Journal of Official Statistics, Vol. 18, No. 2, 2002, pp. 291–303

Multiple-Objective Optimal Designs for the Hierarchical Linear Model

Mirjam Moerbeek[1] *and Weng Kee Wong*[2]

Optimal designs are usually constructed under a single optimality criterion. Such designs are not very realistic in practice because a researcher seldom has just one objective in mind when designing an experiment. This problem can be overcome by multiple-objective designs. Here, we extend previous work and construct multiple-objective designs for situations where the data are correlated using a hierarchical linear model. We present a graphical method for constructing multiple-objective designs and investigate their robustness properties.

Key words: Hierarchical data; sample size; optimality criteria; variance components.

1. Introduction

Data obtained from educational, social, and behavioral studies often have a hierarchical structure, which means that individuals are nested within clusters. Examples are employees nested within work-sites, pupils nested within schools, and students nested within universities. Traditional regression models assume that the outcomes of individuals within a cluster are independent. This assumption, however, is not very realistic, because individuals influence each other's behavior, attitude, and health by direct communication and shared group norms. Furthermore, their outcomes may be influenced by, for example, cluster environment, cluster policy, or cluster leaders. It is therefore reasonable to assume that outcomes of individuals within the same cluster are correlated, while outcomes of individuals within different clusters are independent. An appropriate model for the analysis of such data is the hierarchical linear model (Bryk and Raudenbush 1992), which is also referred to as the multilevel model (Goldstein 1995; Hox 1994; Kreft and De Leeuw 1998; Snijders and Bosker 1999), or random coefficient model (Longford 1995). This type of model contains both fixed and random effects and thus is a mixed-effects model. The assumption is that the clusters represent a random sample from a larger population of clusters so that the results of the study may be generalized to this larger population if their effects are treated as random in the model.

[1] Department of Methodology and Statistics, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, the Netherlands. E-mail: M.Moerbeek@fss.uu.nl
[2] Department of Biostatistics, School of Public Health, University of California at Los Angeles, Box 951772, 10833 Le Conte Avenue, Los Angeles, CA 90095-1772, U.S.A. E-mail: wkwong@sunlab.ph.ucla.edu
Acknowledgments: The research of M. Moerbeek is supported by the Netherlands America Commission for Educational Exchange (NACEE), the Netherlands' Organization for Scientific Research (NWO), and the Carolus Magnus Fund; the research of W.K. Wong is supported by NIH grant R29 AR44177-01A1.

The design of educational, social, and behavioral studies is more complicated for hierarchical data structures because not only the total number of individuals needs to be established, but also the number of clusters and the number of individuals per cluster. During the last decade, a number of papers on sample size calculations for hierarchically structured populations has been published. Some of these assume that either the number of clusters, or the number of individuals per cluster is fixed (Donner 1998; Feng and Grizzle 1992; Hsieh 1998; Lee and Dubin 1994; Liu and Liang 1997; Shih 1997). Another approach is to calculate optimal designs under the restriction that there is a fixed budget for sampling clusters and individuals, which may not be exceeded by the costs of the study. This approach was used by Cohen (1998), Moerbeek, Van Breukelen, and Berger (2000, 2001a, 2001b), Raudenbush (1997), and Snijders and Bosker (1993). Simulation studies to derive the optimal sample sizes have been done by Afshartous (1995) and Mok (1995).

Most of the papers mentioned above use the variance of the treatment effect estimator as the sole design optimality criterion and so the optimal design is called a single-objective optimal design. In practice, there are usually multiple objectives in the study and it is desirable to consider their multi-faceted goals simultaneously in the design stage. An optimal design so constructed is called a multiple-objective optimal design and such a design is especially appropriate when there are competing interests in the study. Recent research on multiple-objective optimal designs focused on models for uncorrelated outcomes (Huang 1996; Huang and Wong 1998a, 1998b; Zhu and Wong 2000a,b). There is hardly any research on multiple-objective optimal designs for correlated outcomes. An exception is Cohen (1998), who briefly discussed a seminal approach to multiple-objective optimal designs for hierarchical linear models.

The aim of our article is to derive multiple-objective optimal designs for hierarchical linear models. In the next section we give a summary of the results on optimal sample sizes for single-objective optimal designs given by Cohen (1998), as well as an account of the robustness of such designs. In the following section we show that a design that is optimal for a certain optimality criterion may be far from optimal if another optimality criterion is used. In such cases, one may want to select a design that has high efficiency for the more important optimality criterion, and at the same time does as well as possible under the other criterion. Such a design is called a constrained optimal design, and we will show how it can be obtained graphically using efficiency plots. Efficiency plots make use of the fact that under certain conditions a constrained optimal design is the same as an optimal design which minimizes a convex combination of the objectives. The latter optimal designs are called compound optimal designs. We show how the weights of the compound optimal design can be chosen to obtain the ''desired'' constrained optimal design. We will also discuss the robustness of multiple-objective optimal designs. The last section of this article contains a discussion and our conclusions.

