Journal of Educational and Behavioral Statistics
Fall 2000, Vol. 25, No. 3, pp. 271–284

Design Issues for Experiments in Multilevel Populations

Mirjam Moerbeek
Gerard J. P. van Breukelen
Martijn P. F. Berger
Maastricht University

Keywords: *multilevel models, experimental designs, level of randomization, allocation of units, power*

For the design of experiments in multilevel populations the following questions may arise: What is the optimal level of randomization? Given a certain budget for sampling and measuring, what is the optimal allocation of units? What is the required budget for obtaining a certain power on the test of no treatment effect? In this article these questions will be dealt with for populations with two or three levels of nesting and continuous outcomes. Multilevel models are used to model the relationship between experimental condition and the outcome variable. The estimator of the regression coefficient associated with treatment condition, a parameter assumed to be fixed in this paper, is of main interest and should be estimated as efficiently as possible. Therefore, its variance is used as a criterion for optimizing the level of randomization and the allocation of units.

In many sciences, experiments are carried out to assess the effect of different treatments on outcome variables of individuals. In such experiments generally two treatment conditions are compared, for example, an intervention and a control. The aim of the experiment is to estimate and test the between-group differences in outcomes. In health sciences, for example, experiments can be carried out to assess the effect of a newly developed smoking prevention program relative to a control on smoking behavior of adolescents. In medical sciences, the effect of a new medicine relative to an old medicine on the recovery of patients may be assessed. In educational sciences, the effect of a new curriculum on the math achievement of pupils may be established by comparing it with an old curriculum. When individuals are not nested within clusters, an important design issue concerns the required number of observations in each treatment group in order to achieve a certain power for the test on treatment effect. The formula for this number of experimental units can be found in, for example, Hsieh (1988). In practice, however, populations often have a multilevel structure with individuals nested within clusters that may themselves be nested within higher order clusters. In the smoking prevention example,

We wish to thank Tom Ambergen for checking the formulas in Table 2.

adolescents might be nested within schools or neighborhoods. For such experiments three design issues may arise.

The first issue concerns the optimal level of randomization to treatment conditions. In principle, randomization can be done at any level of the multilevel data structure. For example, in smoking prevention interventions with pupils nested within classes and classes nested within schools, we may choose to randomize pupils, classes or even entire schools to treatment conditions. Although the purpose of a treatment is mostly to produce an effect on individuals, in practice often entire clusters are randomized to treatment conditions. Reasons for randomization at the cluster level include ethical concerns, political and administrative factors, and the need to reduce costs and treatment group contamination, which occurs when information leaks from the intervention group to the control group (e.g., Donner, Brown, & Brasher, 1990). Apart from these, in community-based interventions, in which the intervention will necessarily affect all members of a cluster, there is no alternative to cluster randomization (e.g., Gail, Mark, Carroll, Green, & Pee, 1996).

The second design issue concerns the optimal allocation of units, that is, the optimal sample sizes at each level of the multilevel data structure given the budget for sampling and measuring units and the level of randomization. Is it preferable to sample a few clusters with a lot of individuals per cluster, or to sample many clusters with just a few individuals per cluster? In practice, a design with few clusters and many individuals per cluster is easier to obtain than a design with many clusters and a small number of individuals per cluster because sampling in an already selected cluster may be less expensive than sampling in a new cluster, and often only a few clusters are willing to participate in a study. On the other hand, the number of individuals to be sampled within a cluster cannot, of course, be larger than the total number of individuals in that cluster.

The third design issue is related to the second and concerns the required budget necessary to obtain a certain power on the test of no treatment effect, given the level of randomization.

So far, there has been little statistical research on these three design issues (see Goldstein 1995, section 11.2). Donner, Birkett, and Buck (1981) and Hsieh (1988) showed that randomization at the cluster level is less efficient in terms of estimator variance than randomization at the individual level and thus tests on treatment effect have less power to detect a real intervention effect. Raudenbush (1997) derived optimal allocations of units for cluster randomized designs with and without covariates and compared the efficiency of these two designs. Snijders and Bosker (1993) derived optimal allocations of units for two level designs with any number of explanatory variables at both levels. Mok (1996) compared two level designs in which the number of individuals per cluster is larger than, equal to, or smaller than the number of clusters by means of simulation studies. She based the comparisons on bias, sampling variance, and MSE of the parameter estimates. Donner, Birkett, and Buck (1981) and

Hsieh (1988) give sample size formulas for obtaining a certain power on the test concerning treatment effect for experiments in populations with two levels of nesting and with randomization to treatment conditions at the cluster level.

