

Table 4: Selected Designs when $n = 25$, $a = 5$

Design	b_1	b_2	b_3	b_4	b_5
1	5	5	5	5	5
2	3	4	5	6	7
3	2	5	6	6	6
4	2	2	2	2	17

compared to one another in terms of the expected length of confidence intervals for ρ using the small-sample calculations. Figure 1 displays the expected lengths of the 90% confidence intervals for ρ where $\alpha/2 = 0.05$. The values of b_i are listed next to each expected length curve. It is interesting to note that the exact results suggest that the balanced design is uniformly better than its competitors, in agreement with the asymptotic calculations.

3.2 Fixed n: Balanced Designs

Suppose the total sample size is fixed and only balanced designs are under consideration. For the balanced case, the asymptotic formula for the expected length of the confidence intervals for ρ is

$$\begin{aligned} E_\rho\left[U(\mathbf{Q}) - L(\mathbf{Q})\right] &= 2Z_{\alpha/2}\sqrt{V(\widehat{\rho})} \\ &= 2Z_{\alpha/2}\left(\sqrt{\frac{2(n-1)}{n(n-b)(b-1)}}(1-\rho)(1+\rho(b-1))\right). \end{aligned} \tag{20}$$

This simple equation is due to the fact that $d = 2$, $Var(\Delta^*) = 0$, and $\overline{\Delta} = b$ for any balanced one-way random effects model. The dependence of the expected length on ρ suggests that there does not exist a single balanced design that has a uniformly minimum expected length across the parameter space. In fact, we can minimize (20) with respect to b and find that the optimal choice is

$$b = \frac{n(1+\rho)+1-\rho}{n\rho+2-\rho}, \tag{21}$$

which, as n goes to infinity, gives $b = 1 + 1/\rho$. This is the same optimal group size recommended by Anderson and Crump (1967) for designs used to estimate σ_1^2/σ_2^2. This asymptotic result suggests that when ρ is small, one should use a few large groups and when ρ is large, one should use many small groups.

If ρ is unknown, or if one is unwilling specify its value, we recommend a group size which has good performance over the entire parameter space. One approach is to find the value of b that minimizes the average expected length over ρ. Brown, Cai, and DasGupta (2001) use this method in an application concerning confidence intervals for proportions. A second approach employs the minimax principle. In this case one selects the value of b that minimizes the maximum asymptotic variance, where the maximum is taken over possible values of ρ. The average expected length of a confidence interval for ρ using the asymptotic formula is

$$\begin{aligned} \int_0^1 E_\rho\left[U(\mathbf{Q}) - L(\mathbf{Q})\right] d\rho &= 2Z_{\alpha/2}\int_0^1 \sqrt{V(\widehat{\rho})}d\rho \\ &= 2Z_{\alpha/2}\left(\frac{1}{6}\sqrt{\frac{2(n-1)}{n}}\frac{b+2}{\sqrt{n-b}\sqrt{b-1}}\right). \end{aligned} \tag{22}$$

This quantity is minimized when

$$b \;=\; \frac{2(2n+1)}{n+5}. \tag{23}$$

If one considers the minimax approach, the maximum value of $\sqrt{V(\widehat{\rho})}$ occurs when $\rho = (b-2)/2(b-1)$, at which point the value of $\sqrt{V(\widehat{\rho})}$ is

$$\frac{1}{4}\sqrt{\frac{2(n-1)}{n}}\frac{b^2}{\sqrt{n-b}(b-1)^{3/2}}. \tag{24}$$

This is minimized with respect to b when

$$b \;=\; \frac{4n}{n+3}. \tag{25}$$

Using either asymptotic criterion, the optimal design is to chose $b = 4$. In other words, four observations per group provides more information about ρ, in an overall sense, than does any other balanced design.

It is interesting to compare the overall recommendation of $b = 4$ with the "best possible" $b = 1 + 1/\rho$ suggested by the asymptotic variance. We will make the comparison using a ratio of square root of asymptotic variances (latter to former), with n going to infinity. The ratio is

$$\frac{2\sqrt{3}\sqrt{\rho}}{1+3\rho}. \tag{26}$$

As ρ goes to zero, this ratio goes to zero, which indicates that at least according to this formula, the overall recommendation could do very poorly. This is not a surprise as the best design according to this formula for very small ρ is two very large groups. For $\rho \geq .083$ the ratio is greater than or equal to 0.8. Anotherwords for the vast majority of ρ we have that the $b = 4$ recommendation is at least "80% efficient", or that the confidence intervals which result will be within 25% of the shortest possible for that ρ, when measured by expected length. So for the vast majority of ρ, $b = 4$ is a good overall recommendation asymptotically. As discussed at the end of Section 2.2, agreement between the asymptotic and exact results are poor when ρ is small. Exact calculations actually suggest that a design having $b = 4$ is a viable contender as ρ goes to zero.

Suppose the experimenter wants to use a balanced design with a group size other than 4. Asymptotically, how much will be lost? The comparison can be done in three ways. One is to compare the average $\sqrt{V(\hat{\rho})}$ for the case $b = 4$ and the candidate value of b, by using a ratio, and then take the limit as the sample size goes to infinity. Using this comparison we find the ratio for a candidate b is

$$\frac{\sqrt{3}(b+2)}{6\sqrt{b-1}}.$$

Of course, if $b = 4$ this ratio is one. A second comparison is to take the ratio of the maximized $\sqrt{V(\hat{\rho})}$ (maximized over ρ), and again take the limit as n goes to infinity. Using this comparison we find the ratio for a candidate b is

$$\frac{3^{3/2}b^2}{4^2(b-1)^{3/2}}.$$

Again, if $b = 4$ this ratio is one. The third and final comparison is to take the ratio of $\sqrt{V(\hat{\rho})}$ for the cases $b = 4$ and a general b, and then maximize it with respect to ρ. This will produce the least favorable comparison of a general b to $b = 4$. Again, we take the limit as n goes to infinity. This maxmized ratio is

$$\frac{\sqrt{3}}{\sqrt{b-1}}$$

for $b < 4$ and

$$\frac{\sqrt{3}b}{4\sqrt{b-1}}$$

for $b > 4$. Table 5 compares $b = 2, 3, 4, 5, 7$ to $b = 4$ using these three criteria.

Table 5: Asymptotic comparison of balanced design with group size b to balanced design with group size 4

group size b	ratio of average $\sqrt{V}$	ratio of max $\sqrt{V}$	max of ratio of $\sqrt{V}$
2	1.16	1.30	1.73
3	1.02	1.03	1.22
4	1	1	1
5	1.01	1.02	1.08
7	1.06	1.08	1.24

An example of how to interpret a table entry is as follows; for $b = 2$ and the max ratio crtierion the table entry of 1.73 means that asymptotically using $b = 2$ can give intervals whose expected length is as much as 73% wider than the intervals from the “best” b of 4. We can see from the table that asymptotically a group size of 5 is an excellent alternative to groups of size 4. Groups of size 3 or 7 are less suitable, although exact calculations suggest that groups of size 3 are suitable in small designs, as we will see later. Asymptotically, it makes little difference whether you use groups of size 4 or 5.