2. Single-Objective Optimal Designs

2.1. Results on single-objective optimal designs

Cohen (1998) derived optimal designs for the model which relates the outcome y_{ij} of student i within school j to a dummy variable at the student level, x_{1ij}, and one at the school level, x_{2j}. Both dummy variables were coded by the values 0 and 1. In total, there were m schools with n pupils each in the sample. The student-level dummy variable was assumed balanced within schools, that is, $x_{1ij}=0$ for half of the students within each school, and $x_{1ij}=1$ for the remaining pupils in the school. The school level dummy variable was assumed balanced overall (i.e., $\frac{1}{2}m$ schools have $x_{2j}=0$ while the other schools have $x_{2j}=1$). A dummy variable may represent an experimental condition (intervention versus control group) or a dichotomous demographic variable (e.g., gender). In the first case it is balanced if randomization is done to treatment groups of equal size, in the second case if samples of equal size are obtained from both values of the demographic variable. We use balanced dummy variables and non-varying school sizes because then the maximum likelihood estimators of the regression coefficients associated with these dummy variables have smallest variance, and the optimal sample size formulae are of practical use.

Because no interaction between schools and the student level dummy variable was assumed, the hierarchical linear model was given by:

$$y_{ij} = \gamma_0 + \gamma_1 x_{1ij} + \gamma_2 x_{2j} + u_{0j} + r_{ij} \tag{1}$$

where the random student-level variables r_{ij} are independently and normally distributed with zero mean and variance σ^2. The random school level variables u_{0j} are independent of each other and the r_{ij}, and have a normal distribution with zero mean and variance τ^2. The intra-school correlation coefficient ρ measures the amount of variance at the school level and is calculated as:

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{2}$$

Choosing a single-objective optimal design means finding a design ξ^* among all the designs ξ such that an optimality criterion $\Theta(\xi)$ is minimized. In this article the optimality criterion $\Theta(\xi)$ is the variance of the estimator of a regression coefficient or a variance component. For the hierarchical linear model, the optimal design ξ^* minimizes this variance, and gives the optimal number of schools and the optimal number of students per school to be included in the study. Cohen (1998) derived optimal designs given the precondition that the costs for measuring and sampling do not exceed the budget. If C_s denotes the cost to include a school into the study, C_k denotes the cost to include a student (‘‘kid’’), and C is the budget, then the cost function is given by

$$C_k nm + C_s m = C \tag{3}$$

Here and throughout we assume $C \gg C_s$. Five different optimality criteria were used by Cohen; each of these seeks to minimize the variance of one of the five parameter estimates. These optimality criteria are a special case of the c-optimality criterion because the parameters to be estimated all have the form $c'\theta$ where $\theta = (\gamma_0, \gamma_1, \gamma_2, \sigma^2, \tau^2)'$. The c-optimal design gives the minimum variance of the estimator of a linear combination of the parameters. When, for example, var($\hat{\gamma}_0$) is to be minimized, the first entry in the vector c is equal to one and the others are equal to zero. Table 1 gives a summary of the results in Cohen’s paper; the subscripts in the optimality criteria refer to the parameter of which the variance of its estimator has to be minimized. Thus, a c_0-, c_1-, or c_2-optimal

Table 1. Optimality criteria and optimal number of students per school

optimality criterion	$\Theta(\xi)$	optimal n
c_0	$\mathrm{var}(\hat{\gamma}_0) = \frac{2n\tau^2+3\sigma^2}{nm}$	$\sqrt{\frac{3C_s(1-\rho)}{2C_k\rho}}$
c_1	$\mathrm{var}(\hat{\gamma}_1) = \frac{4\sigma^2}{nm}$	$\frac{C-10C_s}{10C_k}$
c_2	$\mathrm{var}(\hat{\gamma}_2) = \frac{4n\tau^2+4\sigma^2}{nm}$	$\sqrt{\frac{C_s(1-\rho)}{C_k\rho}}$
c_s	$\mathrm{var}(\hat{\sigma}^2) = \frac{2\sigma^2}{mn-m}$	$\frac{C-10C_s}{10C_k}$
c_t	$\mathrm{var}(\hat{\tau}^2) \approx \frac{2}{mn}\left(\frac{\sigma^4}{n} + 2\tau^2\sigma^2 + n\tau^4\right)$	$\frac{\sqrt{C_k\left(C_k+8C_s\frac{\tau^2}{\sigma^2}\right)}+C_k}{2C_k\frac{\tau^2}{\sigma^2}}$

design minimizes the variance of $\hat{\gamma}_0$, $\hat{\gamma}_1$, or $\hat{\gamma}_2$, respectively, and a c_s- or c_t-optimal design minimizes the variance of $\hat{\sigma}^2$ or $\hat{\tau}^2$, respectively. The optimal m can be derived by substitution of the optimal n in Formula (3). Note that an approximation to the $\mathrm{var}(\hat{\tau}^2)$ was used in order to avoid a formula for the optimal n that is too cumbersome to be of any practical use.