These results will be expanded in this study and the three design issues will be dealt with for populations with two and three levels of nesting and for continuous outcomes. The linear multilevel regression model is used to describe the relation between treatment condition and outcome. Although covariates may be included in the models, in experimental designs the regression coefficient associated with treatment condition is generally of main interest and has to be estimated as efficiently as possible to achieve a maximum power on the test of treatment effect. The optimal level of randomization that minimizes its variance will be presented and the increase in variance if randomization is done at another level will be studied in terms of relative efficiency. Next, the optimal allocation of units at which the variance of this estimator is minimized will be established, given the budget for measuring and sampling and the level of randomization. Thereafter, the required budget to achieve a certain power on the test of the treatment effect will be given. Finally, some conclusions and suggestions for further research will be given.

Models and Estimators

To help clarify the application of the methods in practice, an example is used throughout this paper: a smoking prevention intervention study with pupils nested within classes and classes nested within schools. Pupils are referred to as level one units, classes as level two units, and schools as level three units; we thus have three levels of nesting. Ignoring the nesting of classes within schools results in two levels of nesting. Pupils are indexed by i, classes by j, and schools by k. Throughout this paper, a balanced design will be assumed; that is, we have a sample of n_3 schools, n_2 classes per school, and n_1 pupils per class. Furthermore, it will be assumed that the variances of random effects are known. The models that will be introduced in this section can be used for the comparison of the effects of two different treatments (the smoking prevention intervention and a control) on the outcome of smoking behavior, which is assumed to be measured on a continuous scale. Randomization to these two treatment conditions can be done at the school level, the class level, or the pupil level. If randomization is done at the pupil level, $\frac{n_1}{2}$ pupils are randomized to the control group and $\frac{n_1}{2}$ pupils are randomized to the intervention group within each class, and thus n_1 has to be even. Similarly, if randomization is done at the class level, the number of classes per school has to be even; with randomization at the school level, the number of schools has to be even.

Our derivations will be based upon multilevel models (e.g., Goldstein & McDonald, 1988), also referred to as hierarchical linear models (e.g., Bryk & Raudenbush, 1992), or random coefficient models (e.g., Longford, 1995). Multilevel models can be used when we have a random sample at each level of the multilevel data structure. In the smoking prevention intervention example, we have a sample of schools from the entire population of schools, a sample of classes within each sampled school, and a sample of pupils within each sampled class. Since schools, classes, and pupils are assumed to be sampled randomly, these have to be included as random effects in our model. Treatment condition and covariates, however, are included as fixed effects. So our model contains both fixed and random effects and is a mixed effect model.

In this paper, we deal with models without covariates (but see the end of this section). For three levels of nesting the model that relates smoking behavior, denoted by y, to treatment condition, denoted by x, is given by

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + v_k + u_{jk} + e_{ijk}, \tag{1}$$

where v_k, u_{jk}, and e_{ijk} are random effects at the school (k), class (j), and pupil (i) levels, respectively. These random effects are independently distributed with zero mean and variances σ_v^2, σ_u^2, and σ_e^2, respectively. If randomization is done at the class level the i can be left out of the subscript of x in (1). Similarly, if randomization is done at the school level the i and j can be left out of the subscript of x. β_1 is the regression coefficient associated with treatment condition and is considered to be fixed in this article.

For each level of randomization, the variance of a pupil's outcome, given the fixed part $\beta_0 + \beta_1 x_{ijk}$, consists of three components, the so-called variance components:

$$\text{var}(y_{ijk}) = \sigma_v^2 + \sigma_u^2 + \sigma_e^2. \tag{2}$$

The outcomes of pupils i and i' within the same class j are dependent, and their covariance is

$$\text{cov}(y_{ijk}, y_{i'jk}) = \sigma_v^2 + \sigma_u^2, \text{ for } i \neq i'. \tag{3}$$

Similarly the dependency of outcomes of pupils i and i' within the same school k but within different classes j and j' is given by

$$\text{cov}(y_{ijk}, y_{i'j'k}) = \sigma_v^2, \text{ for } j \neq j'. \tag{4}$$

With two levels of nesting the subscript k, the random effect v_k and the variance component σ_v^2 can be left out of (1), (2), and (3); (4) does not apply.

Since the outcomes are correlated and the variances of the random effects are assumed to be known, the generalized least squares estimator is used for the multilevel model in (1). For a three-level design, β_1 is estimated unbiasedly by

$$\hat{\beta}_1 = \frac{\sum_{ijk} x_{ijk} y_{ijk}}{n_1 n_2 n_3} = \frac{\bar{y}_t - \bar{y}_c}{2}, \tag{5}$$

in which $\bar{y}_t$ and $\bar{y}_c$ are the mean responses in the intervention and control group, respectively. For a two-level design, the subscript k is left out and n_3 is set equal to unity. Formula (5) holds if treatment groups are coded by -1 and 1. If they are coded by 0 and 1, the estimator of β_1 is equal to $\bar{y}_t - \bar{y}_c$. For both two and