Of course, it will not be possible to choose a balanced design with 4 observations per group if the total sample size n is not divisible by 4. Consider the case in which the sample size is 105. Possible balanced designs include those having $b = 3$, 5, 7, 15, 21, and 35. We would like to determine which of these competing balanced designs is best for estimating ρ. Figure 2 displays the optimal values of b for balanced designs based on minimizing (22) for sample sizes ranging from 100 to 200. Sample sizes that are prime numbers are not included. For $n = 105$, the balanced design having $b = 5$ provides more information about ρ than does any other balanced design. Designs based on sample sizes that are divisible by 3, 4 or 5 are better than designs of slightly larger size that are not divisible by 3, 4, or 5. Identical results are obtained if one were to use the asymptotic minimax expected length criterion.

Figure 2 suggests that in some cases an investigator should redesign an experiment to obtain optimal results. For example, suppose total sample sizes of 114 or 115 are under consideration. The best balanced designs for these sample sizes have $b = 3$ and $b = 5$, respectively. As previously mentioned, balanced designs having group sizes of 3 and 5 offer respectable results. If the investigator, however, simply reduced the sample size to 112 with $b = 4$, he would obtain a better design using fewer expermimental units. This is also the case using the asymptotic minimax expected length as a criterion.

In general, one may consider reducing the sample size in order to compare a balanced design with $b = 4$ to the original balanced design which is slightly large in size. Consider the case in which the reduced design with $b = 4$ is compared to a balanced design with $b = 5$ where the sample size is reduced by 3 to get divisibility by 4. For example, $n = 115$ is reduced to $n = 112$ so that $b = 4$ may be used. Reducing the sample size by 3 is the largest possible reduction in order to obtain $b = 4$. The reduced and original designs may be compared using the average expected length or the maximum expected length asymptotic criteria. For average $\sqrt{V(\widehat{\rho})}$, substituting $b = 5$, $n = n$ and $b = 4$, $n = n - 3$ into (22) shows that the reduced design outperforms the original design when $n \geq 102$. Using the same approach for the maximum value of $\sqrt{V(\widehat{\rho})}$ given in (24), the reduced design outperforms the original design when $n \geq 73$.

(Note that

$$\sqrt{\frac{n-4}{n-3}}\frac{6}{\sqrt{3(n-7)}} < \sqrt{\frac{n-1}{n}}\frac{7}{\sqrt{4(n-5)}}. \tag{27}$$

After some algebra we can show this occurs when

$$n^3 - 107n^2 + 559n - 1029 > 0.$$

which is true for $n \geq 102$.)

(We have that the reduced design is better when

$$\frac{(n-1)5^4}{4^3 n(n-5)} > \frac{4^4(n-4)}{3^3(n-3)(n-7)}.$$

This occurs when

$$491n^3 - 38169n^2 + 195445n - 354375 > 0,$$

or when $n \geq 73$.)

Using asymptotic results to measure the quality of designs is appropriate only when sample sizes are large. To compare designs when sample sizes are small, we now turn our attention to exact calculations presented in Section 2.1. Figure 3 displays the optimal values of b for balanced designs based on minimizing (9) for sample sizes ranging from 10 to 100. Sample sizes that are prime numbers are not included. Similar results are obtained if one were to use the minimum maximum expected length criterion.

Consider the case in which the sample size is 48. Possible balanced designs include those having $b = 2$, 3, 4, 6, 8, 16, and 24. We would like to determine which of these competing balanced designs is best for estimating ρ. From Figure 3, the balanced design having $b = 3$ provides more information about ρ than does any other balanced design. For medium sample sizes, balanced designs having $b = 3$ are superior to those having $b = 4$. For small sample sizes, $b = 2$ is the optimal choice of group size. Note that balanced designs having $b = 5$ are inferior to the slightly smaller balanced designs having $b = 3$ or 4.

The actual choice of $b = 2$, 3, or 4 depends on what is meant by small,medium, and large sample sizes as well as the criterion used to measure the quality of the design. Figure 4 displays the optimal balanced design using the exact calculations of the minimum expected length and minimax expected length. For those samples in which group sizes of 2, 3, and 4 are possible, small samples are defined as $n \leq 12$. In this case, use $b = 2$ for those balanced designs in which group sizes of 2 or 3 are options. Medium sample sizes are defined as $12 < n \leq 144$ or $12 < n \leq 60$, depending on the criterion used to judge the design. In this case use $b = 3$ for those balanced designs in which group sizes of 3 or 4 are options. For large sample sizes, the best balanced design has $b = 4$. When the objective of the study is to estimate ρ based on confidence intervals, balanced designs having $b = 5$ do not provide additional information about the parameter. We recommend the investigator simply select a slightly smaller sample size that is divisible by 3 or 4.

3.3 Fixed n: Determine Best Design

In this Section we consider balanced as well as unbalanced designs. For instance, instead of throwing away information in order to obtain a good balanced design, we consider one-way random effects designs where the groups may have an unequal number of measurements. For a fixed value n, we seek the best overall design for estimating ρ. It is not always the case that the best design is a balanced design. Results using both asymptotic and exact calculations are presented.

For a fixed sample size, there does not exist one design that is uniformly better than another design since the asymptotic variance of $\widehat{\rho}$ given by (14) depends on the value of ρ. The average expected length of a confidence interval using asymptotic theory is

$$\begin{aligned}
\int_0^1 E_\rho \left[U(\mathbf{Q}) - L(\mathbf{Q})\right] d\rho &= 2Z_{\alpha/2} \int_0^1 \sqrt{V(\widehat{\rho})} d\rho \\
&= 2Z_{\alpha/2} \sqrt{\frac{2}{(n-a)(a-1)}} \frac{1}{\overline{\Delta}} \left[\frac{(2A+B)\sqrt{A+B+C} - B\sqrt{C}}{4A} \left(1 + \frac{B}{2A}\right) \right. \\
&\quad - \frac{1}{3A} \left((A+B+C)^{3/2} - C^{3/2} \right) \\
&\quad \left. + \left(\frac{4AC - B^2}{8A} \right) \left(1 + \frac{B}{2A}\right) \frac{1}{\sqrt{A}} ln \left(\frac{2\sqrt{A}\sqrt{A+B+C} + 2A + B}{2\sqrt{AC} + B} \right) \right] \qquad (28)
\end{aligned}$$

where A, B, and C are given in Section 2.2.

Theoretical results based on the asymptotic approach are hard to come by. A search of all possible designs numerically (for a given n) using (28) reveals a consistent pattern for $n \geq 36$. The best designs follow a pattern when n is written as $4k - r$, where r is either 0, 1, 2, or 3. These designs consists of $k - r$ groups of size 4, and r groups of size 3. The same pattern appears if one uses the asymptotic minimum maximum expected length criterion. For example, $114 = 4(29) - 2$, so we should have $29 - 2 = 27$ groups of size 4, and 2 groups of size 3. A heuristic way to state this rule is to select a design with as many groups of size 4 as possible, with the proviso that any remaining groups must be of size 3. Note that $n = 114$ could be distributed as 28 groups of size 4 and one group of size 2, which would maximize the number of groups of size 4, but this design violates the rule that remaining groups be of size 3. The unbalanced design of 27 groups of size 4 and 2 groups of size 3 outperforms the best balanced design, which is 38 groups of size 3.