The optimal sample sizes for estimating γ_0, γ_2, and τ^2 with minimal variance depend on the intra-school correlation coefficient, which is generally unknown in practice. The optimal design may be calculated for several realistic values of ρ so that the uncertainty when deriving an optimal design may thus be averaged out.

The $\mathrm{var}(\hat{\gamma}_1)$ and $\mathrm{var}(\hat{\sigma}^2)$ are minimized by sampling just one school with as many students per school as possible such that the costs do not exceed the budget. However, a sample of just one is too small to use the hierarchical linear model. Snijders and Bosker (1999, p. 44) give a rule of thumb for regarding effects of schools as fixed statistical parameters or random variables. With the first approach the school effects are represented by dummy variables and the effects of school level covariates cannot be tested since all between-school variability is already explained by the dummy variables. This is not the case when schools are represented by random variables (i.e., when the hierarchical linear model is used), and for this approach the results from the study may be generalized to a larger population of schools. The rule of thumb says that a minimum of ten schools should be sampled for the hierarchical linear model, and the optimal number of students per school for criteria c_1 and c_s as given in Table 1 satisfy this rule of thumb.

It may happen that the number of students in some of the sampled schools is smaller than the optimal n as given in Table 1. In this case, all students within these schools should be used and the unused part of the budget may be used for sampling extra students from the other schools. Of course, this design is less optimal since it leads to larger variances of the estimated regression coefficients and variance components.

Figure 1 shows the optimal n for criteria c_0, c_2, and c_t as a function of the intra-school correlation coefficient for the parameter values $C = 10{,}000$, $C_s = 30$, and $C_k = 1$, which were also used by Cohen. The optimal value of n is $n = 970$ for $\rho = 0$. The optimal sample sizes for criteria c_1 and c_s do not depend on the intra-school correlation coefficient and

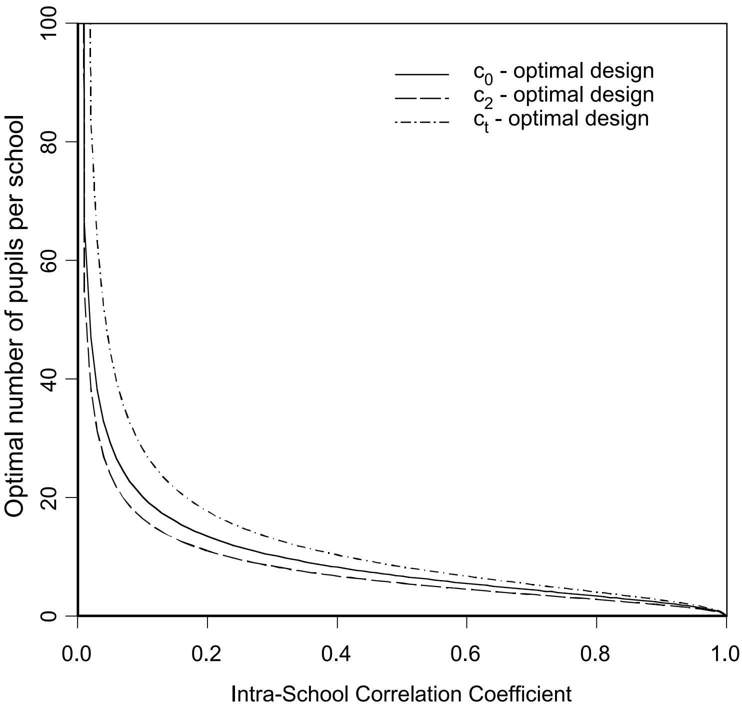

Fig. 1. Optimal number of students per school for c_0-, c_2-, and c_t-optimality. The optimal number of students per school is equal to 970 for c_1- and c_2-optimality

the optimal value of n is $n = 970$. As is clear, the optimal sample sizes for criteria c_1 and c_s are far from optimal for the other three criteria. This suggests optimal sample sizes can vary substantially from one optimality criterion to the other. The robustness of optimal designs against other optimality criteria will be discussed in the next subsection.