TABLE 1
Var($\hat{\beta}_1$) for each Level of Randomization

Number of levels	Level of randomization	Var($\hat{\beta}_1$)
three	pupil	$\frac{\sigma_e^2}{n_1 n_2 n_3}$
three	class	$\frac{n_1\sigma_u^2 + \sigma_e^2}{n_1 n_2 n_3}$
three	school	$\frac{n_1 n_2\sigma_v^2 + n_1\sigma_u^2 + \sigma_e^2}{n_1 n_2 n_3}$
two	pupil	$\frac{\sigma_e^2}{n_1 n_2}$
two	class	$\frac{n_1\sigma_u^2 + \sigma_e^2}{n_1 n_2}$

Note. These variances hold if treatment conditions are coded −1 and 1.

three levels of nesting and for each level of randomization the variance of the estimator associated with treatment condition, that is, var($\hat{\beta}_1$), is given in Table 1. These variances only hold for treatment groups coded by −1 and 1. If they are coded by 0 and 1, the variances as given in Table 1 have to be multiplied by four. Note that var($\hat{\beta}_1$) for two levels of nesting and randomization at the class level is a special case of Equation (37) in Snijders and Bosker (1993).

In our discussion we have not included covariates in our models. In general, however, the outcome variable smoking behavior will be influenced by covariates such as pre-treatment smoking behavior or school type. It can be shown that if such covariates are grand-mean centered and if their coefficients are fixed parameters, the results in Table 1 hold approximately, especially if the sample sizes at each level are not too small. This is due to the experimental (randomized) design, which ensures that treatment condition and covariate are uncorrelated, apart from some sampling error (Moerbeek, van Breukelen, & Berger, 1997). Consequently, the results in the following sections also hold approximately.

Optimal Level of Randomization

As written in the introduction, the optimal level of randomization is that level for which var($\hat{\beta}_1$) is minimized. From Table 1 it is clear that var($\hat{\beta}_1$) decreases if randomization to treatment conditions is performed at a lower level and thus the pupil level is the optimal level of randomization. Randomization at a higher level often leads to higher values for var($\hat{\beta}_1$). This can be seen by considering the relative efficiency of $\hat{\beta}_1$. The relative efficiency of two unbiased estimators

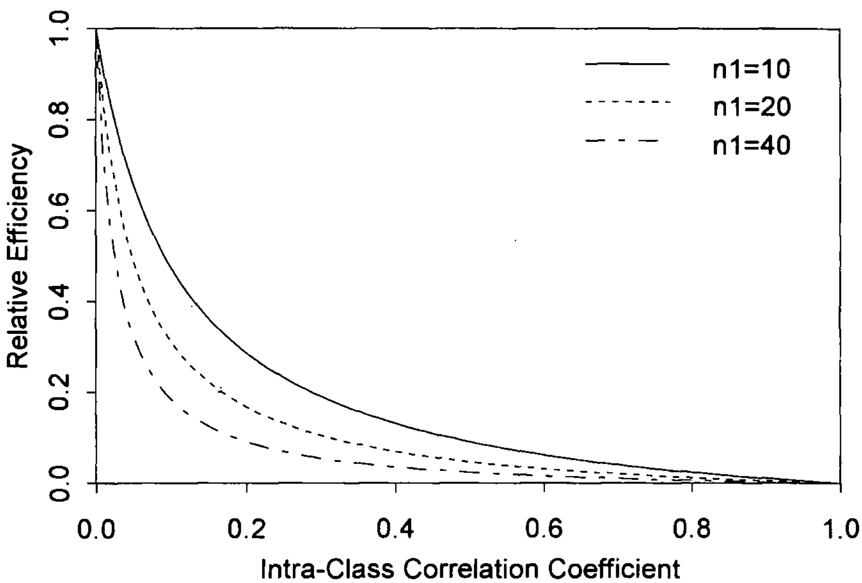

FIGURE 1. *Relative efficiency of* $\hat{\beta}_1$ *for randomization at the class level versus randomization at the pupil level as a function of the intra-class correlation coefficient*

of β_1 is defined as the ratio of the reciprocal of their variances. For two levels of nesting,

$$\text{relative efficiency} = \frac{\sigma_e^2}{n_1\sigma_u^2 + \sigma_e^2},$$

where the var($\hat{\beta}_1$) for randomization at the class level is related to var($\hat{\beta}_1$) for randomization at the pupil level. In Figure 1 the relative efficiency is plotted as a function of n_1 and the intra-class correlation coefficient ρ, which measures the amount of variance between classes, that is $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$. Figure 1 shows that if ρ or n_1 increase, the relative efficiency decreases. For low values of ρ the decrease in relative efficiency is already considerable. For $\rho = 0.1$ the relative efficiency is equal to 0.47, 0.31, and 0.18, for $n_1 = 10$, $n_1 = 20$, and $n_1 = 40$, respectively. When ρ approaches unity the relative efficiency goes to zero. Similarly, it can be inferred from Table 1 that for three levels of nesting the relative efficiency of $\hat{\beta}_1$ obtained with randomization at the school level versus randomization at the pupil level decreases with n_1, n_2, the intra-class correlation coefficient, and the intra-school correlation coefficient, which measures the amount of variance between schools.