For large n, these unbalanced designs also outperform the reduced balanced designs having $b = 4$. In practice, however, the quality and simplicity of the balanced design may persuade the user to forgo the small gain in performance offered by the unbalanced design. This is apparent if one compares the reduced balanced design to the mathematically best design, which is unattainable, being a balanced design based on n measurements with a fractional number of observations per group. Using the average expected asymptotic confidence interval width given in (22), the worst scenario for the reduced design occurs when n is reduced by 3. In this case the comparison is between the reduced balanced design having $n = n - 3, b = 4$ and the mathematically best design having $n = n, b = 2(2n + 1)/(n + 5)$. Using (22), the ratio of the former to the latter is

$$\sqrt{\frac{n(n-1)(n-4)}{(n-3)(n-7)(n+2)}}. \tag{29}$$

This is monotonic in n and for $n \geq 40$ is less than 1.05. This means that for $n \geq 40$, the reduced balanced design having $b = 4$ will be within a few percent of the best possible design when performance is measured by the average expected asymptotic confidence interval width.

One may also compare these designs using the maximum value of $\sqrt{V(\widehat{\rho})}$ given in (24). In this case, the comparison is between the reduced balanced design having $n = n - 3, b = 4$ and the mathematically best design having $n = n, b = 4n/(n + 3)$. Using (24), the ratio of the former to latter is

$$\frac{(n-1)^{3/2}\sqrt{n-4}}{n\sqrt{n-3}\sqrt{n-7}}. \tag{30}$$

This is also monotonic in n and for $n \geq 37$ is less than 1.05, indicating that the reduced balanced design having $b = 4$ does not give up an appreciable amount of information when estimating ρ.

A final way to compare the reduced design to the unattainable theoretical best is to take the ratio of $\sqrt{V}$ and maximize with respect to ρ. If we do this we obtain

$$\sqrt{\frac{(n-4)(n-1)}{(n-3)(n-7)}}.$$

This is also monotonic in n and for $n \geq 56$ is less than 1.05. All three of these comparisons show that although the best asymptotic unbalanced design is fairly simple to construct, a practitioner will lose little if he or she reduces the sample size to obtain a balanced design with $b = 4$.

As was the case using the asymptotic approach, theoretical results based on exact calculations are not easy to produce. A search of all possible designs numerically (for a given n) using (9) reveals a set of patterns depending on the sample size. The following rule of thumb selects a design which is usually the best possible (although not always, see the example discussed in the next paragraph). The rule is to select a design as follows: If $n < 18$, use as many groups of size 2 as possible, with the proviso that any remaining groups must be of size 3; if $18 \leq n < 36$, use as many groups of size 3 as possible, with the proviso that any remaining groups must be of size 2; if $36 \leq n < 108$, use as many groups of size 3 as possible, with the proviso that any remaining groups must be of size 4; and if $n \geq 108$, use as many groups of size 4 as possible, with the proviso that any remaining groups must be of size 3. Similar results are obtained using the exact minimax expected length calculations. When using the minimax criterion, the cutoff value for recommending using as many groups of size 4 as possible, with the proviso that any remaining groups must be of size 3, is reduced from 108 to 64.

The general strategy used to select designs is based on the result that groups should be primarily of sizes 2, 3, or 4 depending on the sample size. This rule selects designs that are very good but not necessarily the single best design. For example, while the rule selects 6 groups of size 3 for $n = 18$, both the minimum average and minimax expected length calculations select a design having 3 groups of size 2 and 4 groups of size 3. The ratio of the minimum average expected length of the first design to the second is 1.006. Similarly, the ratio of the minimax expected length of the first design to the second is 1.003. The rule simply provides an easy way to select quality designs in order to estimate ρ.

For intermediate sample sizes, the design selected using the asymptotic approach is different from the design selected using the exact approach. The asymptotic approach sets out to create group sizes of 4 whereas the priority of the exact approach is to create groups of size 3. To compare designs recommended by the two approaches, consider the percent relative difference of the average expected length (or minimax expected length) of the two designs. Specifically,

$$100\frac{\{\int_0^1 E_\rho\left[U(\mathbf{Q})-L(\mathbf{Q})\right]d\rho\}_{Design2}-\{\int_0^1 E_\rho\left[U(\mathbf{Q})-L(\mathbf{Q})\right]d\rho\}_{Design1}}{\{\int_0^1 E_\rho\left[U(\mathbf{Q})-L(\mathbf{Q})\right]d\rho\}_{Design1}} \tag{31}$$

where $E_\rho\left[U(\mathbf{Q})-L(\mathbf{Q})\right]$ is computed using (9), Design 2 is the best design using the asymptotic method and Design 1 is the best design using the exact method. Figure 5 displays the value of (31) for 90% confidence intervals as a function of sample size for $n \geq 36$. Figure 5 suggests that the performance of the design selected using the asymptotic approach is similar to the peformance of the design selected using the exact approach. That is, designs predominantly made up of groups of size 4 with remaining groups of size 3 are similiar in quality to designs predominantly made up of groups of size 3 with remaining groups of size 4. Specifically, the asymptotic design is within 2% of the best exact one for $n > 40$ and within 1% of the best exact one for $n \geq 75$.

We now return to the example presented in Vangel (1992) and consider unbalanced designs which outperform the balanced design having $a = 5$ and $b = 5$. Figure 6 displays the expected length of the 90% confidence intervals for ρ where $\alpha/2 = 0.05$ for selected designs. The values of b_i are listed next to each expected length curve. Straddling designs having group sizes of 2 and 3 or 3 and 4 uniformily outperform the balanced design. It is interesting to note that the design (3, 3, 3, 3, 3, 3, 3, 4) uniformly outperforms the design (3, 3, 3, 4, 4, 4, 4). Although the design (2, 2, 3, 3, 3, 3, 3, 3, 3) does not uniformily outperform the other unbalanced designs, it has the smallest average expected length as well as the minimax expected length. Using (31) to compare designs, the percent relative difference of the average expected length for design (3, 3, 3, 4, 4, 4, 4) compared to design (2, 2, 3, 3, 3, 3, 3, 3, 3) is 2.12%. Likewise, the percent relative difference of the average expected length for design (3, 3, 3, 3, 3, 3, 3, 4) compared to design (2, 2, 3, 3, 3, 3, 3, 3, 3) is 0.44%. Note that the actual design used, the (5,5,5,5,5) design, is 9.40% worse than the best possible design based on average expected length.