2.2. *Robustness of single-objective optimal designs*

Robustness properties of optimal designs are generally evaluated by their relative efficiency. The efficiency of design ξ relative to the optimal design ξ^* for the criterion $\Theta(\xi)$ is given by

$$Eff_{\Theta(\xi)} = \frac{\Theta(\xi^*)}{\Theta(\xi)} \tag{4}$$

This ratio is a constant between zero and one. The interpretation of the relative efficiency is that if N observations are used in design ξ^*, the same amount of information can be obtained from design ξ by taking $N/Eff_{\Theta(\xi)}$ observations. Of course, high efficiencies are preferable. In this section we discuss two types of robustness: robustness of optimal designs against an incorrectly specified intra-school correlation coefficient, and robustness of optimal designs against other optimality criteria.

2.2.1. Robustness against an incorrectly specified intra-school correlation coefficient

This is an essential consideration because the optimal sample sizes for most optimality criteria in Table 1 depend on the intra-school correlation coefficient ρ, which is generally

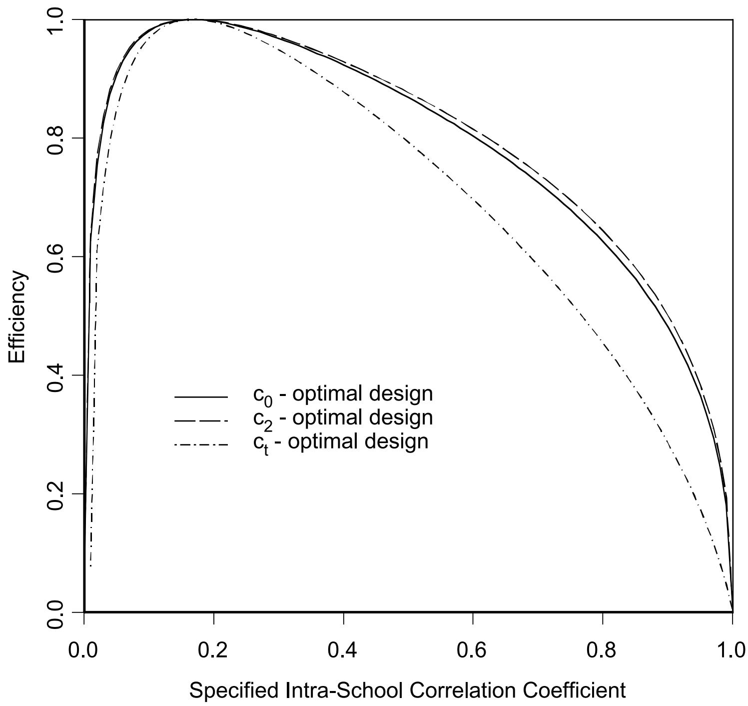

Fig. 2. Robustness of optimal designs for the "c_0, c_2, and c_t optimal" designs against an incorrectly specified intra-school correlation coefficient ρ when the true $\rho = 0.167$

not known in practice. Thus, a prior estimate obtained from a pre-study or a reasonable guess has to be specified in advance to calculate the optimal sample sizes. Figure 2 shows the relationship between the efficiency and the specified ρ for the "c_0, c_2, and c_t-optimal" designs, when the assumed parameter values are $C = 10{,}000$, $C_s = 30$, and $C_k = 1$, $\sigma^2 = 5$, and $\tau^2 = 1$. This implies that the true $\rho = 0.167$. Because the optimal sample sizes for criteria c_1 and c_s do not depend on ρ, the efficiency for the optimal designs for these two criteria is always equal to one. The optimal designs under the criteria c_0, c_2, and c_t have efficiencies equal to one if the specified ρ is equal to the true ρ. For these criteria the optimal designs are more sensitive to underspecification of the ρ than to overspecification, but another relationship may probably hold for other parameter values. Of these three criteria, the optimal design for criterion c_2 is less sensitive to incorrectly specified ρ, whereas the optimal design for criterion c_t is most sensitive.

2.2.2. Robustness of optimal designs under different optimality criteria

Figure 3 shows the c_1- and the c_2-efficiency of optimal designs for the other optimality criteria as a function of the intra-school correlation coefficient. The c_1-efficiency of c_s-optimal designs is always equal to one, because the optimal sample sizes for both optimality criteria are equal and independent of ρ. The c_1-efficiencies of c_0-, c_2-, and c_t-optimal designs, however, are very small because their optimal sample sizes are much different than those for a c_1-optimal design. For the same reason, the c_2-efficiency of c_1- and c_s-optimal designs is very low, whereas the c_2-efficiency of c_0- and c_t-optimal designs is quite high.

