Statistica Neerlandica (2004) Vol. 58, nr. 1, pp. 111–124

The effectiveness of randomized complete block design

Gwowen Shieh*

Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan 30050, R.O.C.

Show-Li Jan†

Department of Applied Mathematics, Chung Yuan Christian University, Chungli, Taiwan 32023, R.O.C.

This paper examines the relative efficiency of randomized complete block design as compared with that of completely randomized design. The most widely used measure of relative efficiency considers only the error variances of the two designs, therefore it does not provide the complete information concerning the sensitivity of the experiment in the final results. We study three alternative criteria related to the sensitivity issue and design planning consideration. The proposed relative measures employ the *p*-value, Scheffé confidence interval estimation and power of both designs. The distinct feature of this study is the focus on the estimated relative efficiency measures and their relation with the coefficient of partial determination between responses and block effects, given that treatment effects are present in the models. Furthermore, informative visual representations and numerical assessments of various aspects of their properties are also presented.

Key Words and Phrases: Completely randomized design, observed significance level, paired-sample *t* test, power, relative efficiency, Scheffé confidence interval, two-sample *t* test.

1 Introduction

The most basic type of statistical design for making inferences about treatment means is the completely randomized design (CRD), where all treatments under investigation are randomly allocated to the experimental units. The CRD is appropriate for testing the equality of treatment effects when the experimental units are relative homogeneous with respect to the response variable. When the experimental units are heterogeneous, the notion of blocking is used to control the extraneous sources of variability. The major criteria of blocking are characteristics

*gwshieh@mail.nctu.edu.tw

†sljan@math.cycu.edu.tw

associated with the experimental material and the experimental setting. The purpose of blocking is to sort experimental units into blocks, so that the variation within a block is minimized while the variation among blocks is maximized. An effective blocking not only yields more precise results than an experimental design of comparable size without blocking, but also increases the range of validity of the experimental results. One can use a randomized complete block design (RCBD) to compare treatment means when there is an extraneous source of variability. In such cases, treatments are randomly assigned to experimental units within a block, with each treatment appearing exactly once in every block. As planning and conducting an experiment with the RCBD requires extra effort relative to the CRD, a natural question of interest is how well the blocking has worked or how much has been saved by using an RCBD rather than a CRD with the same number of experimental units. The answer helps to justify the effectiveness of blocking in the experiment being conducted and is also useful for future studies using the same or similar experimental units. It is well recognized that the gain from using an RCBD instead of a CRD is a reduction in error variance, while the loss is a decrease in error degrees of freedom. The latter results in a decrease in the sensitivity or, more exactly, the probability of obtaining a significant result with respect to treatment comparisons, therefore, it may not be advantageous to conduct an experiment in RCBD. The most widely accepted criterion for performing this evaluation is the measure of relative precision in terms of the ratio of error variance estimates associated with the two designs. However, many standard textbooks fail to address the fundamental discrepancy between the concepts of sensitivity and precision. Consequently, the use of relative precision does not resolve the issue of sensitivity. As sensitivity is essential to the results of analyses, it is open to question that the relative precision is most useful for evaluating the relative efficiency of RCBD to CRD. To compare the sensitivity of two experiments, the notions of the width of confidence interval and power were proposed in COCHRAN and COX (1957, p. 32). Despite the great applicability in assessing the relative efficiency for more general setup, their study is limited to the case of comparing the difference between the effects of two treatments along with some simplified assumptions. It should be informative to extend the paradigm to the general situation of testing the difference of several treatment effects. In summary, previous work did not provide specific guidance to the construction of efficiency measures regarding sensitivity in the context of RCBD relative to CRD.

This article aims to investigate the effectiveness of blocking in RCBD through three relative measures. They are the relative efficiency of RCBD as compared with that of CRD evaluated in terms of the observed significance level (p-value), the squared half width of the Scheffé confidence interval and the power of detecting treatment effects. The distinct feature of this study is the focus on the estimates of the relative efficiency measures and their relation to the coefficient of partial determination between responses and block effects, given that treatment effects are present in the models. The similarities and differences among the proposed relative efficiency measures and the relative precision are described in an attempt to provide some guidance in the choice of appropriate measure of effective blocking in RCBD. In order to enhance the clarity and usefulness of the findings, both numerical illustration and visual supplement are provided. For more complex models, MORRISON (1972) and VONESH (1983) employed the notion of expected squared half width of Scheffé confidence interval for comparing the sensitivity, while JENSEN (1980, 1982) conducted the asymptotic efficiency comparisons with the ratio of noncentrality parameters in the context of repeated measures designs.