4 Conclusions and Discussion

We have presented rules for determining optimal designs based on expected length of confidence intervals for the intraclass correlation coefficient ρ. For simplicity, we base our recommendations primarily on theoretical results obtained from the asymptotic normal approximation to the ANOVA estimator of ρ. We have also performed extensive numerical exact calculations of expected lengths for various sample sizes to bolster these conclusions. We find that if both the number of observations (n) and the number of groups (a) are fixed then one should choose a balanced design, or the design closest to balanced. These recommendations are consistent with those of Anderson and Crump (1967) who considered estimation of the variance components themselves, or their ratio. If only the total sample size is fixed, then the optimal allocation of experimental units in a one-way random effects model depends on ρ. In the absence of knowledge of ρ, we suggest two methods of removing ρ from the decision; averaging the expected length over ρ, and minimizing the expected length over ρ. Using asymptotic results, both methods suggest a group size of 4, or slightly less than 4, if $n < 100$.

In general, n will not be divisible by 4, so a balanced design with groups of size 4 is not possible. In these cases, we can with little penalty discard experimental units to get the desired balanced design, or alternatively pick an unbalanced design consisting of groups of size 3 and groups of size 4 as outlined in Section 3.3.

Roughly speaking, the exact small-sample calculations we have done agree with the asymptotic recommendations for $n > 100$. For smaller n, the best designs typically have slightly fewer observations per group than the asymptotic formula suggests. For instance, the best balanced designs have group sizes of 2, 3, or 4 depending on the sample size. However, for intermediate to large sample samples, say $n \geq 36$, even here one will typically not go seriously wrong following the asymptotic recommendation of group sizes of 4 whenever possible.

For a given sample size, the recommendations in this article are easy for the practitioner to implement and appear to work well for all values of ρ. Not addressed are issues related to different costs associated with sampling and subsampling. In addition, we have not computed the sample size required to obtain a prespecified value of expected length. This subject is important when planning a study and selecting a sample size that yields quality information about ρ. These are topics that need to be explored in the future.

APPENDIX

EIGENVALUES FOR UNBALANCED DESIGN HAVING TWO DISTINCT GROUP SIZES

We begin with the fact that if a group of size x is repeated c times, the corresponding eigenvalue is x having replication $c-1$. Similarly, if a group of size y $(x < y)$ is repeated $a-c$ times, the corresponding eigenvalue is y having replication $a-c-1$. Recall that $\Delta_1 = 0$, $r_1 = n-a$, and $cx + (a-c)y = n$. Then

$$\sum_{m=1}^{d} r_m = (n-a) + (c-1) + (a-c-1) + \text{remaining } \ r_m\text{'s} \tag{32}$$

$$= n-1 \tag{33}$$

which implies there is one additional eigenvalue having a replication of one. Call this eigenvalue z. From (12), it follows that

$$(c-1)x + (a-c-1)y + z = cx + (a-c)y - \frac{\sum_i b_i^2}{n} \tag{34}$$

and thus

$$\sum_i b_i^2 = n(x+y-z). \tag{35}$$

However,

$$\sum_i b_i^2 = cx^2 + (a-c)y^2. \tag{36}$$

Equating (35) and (36) yields $z = axy/n$. Note that $x < z < y$.

PROOF of THEOREM 1

Let $V_{BAL}(\widehat{\rho})$ and $V_{UNB}(\widehat{\rho})$ denote the asymptotic variances for the balanced design and any unbalanced design, respectively. Also, $b = n/a$ is the number of observations per group in the balanced design. Then

$$V_{UNB}(\widehat{\rho}) - V_{BAL}(\widehat{\rho}) = \frac{2(1-\rho)^2}{(n-a)(a-1)}(A\rho^2 + B\rho + C) \tag{37}$$

where

$$A = \frac{(n-a)Var(\Delta^*) + (n-1)(\overline{\Delta}-1)^2}{\overline{\Delta}^2} - \frac{(n-1)(b-1)^2}{b^2} \tag{38}$$

$$B = 2(n-1)\left(\frac{\overline{\Delta}-1}{\overline{\Delta}^2} - \frac{b-1}{b^2}\right) \tag{39}$$

$$C = (n-1)\left(\frac{1}{\overline{\Delta}^2} - \frac{1}{b^2}\right) \tag{40}$$

and $Var(\Delta^*)$ and $\overline{\Delta}$ refer to quantities in the unbalanced design. We only need show that $A\rho^2 + B\rho + C > 0$ for all ρ in $[0, 1)$ in order to satisfy the theorem. Note that $\overline{\Delta} < b$. This was shown by Anderson and Crump (1967) and follows from the fact that

$$b - \overline{\Delta} = \frac{\sum_{i=1}^{a}(b_i - b)^2}{n(a-1)} \tag{41}$$

using (11) and (12). $b - \overline{\Delta}$ is related to the variance of the group sizes, which is always positive for unbalanced designs. From $\overline{\Delta} < b$ it follows that $C > 0$ and that $B + 2C > 0$. Furthermore, $A + B + C = (n-a)Var(\Delta^*)/\overline{\Delta}^2 > 0$. These facts are sufficient to obtain our result since $A\rho^2 + B\rho + C = (A+B+C)\rho^2 + (B+2C)(\rho-\rho^2) + C(1-\rho)^2 > 0$ for ρ in $[0, 1)$. It follows that $V_{UNB}(\hat{\rho}) - V_{BAL}(\hat{\rho}) > 0$. Hence the balanced design is uniformly better than any unbalanced design using the asymptotic variance.

PROOF of THEOREM 2

If $V_1(\hat{\rho})$ and $V_2(\hat{\rho})$ denote the asymptotic variances for any two designs, then

$$V_1(\hat{\rho}) - V_2(\hat{\rho}) = \frac{2(1-\rho)^2}{(n-a)(a-1)}(A\rho^2 + B\rho + C) \tag{42}$$

where

$$A = \frac{(n-a)Var_1(\Delta^*) + (n-1)(\overline{\Delta}_1-1)^2}{\overline{\Delta}_1^2} - \frac{(n-a)Var_2(\Delta^*) + (n-1)(\overline{\Delta}_2-1)^2}{\overline{\Delta}_2^2} \tag{43}$$

$$B = 2(n-1)\left(\frac{\overline{\Delta}_1-1}{\overline{\Delta}_1^2} - \frac{\overline{\Delta}_2-1}{\overline{\Delta}_2^2}\right) \tag{44}$$

$$C = (n-1)\left(\frac{1}{\overline{\Delta}_1^2} - \frac{1}{\overline{\Delta}_2^2}\right) \tag{45}$$

where $Var_1(\Delta^*), \overline{\Delta}_1$ and $Var_2(\Delta^*), \overline{\Delta}_2$ refer to quantities associated with designs 1 and 2, respectively. Anderson and Crump (1967) showed that the straddling design maximizes $\overline{\Delta}$, so if the straddling design is design 2, we have that $\overline{\Delta}_2 > \overline{\Delta}_1$. It follows that $C > 0$ and $B + 2C > 0$. Furthermore,

$$A + B + C = (n - a)\left(\frac{Var_1(\Delta^*)}{\overline{\Delta}_1^{\,2}} - \frac{Var_2(\Delta^*)}{\overline{\Delta}_2^{\,2}}\right). \quad (46)$$