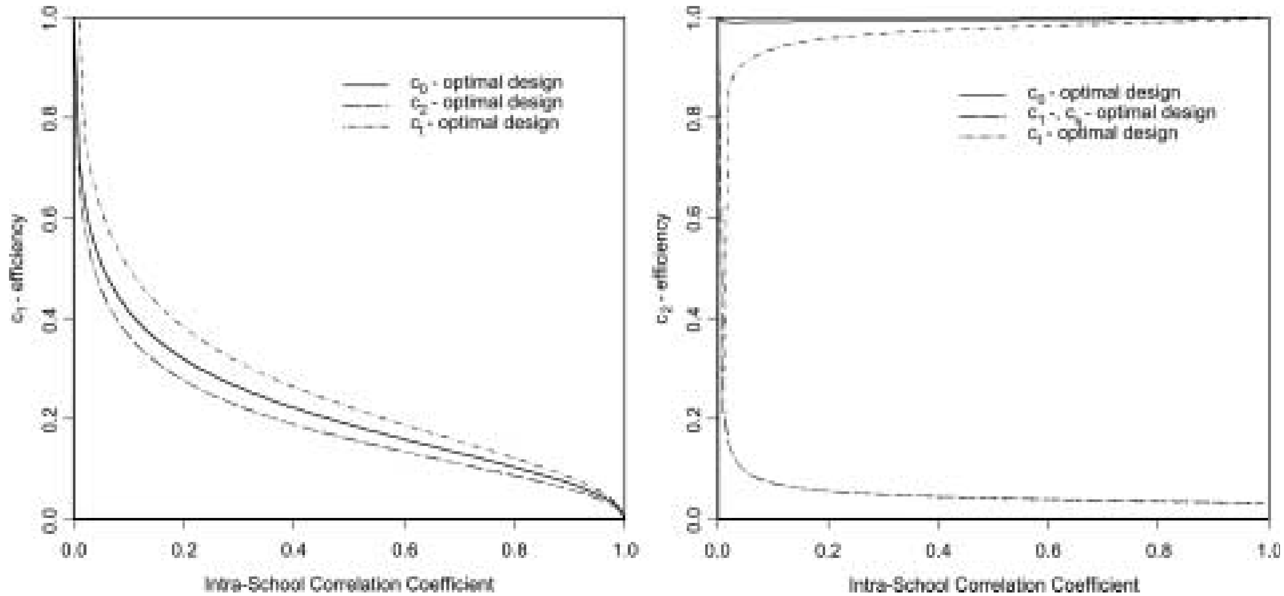

Fig. 3. Robustness of optimal designs against other optimality criteria. Left: c_1-efficiencies of c_0-, c_2-, and c_t-optimal designs. Right: c_2-efficiencies of c_0-, c_1-, c_t- and c_s-optimal designs

3. Two-Objective Optimal Designs

3.1. Constrained and compound optimal designs

When designing a study a researcher mostly has more than one objective in mind. In this section we will focus, for simplicity, on two-objective optimal designs. The construction of designs with two or more objectives is straightforward in principle. In principle we denote the two objectives by Θ_1 and Θ_2 and assume Θ_1 is the more important objective. The design sought is the one that does best under the objective Θ_2 subject to the constraint that this design has a Θ_1 value smaller than a user-specified constant c. Thus, the optimality criterion is

$$\text{minimize } \Theta_2(\xi) \text{ subject to } \Theta_1(\xi) \leq c \tag{5}$$

and is called a constrained optimal design. For convenience, the criterion (5) is often rewritten as

$$\text{minimize } \Theta_2(\xi) \text{ subject to } Eff_{\Theta_1(\xi)} \geq e \tag{6}$$

where $Eff_{\Theta_1(\xi)}$ is the design efficiency of ξ under Θ_1 alone. Thus, the less important objective is minimized under the restriction that the efficiency of the more important objective is at least equal to a user-selected constant e. Designs that satisfy (6) are constrained optimal designs and such a design is sometimes called a Θ_1-restricted Θ_2-optimal design. Because constrained optimal designs are difficult to find, one may want to construct a compound optimal design to minimize

$$\Theta(\xi|\lambda) = \lambda\Theta_1(\xi) + (1-\lambda)\Theta_2(\xi) \tag{7}$$

This optimization is easier to solve, either analytically or numerically. Cook and Wong (1994) showed that under convexity and differentiability constrained and compound optimal designs are equivalent. Thus, the desired constrained optimal design may be found by first forming a compound optimal design as a function of the weight λ in (7). The constrained optimal design is found by drawing an efficiency plot, in which the relation between λ and the efficiencies for both objectives are drawn, and selecting the λ for which (6) is satisfied and $Eff_{\Theta_2(\xi)}$ is maximized. The objectives Θ_1 and Θ_2 are often standardized (i.e., divided by their minimal possible values) so that the two components in (7) are of comparable magnitude.