Optimal Allocation of Units

Although the lowest level of the multilevel data structure is the optimal level of randomization, randomization at this level is not always possible. Therefore, the optimal allocation of units will be derived for each level of randomization.

As explained in the introduction, the optimal allocation of units consists of those values n_1, n_2 and n_3 for which $\mathrm{var}(\hat{\beta}_1)$ is minimized. Optimal allocations are now derived under the condition that the budget for sampling and measuring units is fixed to C. We require to

$$\text{minimize } \mathrm{var}(\hat{\beta}_1),$$

subject to the condition

$$C = c_1 n_1 n_2 n_3 + c_2 n_2 n_3 + c_3 n_3 \quad (c_l > 0,\ n_l \geq 2 \text{ for } l = 1, 2, 3) \tag{6}$$

for three levels of nesting, or to the condition

$$C = c_1 n_1 n_2 + c_2 n_2 \quad (c_l > 0,\ n_l \geq 2 \text{ for } l = 1, 2) \tag{7}$$

for two levels. Here c_3 are the costs for sampling and measuring a school, c_2 are the costs for sampling and measuring a class in an already sampled school, and c_1 are the costs for sampling and measuring a pupil in an already sampled class. c_3 may vary across school, for instance, because of different travel times. However, at the time of planning the experiment it is not known which schools will be sampled. Therefore, an average value of c_3 has to be used for planning the experiment. In this article, C, c_1, c_2, and c_3 are the budget and costs in dollars, but of course any other currency can also be used. We restrict n_1, n_2, and n_3 to be at least two in order to obtain a multilevel data structure. For three levels of nesting the optimal n_l can be obtained by expressing either n_1, or n_2, or n_3 in terms of the other two n_l and the costs using (6), substituting this expression in the formula for $\mathrm{var}(\hat{\beta}_1)$ as given in Table 1, and solving for the other two n_l. For two levels of nesting the procedure is comparable. The optimal values of the n_l thus obtained and the minimum $\mathrm{var}(\hat{\beta}_1)$ under the optimal allocation are given in Tables 2 and 3 respectively. A lower $\mathrm{var}(\hat{\beta}_1)$ can be obtained by using a higher budget C, which results in sampling more units only at the level at which randomization is done since the sample sizes at the other levels are independent of C (see Table 2). For two levels of nesting with randomization at the class level, the optimal allocation corresponds with equation (39) in Snijders and Bosker (1993). Note that the values of n_l other than 2 in Table 2 are obtained by treating the n_l as continuous variables, which actually are integers. Thus the values of the n's obtained with the formulas in Table 2 need to be rounded to the nearest integer in such a way that the budget is not exceeded; $\mathrm{var}(\hat{\beta}_1)$ may somewhat exceed the minimum value given by Table 3. In the next section an example will illustrate these results.

An Example

In a multilevel design the costs of sampling and measuring at a higher level are often larger than the costs of samplings and measuring at a lower level. Furthermore, there is often more variation between lower level units than between higher level units. The following set of values is an example of these

TABLE 2
Optimal Allocation of Units for each Level of Randomization

Number of levels	Level of random-ization	n_1	n_2	n_3
three	pupil	$\frac{C - 4c_2 - 2c_3}{4c_1}$	2	2
three	class	$\frac{\sigma_e}{\sigma_u}\sqrt{\frac{c_2}{c_1}}$	$\frac{\frac{1}{2}C - c_3}{\frac{\sigma_e}{\sigma_u}\sqrt{c_1c_2} + c_2}$	2
three	school	$\frac{\sigma_e}{\sigma_u}\sqrt{\frac{c_2}{c_1}}$	$\frac{\sigma_u}{\sigma_v}\sqrt{\frac{c_3}{c_2}}$	$\frac{C}{c_3 + \frac{\sigma_u}{\sigma_v}\sqrt{c_2c_3} + \frac{\sigma_e}{\sigma_v}\sqrt{c_1c_3}}$
two	pupil	$\frac{C - 2c_2}{2c_1}$	2	—
two	class	$\frac{\sigma_e}{\sigma_u}\sqrt{\frac{c_2}{c_1}}$	$\frac{C}{\frac{\sigma_e}{\sigma_u}\sqrt{c_1c_2} + c_2}$	—

Note. For two level designs the nesting of classes within schools is omitted and thus the optimal n_3 is not given. Treatment conditions are coded −1 and 1.

cases: $c_1 = \$1$, $c_2 = \$2$, $c_3 = \$3$, $C = \$200$, $\sigma_e^2 = 16$, $\sigma_u^2 = 2$, and $\sigma_v^2 = 0.5$. For these values the optimal allocation of units can be calculated from Table 2 and is shown in Figures 2, 3, and 4.