If $A + B + C > 0$, then the quadratic $A\rho^2 + B\rho + C > 0$ for all ρ in $[0, 1)$ and hence the result follows. Now

$$\frac{Var(\Delta^*)}{\overline{\Delta}^2} = \frac{\sum_m r_m \Delta_m^2}{(a-1)\overline{\Delta}^2} - 1 \quad (47)$$

so $Var(\Delta^*)/\overline{\Delta}^2$ is minimum when $\sum_m r_m \Delta_m^2/\overline{\Delta}^2$ is minimum. A small amount of algebra shows that $\sum_m r_m \Delta_m^2/\overline{\Delta}^2$ is proportional to

$$\frac{S_2^2 - 2nS_3 + n^2 S_2}{(n^2 - S_2)^2} \quad (48)$$

where $S_2 = \sum_i b_i^2$ and $S_3 = \sum_i b_i^3$. Anderson and Crump (1967) in a three page proof showed that the staddling design minimizes this quantity, and hence applying this result we have that $A + B + C > 0$. Thus, in the absence of a balanced design, the straddling design is uniformly best in terms of asymptotic variance.

5 References

Airy, G. B. (1861), *On the Algebraical and Numerical Theory of Errors of Observations and the Combinations of Observations*, London: MacMillan.

Anderson, R. L., and Crump, P. P. (1967), "Comparisons of designs and estimation procedures for estimating parameters in a two-stage nested process," *Technometrics*, 9, 499–516.

Bessel, F. W. (1838), "The parallax of 61 Cygni," *Monthly Notices of the Royal Astronomical Society.* Reprinted in *Source Book in Astronomy* (1929), eds. H. Shapley and H. E. Howarth, New York: McGraw-Hill.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001), "Interval estimation for a binomial proportion," *Statistical Science*, 16, 101-133.

Burch, B. D., and Harris, I. R. (2001), "Closed-form approximations to the REML estimator of a variance ratio (or heritability) in a mixed linear model," *Biometrics*, 57, 1148–1156.

Burch, B. D., and Iyer, H. K. (1997), "Exact confidence intervals for a variance ratio (or heritability) in a mixed linear model," *Biometrics*, 53, 1318–1333.

Chauvenet, W. (1863a), *A Manual of Spherical and Practical Astronomy, 1: Spherical Astronomy*, Philadelphia: Lippincott, pp. 363–367.

Chauvenet, W. (1863b), *A Manual of Spherical and Practical Astronomy, 2: Theory and Use of Astronomical Instruments*, Philadelphia: Lippincott, pp. 194-197.

Donner, A., and Koval, J. J. (1982), "Design considerations in the estimation of intraclass correlation," *Annals of Human Genetics*, 46, 271–277.

Gibbons, R. D., and Bhaumik, D. K. (2001), "Weighted random-effects regression models with application to interlaboratory calibration," *Technometrics*, 43, 192–198.

Hammersley, J. M. (1949), "The unbiased estimate and standard error of the interclass variance," *Metron*, 15, 189–205.

Koziel, K. (1985), "Bessel and the librations of the moon," *Astrophysics and Space Science*, 110, 193-196.

LaMotte, L. R. (1976), "Invariant quadratic estimators in the random, one-way ANOVA model," *Biometrics*, 32, 793–804.

Lohr, S. L. (1995), "Optimal Bayesian design of experiments for the one-way random effects model," *Biometrika*, 82, 175–186.

Pratt, J. W. (1961), "Length of confidence intervals," *Journal of the American Statistical*

Association, 56, 549–567.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: John Wiley.

Scheffé, H. (1956), "Alternative models for the analysis of variance," *The Annals of Mathematical Statistics*, 27, 251–271.

Shoukri, M. M., and Ward, R. H. (1984), "On the estimation of the intraclass correlation," *Communications in Statistics - Theory and Methods*, 13, 1239–1255.

Tippett, L. H. C. (1931), *The Methods of Statistics* (1st ed.), London: Williams and Norgate.

Vangel, M. G. (1992), "New methods for one-sided tolerance limits for a one-way balanced random-effects ANOVA model," *Technometrics*, 34, 176–185.

Wald, A. (1940), "A note on the analysis of variance with unequal class frequencies," *The Annals of Mathematical Statistics*, 11, 96–100.

Walter, S. D., Eliasziw, M., and Donner, A. (1998), "Sample size and optimal designs for reliability studies," *Statistics in Medicine*, 17, 101-110.