An example. We use the same example described earlier to illustrate how the efficiency plot can help us find the constrained optimal design. In Figure 4 efficiency plots for a c_2-restricted c_0-optimal design and a c_2-restricted c_1-optimal design are given. The efficiencies for both objectives are high for a c_2-restricted c_0-optimal design, because the optimal sample sizes for these two objectives are almost similar (see Figure 1). Thus the criteria are not competitive in the sense that not much sacrifice in efficiency has to be accepted for a gain in the other criterion. On the other hand both efficiencies for a c_2-restricted c_1-optimal design are much lower because the optimal sample sizes are very different. These criteria are thus incompatible. The desired constrained optimal design as given in (6) can be found as follows. We draw a horizontal line at e as given in (6) to intersect the graph of $Eff_{\Theta_1(\xi)}$. Then a vertical line is drawn from the point of intersection to meet the λ-axis. The resulting value of λ is then used to find the compound optimal design.

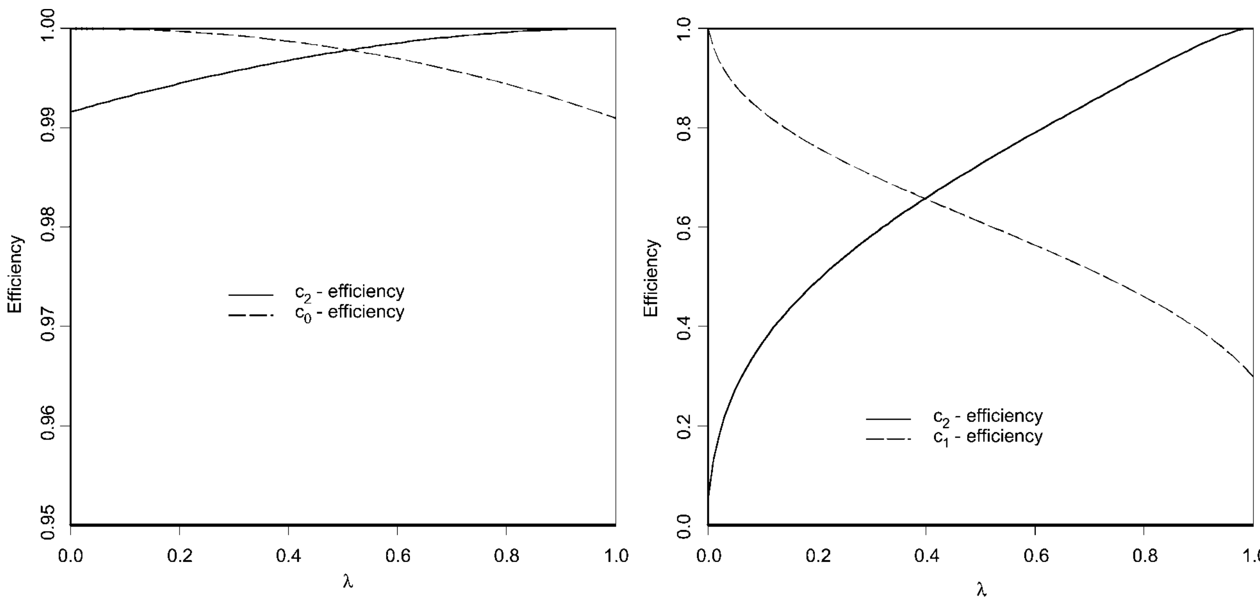

Fig. 4. *Efficiency plots for two-objective designs. Left: c_2-restricted c_0-optimal design. Right: c_2-restricted c_1-optimal design. Note that the ranges of the vertical axes in the two figures do not correspond*

We can draw some general conclusions from Figure 4. If $\lambda = 0$ the design is a Θ_2-optimal design and $Eff_{\Theta_2(\xi)} = 1$. On the other hand, if $\lambda = 1$ the design is Θ_1-optimal and $Eff_{\Theta_1(\xi)} = 1$. $Eff_{\Theta_1(\xi)}$ is non-decreasing if λ increases, and $Eff_{\Theta_2(\xi)}$ is non-increasing if λ increases. When the lines meet, the compound optimal design constructed for that point actually maximizes the minimum efficiencies under the two criteria (Imhof and Wong 2000). The two lines do not necessarily meet at the point $\lambda = 0.5$, as Figure 4 shows.