In Figure 2, contour lines of var($\hat{\beta}_1$) for randomization at the pupil level are plotted in the (n_1, n_2) plane for the following values of var($\hat{\beta}_1$): 0.09, 0.095, 0.10, 0.15, 0.11. The lines $n_1 = 2$, $n_2 = 2$, and $n_3 = 2$ are shown as dotted lines. In the area enclosed by these three dotted lines the conditions given in (6) are satisfied; thus the optimal allocation of units under these conditions is located within this area. Clearly, var($\hat{\beta}_1$) decreases as n_1 increases, and the optimal allocation is given by point A in Figure 2, where $n_2 = n_3 = 2$, $n_1 = 46.5$, and var($\hat{\beta}_1$) = 0.086 from Table 3. We take $n_1 = 46$, as it has to be an integer, giving $C = c_1n_1n_2n_3 + c_2n_2n_3 + c_3n_3 = 198$ and var($\hat{\beta}_1$) = 0.087, virtually unchanged. However, a value of 46 pupils per class may be too large for practical purposes and different allocations of units may be more realistic. In this example, the minimum is a flat one, and so certain other allocations of units for which the budget is not exceeded will give much the same values for var($\hat{\beta}_1$). For example, if $n_1 = 21$, $n_2 = 2$, $n_3 = 4$, then $C = 196$ and var($\hat{\beta}_1$) = 0.095, while if

TABLE 3
$Var(\hat{\beta}_1)$ under the Optimal Allocation of Units and for Each Level of Randomization

Number of levels	Levels of randomization	$Var(\hat{\beta}_1)$
three	pupil	$\dfrac{\sigma_e^2 c_1}{C - 4c_2 - 2c_3}$
three	class	$\dfrac{(\sigma_u\sqrt{c_2} + \sigma_e\sqrt{c_1})^2}{C - 2c_3}$
three	school	$\dfrac{(\sigma_v\sqrt{c_3} + \sigma_u\sqrt{c_2} + \sigma_e\sqrt{c_1})^2}{C}$
two	pupil	$\dfrac{\sigma_e^2 c_1}{C - 2c_2}$
two	class	$\dfrac{(\sigma_u\sqrt{c_2} + \sigma_e\sqrt{c_1})^2}{C}$

Note. Treatment conditions are coded −1 and 1.

$n_1 = 10$, $n_2 = 8$, $n_3 = 2$, then $C = 198$ and $var(\hat{\beta}_1) = 0.100$. If the difference between a value such as 0.095 or 0.100 and the formal minimum 0.087 is not worth bothering, these allocations may be considered as an alternative to the formal optimum (46, 2, 2).

In Figure 3, contour lines of $var(\hat{\beta}_1)$ for randomization at the class level are plotted for $var(\hat{\beta}_1) = 0.19$, 0.20, 0.21, 0.22, and 0.23. Again, the conditions given in (6) are satisfied in the area enclosed by the dotted lines. Figure 3 shows that $var(\hat{\beta}_1)$ decreases as n_1 increases and then increases, while for given n_1, $var(\hat{\beta}_1)$ decreases as n_2 increases. The optimal allocation is given by point A where $n_1 = 4$, $n_2 = 16.2$, $n_3 = 2$, and $var(\hat{\beta}_1) = 0.186$, as follows from Tables 2 and 3. Rounding to $n_2 = 16$ gives $C = c_1 n_1 n_2 n_3 + c_2 n_2 n_3 = c_3 n_3 = 198$ and $var(\hat{\beta}_1)$ increase somewhat to 0.188.

For randomization at the school level, contour lines are plotted in Figure 4 for the following values of $var(\hat{\beta}_1)$: 0.27, 0.30, 0.33, and 0.36. The optimal allocation of units is given by point A where $n_1 = 4$, $n_2 = 2.4$, $n_3 = 11.3$, and $var(\hat{\beta}_1) = 0.261$ (see Tables 2 and 3). As n_1, n_2, and n_3 have to be integers, we take $n_2 = 2$ and $n_3 = 12$ for which $C = c_1 n_1 n_2 n_3 + c_2 n_2 n_3 + c_3 n_3 = 180$ and $var(\hat{\beta}_1) = 0.292$. However, a lower $var(\hat{\beta}_1)$ can be obtained with $n_1 = 6$, $n_2 = 2$, and $n_3 = 10$, giving $C = c_1 n_1 n_2 n_3 + c_2 n_2 n_3 + c_3 n_3 = 190$ and $var(\hat{\beta}_1) = 0.283$.