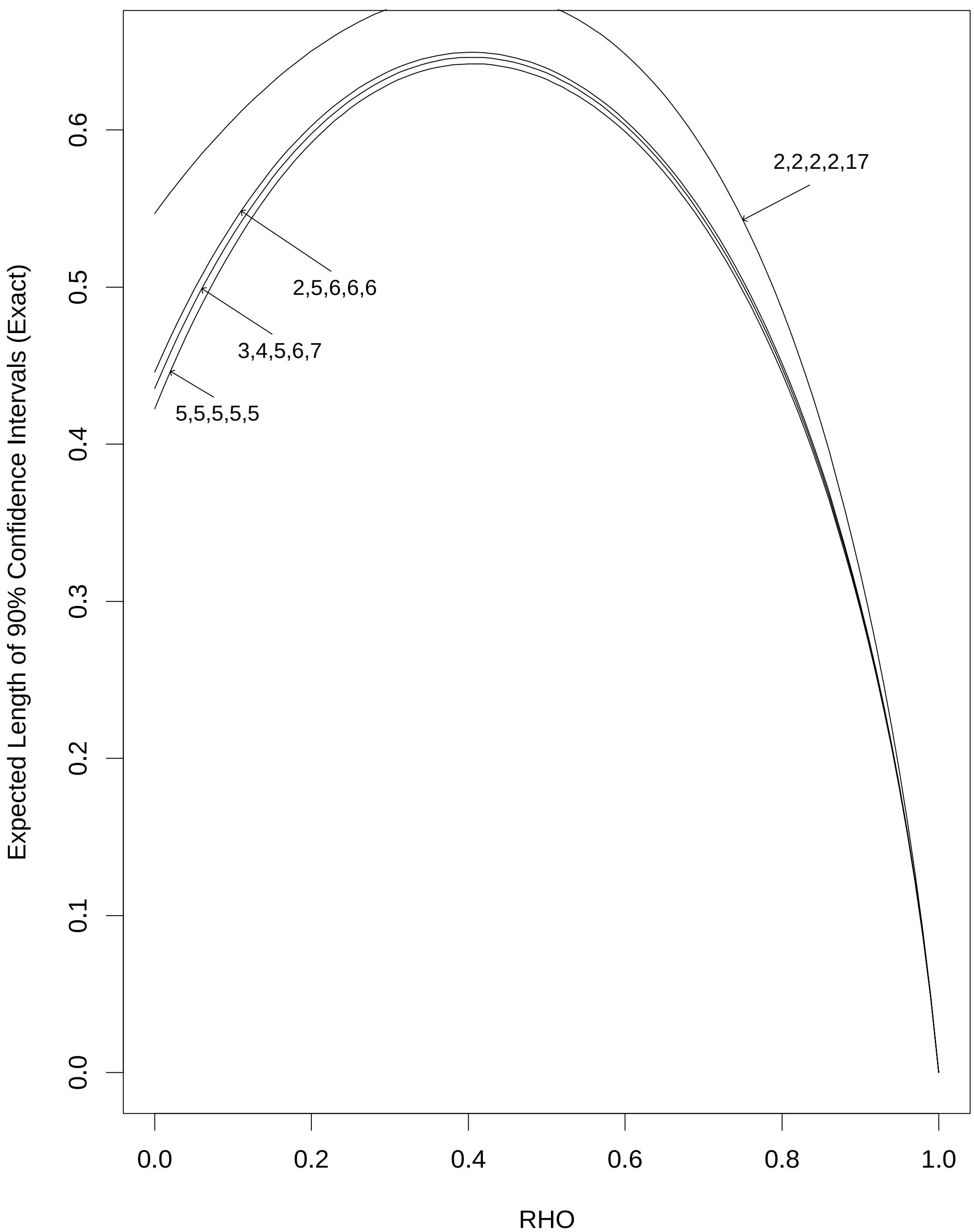

Figure 1: Comparing Balanced and Unbalanced Designs when $n = 25$, $a = 4$

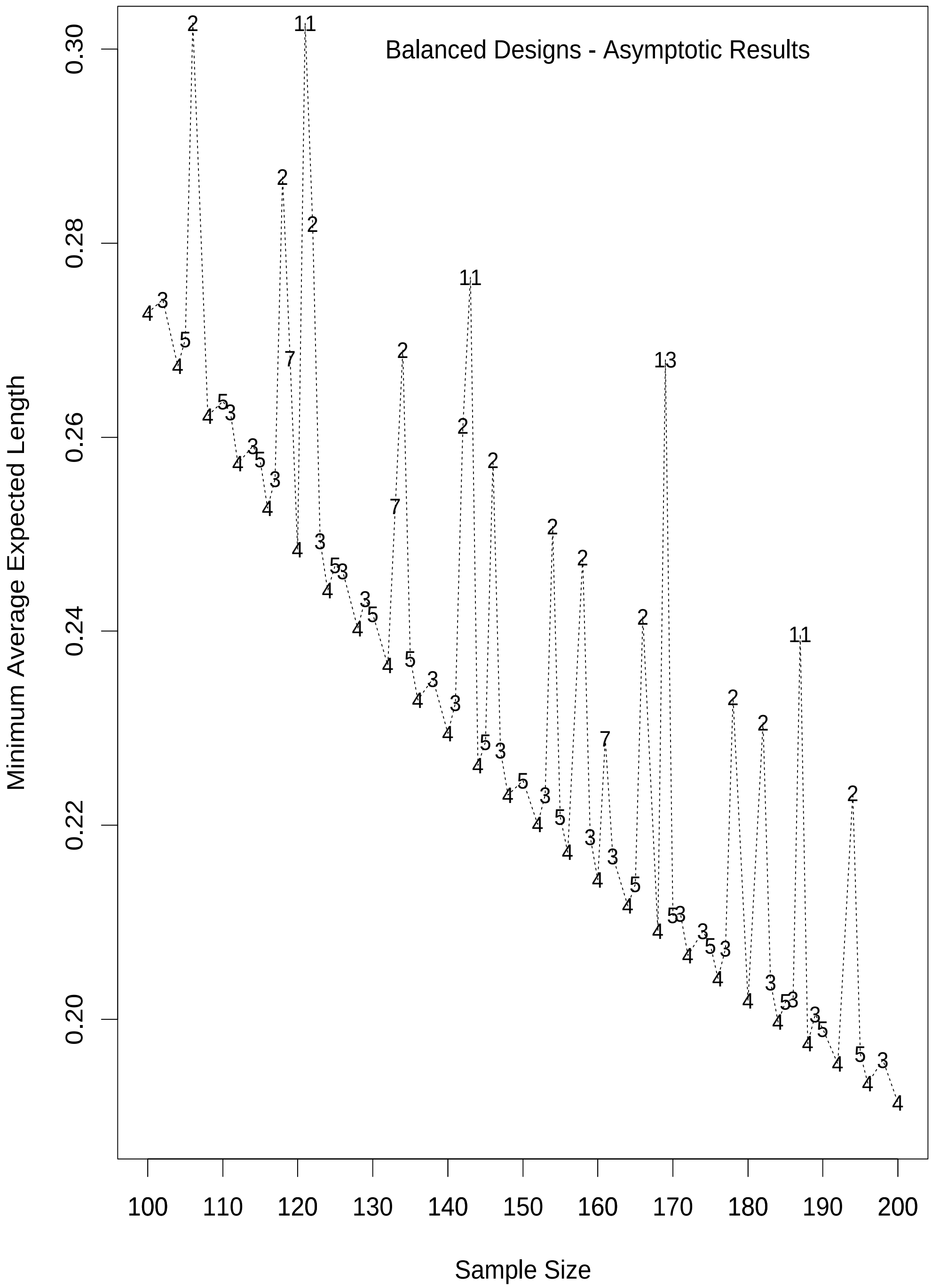

Figure 2: Optimal Group Sizes for Balanced Designs - Asymptotic Results