3.2. *Robustness of two-objective optimal designs*

As with single-objective optimal designs, the value of the intra-school correlation coefficient ρ needs to be specified to calculate a two-objective optimal design. The sensitivity issue of the two-objective optimal design to misspecification in the intra-school correlation coefficient can be studied in a similar manner. As an example we study the robustness of c_2-restricted c_0-optimal designs, and c_2-restricted c_1-optimal designs against an incorrectly specified ρ as a function of λ. As in Section 2.2.1., $C = 10{,}000$, $C_s = 30$, and $C_k = 1$, $\sigma^2 = 5$, and $\tau^2 = 1$, which implies that the true $\rho = 0.167$. The efficiencies of these designs are plotted in Figure 5 for $\rho = 0.125$, 0.083, and 0.042. Just like in single-objective optimal designs, the efficiency decreases if ρ departs from the true ρ. We note that the efficiency of c_2-restricted c_0-optimal designs increases slightly with λ. This is because the c_2-optimality criterion is a bit more robust to incorrect specification of ρ than the c_1-optimality criterion (see Figure 2) and a larger value of λ implies a higher weight is given to the c_2-optimality criterion in the compound criterion. If $\lambda = 0$ the c_2-restricted c_1-optimal design is a c_1-optimal design which does not depend on ρ, i.e., the efficiency is equal to 1. The efficiency decreases rapidly when λ increases. This is also obvious because a higher weight is given to the c_2-optimality criterion in the combination of the two objectives and the c_2-optimality criterion is not very robust to incorrectly specified ρ.

4. Discussion and Conclusions

We have constructed a two-objective optimal design for the hierarchical linear model using the model from Cohen (1998) as an illustrative example. Such designs may, of course, be constructed for any hierarchical linear models with either continuous or dichotomous outcomes and any number and type of explanatory variables. Starting points are the analytical formulae for single-objective optimal designs, which may be obtained either from the literature (see the papers referred to in the introduction of this article) or by derivation of the variances of maximum likelihood estimators. Once the relative importance of the two objectives has been chosen the constrained optimal designs may be derived from the compound optimal design, either analytically or numerically. Most designs will depend on the intra-school correlation coefficient and so a reasonable prior guess must be given. The robustness of the design against misspecification of this intra-school correlation coefficient may be studied as shown in Section 3.2.

In this article we have focused on two-objective optimal designs but an extension to more than two objectives is possible in principle. The problem is more complicated

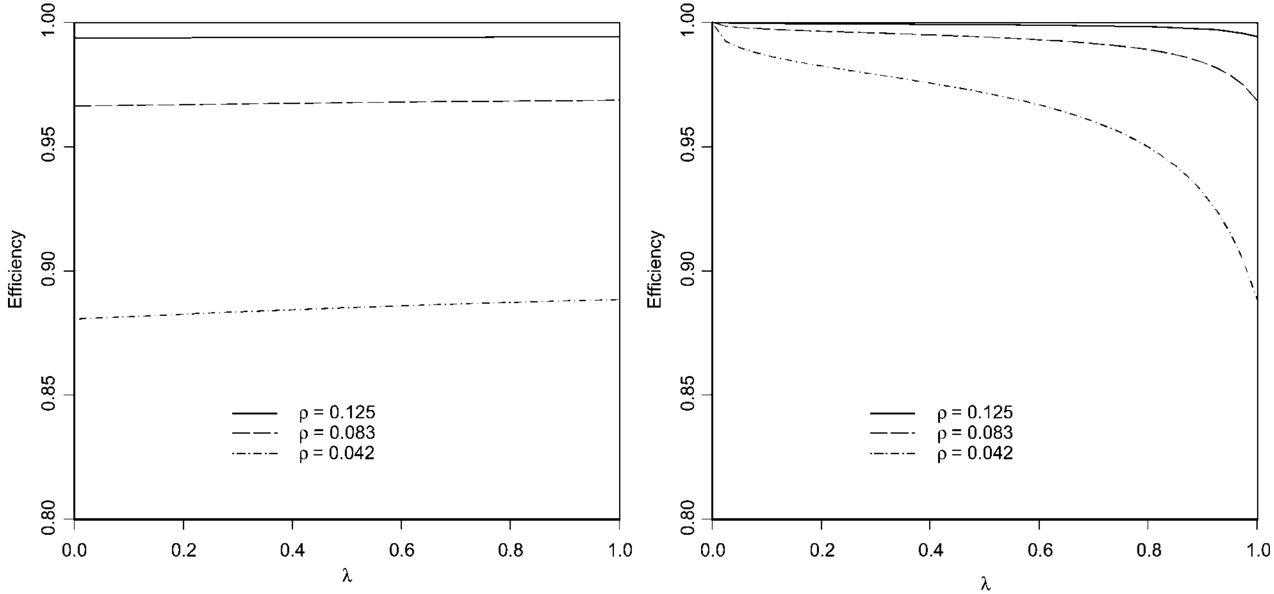

Fig. 5. *Robustness of two-objective optimal designs against an incorrectly specified* λ *when the true* $\rho = 0.167$. *Left:* c_2*-restricted* c_0*-optimal design. Right:* c_2*-restricted* c_1*-optimal design*

because the efficiency plots are multidimensional and thereby less easily interpretable. An alternative is to adopt a sequential approach as was done by Huang and Wong (1998b). Multiple-objective optimal designs are an improvement of the single-objective optimal designs because they account for the multiple objectives in the study. In addition, the formulation reflects more adequately what is needed in reality. Therefore they should be used more often in practice. It should, however, be noted that there is no guarantee that the optimal design always exists. If high efficiencies are required for all or most of the criteria, then the constrained optimal design may not exist.