Minimizing the Budget

So far, the allocation of units minimizing $var(\hat{\beta}_1)$ under the condition that the budget for measuring the sampling is fixed to a certain value has been derived.

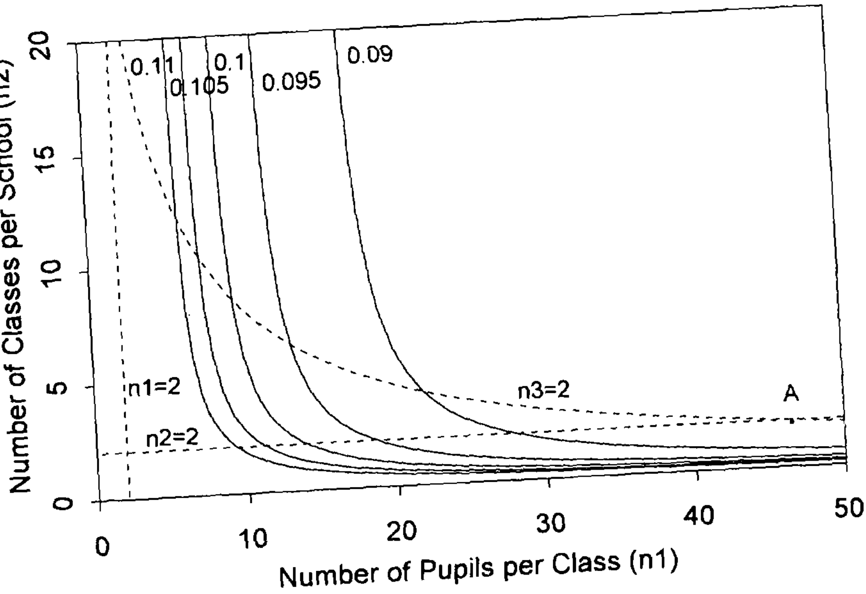

FIGURE 2. *Preconditions (dashed lines) and contour lines for* $var(\hat{\beta})$ *(solid lines) in the* (n_1, n_2)*-plane; three levels of nesting, randomization at the pupil level, and* $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $C = 200$, $\sigma_e^2 = 16$, $\sigma_u^2 = 2$, $\sigma_v^2 = 0.5$

But it may also be worthwhile to derive the allocation of units for minimizing the budget for measuring and sampling to obtain a certain value $var(\hat{\beta}_1)$. Again, we have an optimization problem: The budget for measuring and sampling as given in (6) or (7) has to be minimized under the condition that $var(\hat{\beta}_1)$ as given in Table 1 is fixed. The solution to this problem can directly be derived from Tables 2 and 3 by considering the following. If the budget for measuring and sampling is equal to C, the allocation of units as given in Table 2 results in a minimal value of $var(\hat{\beta}_1)$ as given in Table 3. So, this budget C is also the minimal budget to obtain that particular value of $var(\hat{\beta}_1)$, since if there existed a smaller budget with other allocation yielding the same $var(\hat{\beta}_1)$, then our allocation given C would not be optimal. Thus, if a value of $var(\hat{\beta}_1)$ equal to V is required, the minimal budget to obtain this $var(\hat{\beta}_1)$ follows by setting $var(\hat{\beta}_1)$ (as given in Table 3) equal to V and solving for the budget C. The corresponding optimal allocation of units then follows from Table 2. For example, assume that randomization is done at the school level, $c_1 = \$1$, $c_2 = \$2$, $c_3 = \$3$, $\sigma_e^2 = 16$, $\sigma_u^2 = 2$, and $\sigma_v^2 = 0.5$, and $var(\hat{\beta}_1)$ has to be equal to 0.2. The budget then has to be equal to \$261 with corresponding optimal allocation (before rounding) $n_1 = 4$, $n_2 = 2.4$, and $n_3 = 14.7$.