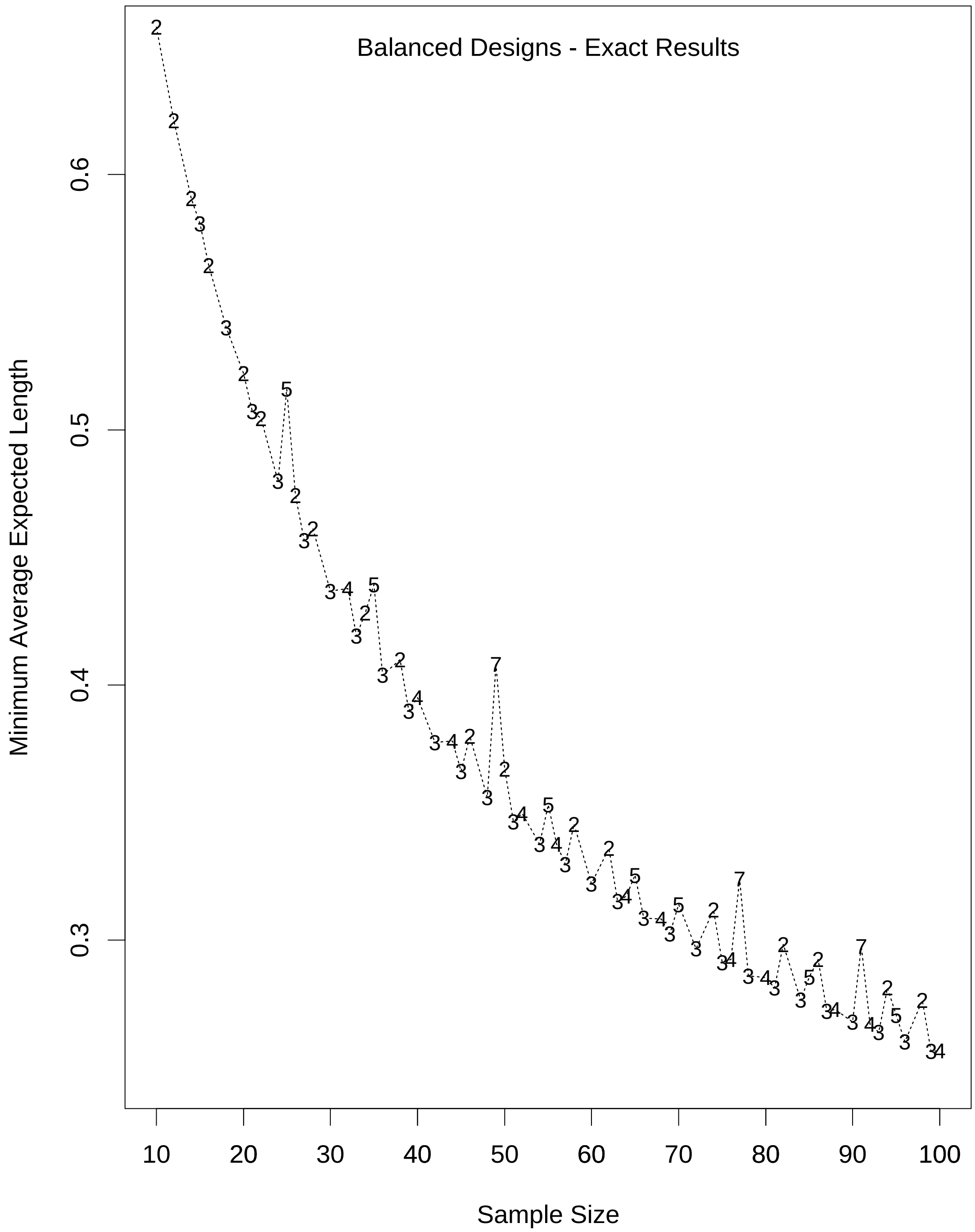

Figure 3: Optimal Group Sizes for Balanced Designs - Exact Results

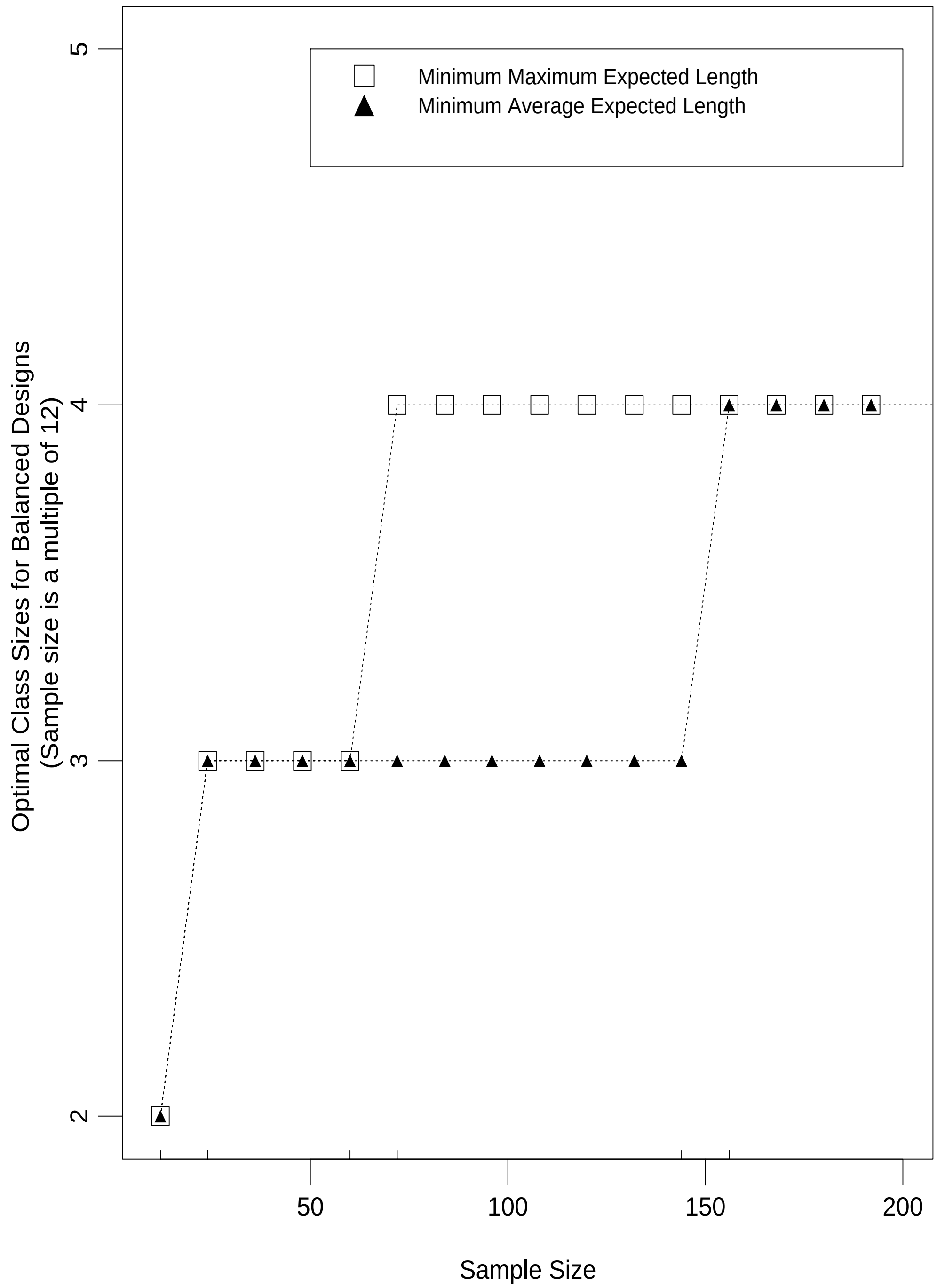

Figure 4: Optimal Group Sizes when $b = 2$, 3, or 4