5. References

Afshartous, D. (1995). Determination of Sample Size for Multilevel Model Design. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Bryk, A.S. and Raudenbush, S.W. (1992). Hierarchical Linear Models. Newbury Park: Sage Publications.

Cohen, M.P. (1998). Determining Sample Sizes for Surveys with Data Analyzed by Hierarchical Linear Models. Journal of Official Statistics, 14, 267–275.

Cook, R.D. and Wong, W.K. (1994). On the Equivalence of Constrained and Compound Optimal Designs. Journal of the American Statistical Association, 89, 687–692.

Donner, A. (1998). Some Aspects of the Design and Analysis of Cluster Randomization Trials. Applied Statistics, 47, 95–113.

Feng, Z. and Grizzle, J.E. (1992). Correlated Binomial Variates: Properties of Estimator of Intraclass Correlation and Its Effect on Sample Size Calculation. Statistics in Medicine, 11, 1607–1614.

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

Hox, J.J. (1994). Applied Multilevel Analysis. Amsterdam: TT-Publikaties.

Hsieh, F.Y. (1988). Sample Size Formulae for Intervention Studies with the Cluster as Unit of Randomization. Statistics in Medicine, 8, 1195–1201.

Huang, Y.-C. (1996). Multiple-objective Optimal Designs. Unpublished PhD-thesis. University of California at Los Angeles.

Huang, Y.-C. and Wong, W.-K. (1998a). Multiple-objective Optimal Designs. Journal of Biopharmaceutical Studies, 8, 635–643.

Huang, Y.-C. and Wong, W.-K. (1998b). Sequential Construction of Multiple-objective Optimal Designs. Biometrics, 54, 1388–1397.

Imhof, L. and Wong, W.K. (2000). A Graphical Method for Finding Maximin Efficiency Designs. Biometrics, 56, 113–117.

Kreft, I. and de Leeuw, J. (1998). Introducing Multilevel Modeling. London: Sage Publications.

Lee, E.W. and Dubin, N. (1994). Estimation and Sample Size Considerations for Clustered Binary Data. Statistics in Medicine, 13, 1241–1252.

Liu, G. and Liang, K.-Y. (1997). Sample Size Calculations for Studies with Correlated Observations. Biometrics, 53, 937–947.

Longford, N.T. (1995). Random Coefficient Models. Oxford: Clarendon Press.

Moerbeek, M., Van Breukelen, G.J.P., and Berger, M.P.F. (2000). Design Issues for

Experiments in Multilevel Populations. Journal of Educational and Behavioral Statistics, 25, 271–284.
Moerbeek, M., Van Breukelen, G.J.P., and Berger, M.P.F. (2001a). Optimal Experimental Designs for Multilevel Logistic Models. The Statistician, 50, 17–30.
Moerbeek, M., Van Breukelen, G.J.P., and Berger, M.P.F. (2001b). Optimal Experimental Designs for Multilevel Models with Covariates. Communications in Statistics – Theory and Methods, 30, 2683–2698.
Mok, M. (1995). Sample Size Requirements for 2-level Designs in Educational Research. Multilevel Modeling Newsletter, 7, 11–15.
Raudenbush, S.W. (1997). Statistical Analysis and Optimal Design for Cluster Randomized Trials. Psychological Methods, 2, 173–185.
Shih, W.J. (1997). Sample Size and Power Calculations for Periodontal and Other Studies with Clustered Samples Using the Method of Generalized Estimating Equations. Biometrical Journal, 39, 899–908.
Snijders, T.A.B. and Bosker, R.J. (1993). Standard Errors and Sample Sizes for Two-level Research. Journal of Educational Statistics, 18, 237–259.
Snijders, T.A.B. and Bosker, R.J. (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. London: Sage Publications.
Zhu, W. and Wong, W.K. (2000a). Optimum Treatment Allocation for Dual-objective Clinical Trials with Binary Outcomes. Communiations in Statistics – Theory and Methods, 29, 957–974.
Zhu, W. and Wong, W.K. (2000b). Optimal Treatment Allocation in Comparative Biomedical Studies. Statistics in Medicine, 19, 639–648.

Received April 2001
Revised March 2002