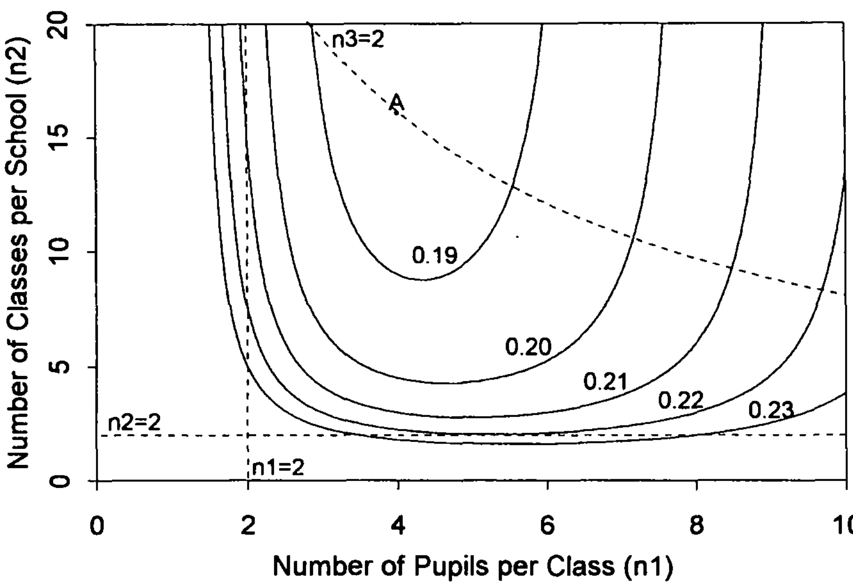

FIGURE 3. *Preconditions (dashed lines) and contour lines for* var($\hat{\beta}_1$) *(solid lines) in the* (n_1, n_2)*-plane; three levels of nesting, randomization at the class level, and* $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $C = 200$, $\sigma_e^2 = 16$, $\sigma_u^2 = 2$, $\sigma_v^2 = 0.5$

Required Budget to Obtain a Certain Power for the Test of No Treatment Effect

Suppose we are testing the hypothesis that the population value for the regression coefficient β_1 associated with treatment effect is zero, that is, H_0: $\beta_1 = 0$. The *third design issue* is the budget required for this test to have a specified power when β_1 has a given value, say $\beta_1 = \delta$. The test statistic is $z = \hat{\beta}_1/\sqrt{\text{var}(\hat{\beta}_1)}$; under H_0, z has the standard normal distribution since the variance components are assumed known. For one-sided tests with alternatives, H_1: $\beta_1 > 0$ or H_2: $\beta_1 < 0$, with power $1 - \gamma$ at $\beta_1 = \delta$, and with significance level α, we have (e.g., Cochran, 1983):

$$\sqrt{\text{var}(\hat{\beta}_1)} = \frac{|\delta|}{z_{1-\alpha} + z_{1-\gamma}}, \qquad (8)$$

where $z_{1-\alpha}$ and $z_{1-\gamma}$ are the $100(1 - \alpha)\%$ and $100(1 - \gamma)\%$ standard normal deviates; δ, of course, is positive for alternative H_1 and negative for alternative H_2. For two-sided tests with alternative H_3: $\beta_1 \neq 0$, $z_{1-\alpha}$ in (8) is replaced by $z_{1-(\alpha/2)}$.

In planning the experiment, δ is first chosen (bearing in mind that β_1 denotes *half* of the true between-group difference, because treatment conditions are coded -1 and 1); the required value of var($\hat{\beta}_1$) is then given by (8), and the

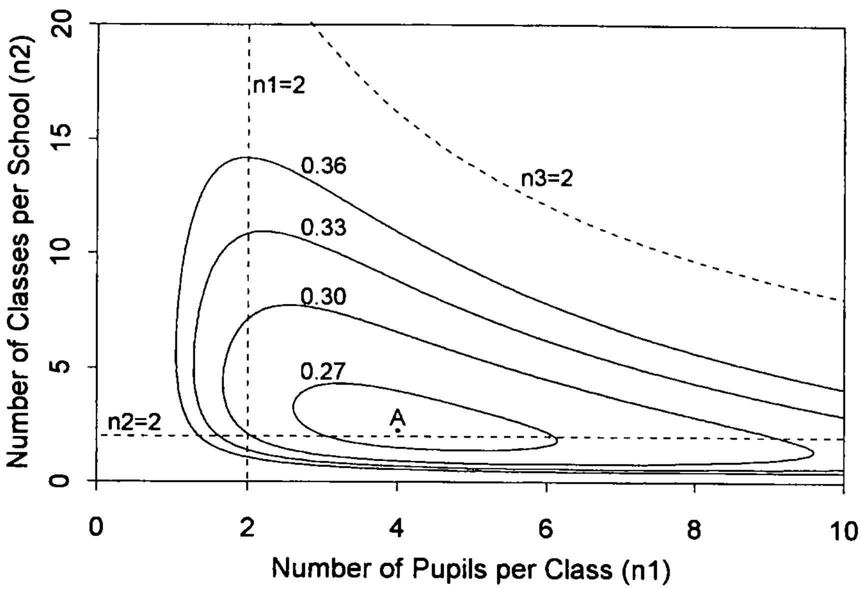

FIGURE 4. *Preconditions (dashed lines) and contour lines for var*$(\hat{\beta}_1)$ *(solid lines) in the* (n_1, n_2)*-plane; three levels of nesting, randomization at the school level, and* $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $C = 200$, $\sigma_e^2 = 16$, $\sigma_u^2 = 2$, $\sigma_v^2 = 0.5$

required budget and allocation of units follow as in the previous section. This is illustrated in the following example.