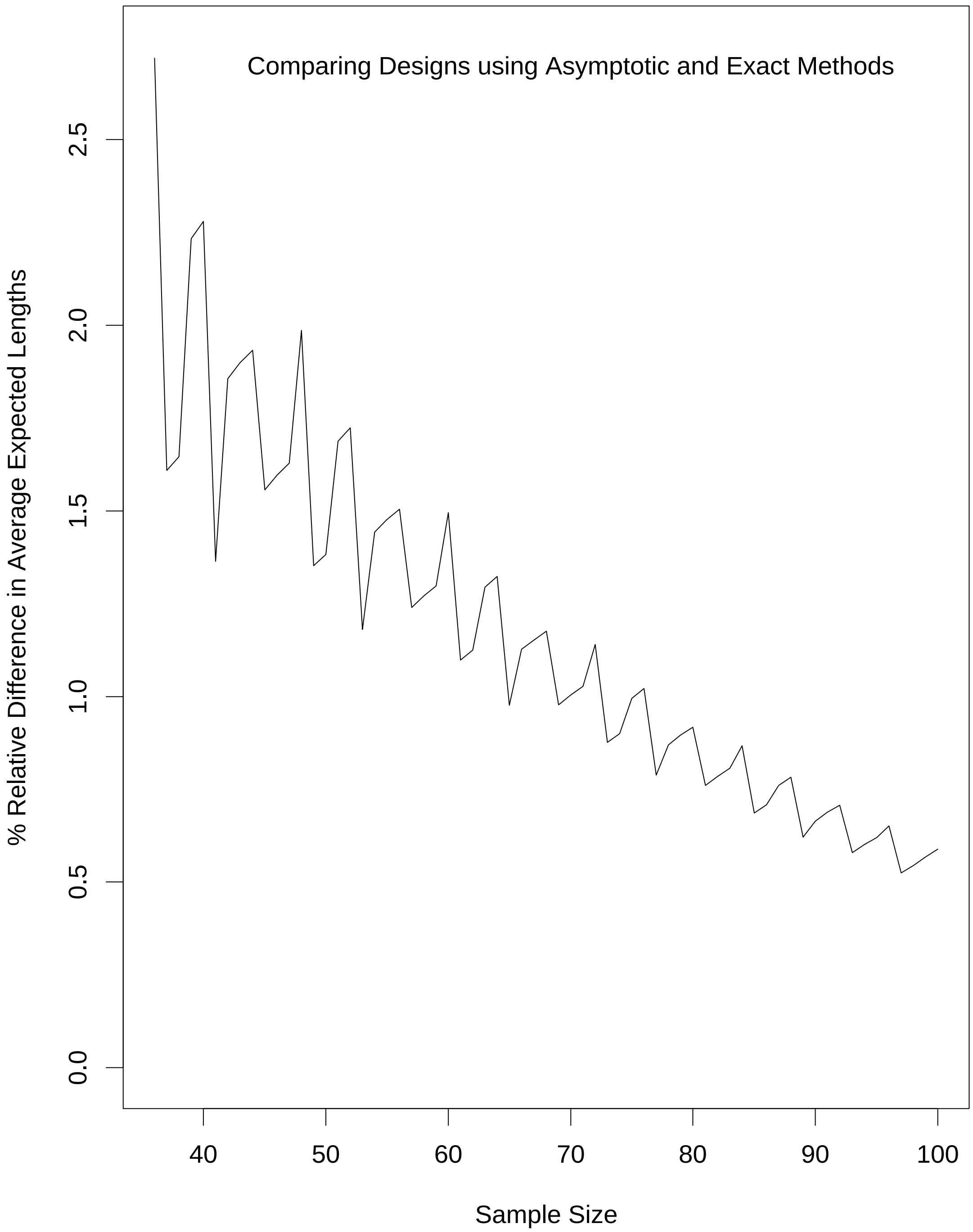

Figure 5: Comparing Designs using Asymptotic and Exact Methods

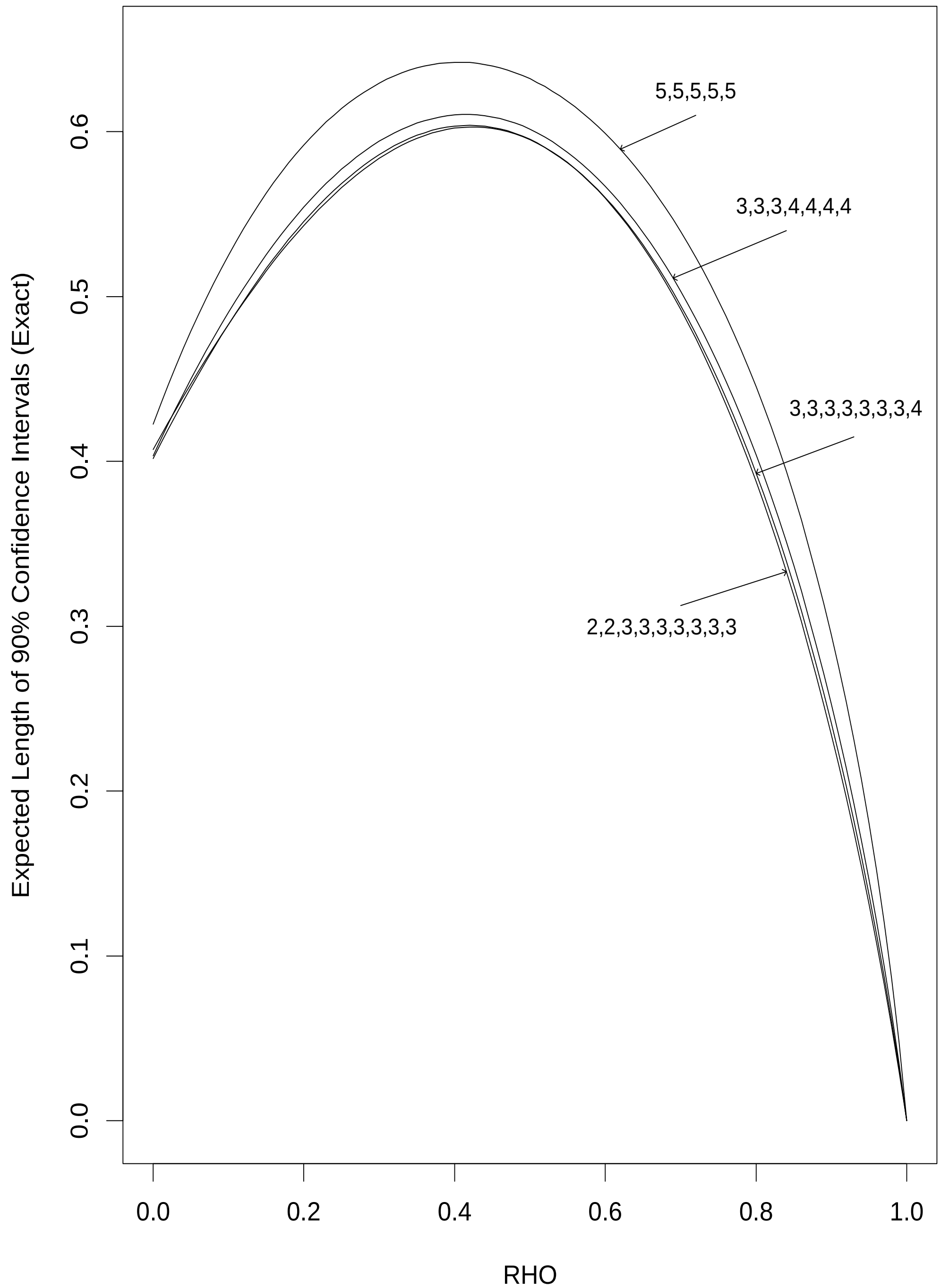

Figure 6: Comparing Balanced and Unbalanced Designs when $n = 25$