An Example

Suppose that three levels of nesting are distinguished and that we want to test H_0: $\beta_1 = 0$ versus H_a: $\beta_1 > 0$ at a significance level $\alpha = 0.05$ and with power $1 - \gamma = 0.90$ at the value $\delta = 1$ for β_1. From (8), $\text{var}(\hat{\beta}_1) = [1/(1.645 + 1.282)]^2 = 0.1167$. Suppose also that $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $\sigma_e^2 = 16$, $\sigma_u^2 = 2$, and $\sigma_v^2 = 0.5$, which is the set of parameter values we used in the previous section. If randomization is done at the pupil level a budget of $C = \$151$ is needed for this test. If randomization is done at the class level, we need a budget $C = \$314$ and, if randomization is done at the school level, a budget of $C = \$447$.

Discussion and Conclusion

The first step in designing experiments in multilevel populations is the establishment of the optimal level of randomization, the lowest level. However, there may be practical reasons for randomization at a higher level, resulting in a higher required budget to obtain a certain power on the test of treatment effect. Once the required budget is calculated using (8), the optimal allocation follows from Table 2.

In experiments the effect of the treatment condition is generally of main interest, although covariates may be included in the multilevel regression model. Therefore, the derivation of the optimal level of randomization and the optimal allocation of units was based on the regression coefficient of this treatment factor. When the covariates, which may be fixed or random, are also of interest a D-optimal design would probably be more appropriate.

It must be emphasized that this study assumed a balanced design and randomization into equal numbers of cases and controls, but it is also useful to derive optimal designs for more general cases. Furthermore, it was assumed that the variance components were known, but it is also necessary to establish the effect of unknown variance components on var($\hat{\beta}_1$) and on the distribution of the test statistic for the test of no treatment effect. Research on these issues is now in progress. First results show that the formulas in Table 1 still hold for unknown variance components if σ_e^2, σ_u^2, and σ_v^2 are replaced by their estimates. Furthermore, the test statistic for the test on treatment effect seems to be t-distributed if the variance components are unknown.

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park: Sage Publications.

Cochran, W. G. (1983). *Planning and analysis of observational studies.* New York: Wiley.

Donner, A., Birkett, N., & Buck, C. (1981). Randomization by cluster, sample size requirements and analysis. *American Journal of Epidemiology, 114*, 906–914.

Donner, A., Brown, K. S., & Brasher, P. (1990). A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *International Journal of Epidemiology, 19*, 795–800.

Gail, M. H., Mark, S. D., Carroll, R. J., Green, S. B., & Pee, D. (1996). On design considerations and randomization-based inference for community intervention trials. *Statistics in Medicine, 15*, 1069–1092.

Goldstein, H. (1995). *Multilevel Statistical Models* (2nd ed.). London: Edward Arnold.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53*, 455–467.

Hsieh, F. Y. (1988). Sample size formulae for intervention studies with the cluster as unit of randomization. *Statistics in Medicine, 8*, 1195–1201.

Longford, N. T. (1995). *Random coefficient models.* Oxford: Clarendon Press.

Moerbeek, M., Van Breukelen, G.J.P., & Berger, M.P.F. (1997). *A comparison of the mixed effect, the fixed effect and the data aggregation model for multilevel data.* (Tech. Rep. No. 9701). Maastricht, the Netherlands: Maastricht University, Department of Methodology and Statistics.

Mok, M. (1996). Sample size requirements for 2-level designs in educational research. *Multilevel Modelling Newsletter, 7*, 11–16.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173–185.

Snijders, T.A.B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*, 237–259.

Authors

MIRJAM MOERBEEK is currently employed as a biostatistician at the National Institute of Public Health and the Environment, P. O. Box 1, 3720 BA Bilthoven, the Netherlands; Mirjam.Moerbeek@RIVM.NL. Her research interests include multilevel models and optimal experimental designs.

GERARD J. P. VAN BREUKELEN is Assistant Professor, Department of Methodology and Statistics, Maastricht University, P. O. Box 616, 6200 MD Maastricht, the Netherlands; Gerard.vBreukelen@STAT.UNIMAAS.NL. His research interests include mixed effect regression in the context of field experiments and psychomterics.

MARTIJN P. F. BERGER is Professor, Department of Methodology and Statistics, Maastricht University, P. O. Box 616, 6200 MD Maastricht, the Netherlands; Martijn.Berger@STAT.UNIMAAS.NL. He specializes in statistical models, optimal design, and psychometrics.