Section 2 specifies the model for RCBD and gives a review of the properties of the usual relative precision estimate that motivates this study. In Section 3, the results for the relative efficiency of RCBD to CRD obtained by the proposed measures are presented. The special case of two treatments is also discussed and this situation essentially reduces to the classical phenomenon of paired-sample t test versus two-sample t test. As suggested by a referee, we also present the multiple-objective approach to the evaluation of efficiency. Section 4 contains a brief summary.

2 The model and motivation

Consider the standard linear model for an RCBD with both the block and treatment effects fixed and without interaction effects as follows:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

where Y_{ij} denotes the response for the experimental unit with the ith treatment in the jth block, μ is the overall mean, τ_i is the treatment effect, β_j is the block effect, and ε_{ij} is the error with $i = 1,\ldots,t$ and $j = 1,\ldots,b$. As usual, the treatment and block effects are subject to the restrictions that $\sum_{i=1}^{t} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$, respectively, and the errors are assumed to be independent $N(0, \sigma^2_{\mathrm{RCBD}})$. With the usual notation of an overbar and a dot that denote averaging over a subscript, the analysis of variance for the RCBD is presented in Table 1. Once an experiment has been conducted in an RCBD, it is natural to question whether the RCBD provided a better comparison for the treatments than the CRD with the same number of experimental units. To facilitate the presentation, it is convenient to define the statistical model for CRD with the same treatment effects and sample size as

Table 1. Analysis of variance table for RCBD and CRD.

Source	*SS*	*df*	*MS*
RCBD			
Treatment	$SSTR = b\sum_{i=1}^{t}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$t - 1$	$MSTR = \frac{SSTR}{t-1}$
Block	$SSBL = t\sum_{j=1}^{b}(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$	$b - 1$	$MSBL = \frac{SSBL}{b-1}$
Error	$SSE = \sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2$	$(t - 1)(b - 1)$	$MSE = \frac{SSE}{(t-1)(b-1)}$
CRD			
Treatment	$SSBT = b\sum_{i=1}^{t}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$t - 1$	$MSBT = \frac{SSBT}{t-1}$
Error	$SSWT = \sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$t(b - 1)$	$MSWT = \frac{SSWT}{t(b-1)}$

$$Y_{ij} = \mu^* + \tau_i + e_{ij},$$

where Y_{ij} denotes the response for the jth experimental unit with the ith treatment, μ^* is the overall mean, τ_i is the fixed treatment effect with $\sum_{i=1}^{t} \tau_i = 0$ as above, and the error e_{ij} is independently distributed as $N(0, \sigma^2_{\text{CRD}})$ for $i = 1,\ldots,t$ and $j = 1,\ldots,b$. The partition of the between- and within-treatment sum of squares and corresponding degrees of freedom for CRD is also listed in Table 1. Under the null hypothesis of no treatment effects, the test statistics are $F_{\text{RCBD}} = MSTR/MSE \sim F^{t-1}_{(t-1)(b-1)}$ and $F_{\text{CRD}} = MSBT/MSWT \sim F^{t-1}_{t(b-1)}$ for RCBD and CRD, respectively, where F^{df1}_{df2} represents an F distribution with $df1$ and $df2$ degrees of freedom. However, it should be noted that the CRD has never been conducted. The mean sum of squares $MSBT$ and $MSWT$ must be estimated from $MSTR$, $MSBL$ and MSE for the RCBD. An unbiased estimator of σ^2_{CRD} is (see, for example, COCHRAN and COX, 1957, p. 112; HINKELMANN and KEMPTHORNE, 1994, p. 261)

$$MSWT^* = \frac{b(t-1)MSE + (b-1)MSBL}{(bt-1)}. \tag{1}$$

Since $E(MSTR) = \sigma^2_{\text{RCBD}} + b\sum_{i=1}^{t} \tau_i^2/(t-1)$ and $E(MSE) = \sigma^2_{\text{RCBD}}$, one can proceed to estimate $E(MSBT) = \sigma^2_{\text{CRD}} + b\sum_{i=1}^{t} \tau_i^2/(t-1)$ with the unbiased estimator

$$MSBT^* = MSWT^* + MSTR - MSE. \tag{2}$$

The success of blocking is best measured by the relative efficiency of the RCBD as compared with that of the CRD. In general, the relative efficiency is a positive number that can be interpreted as the ratio by which the sample size of the CRD would have to be in order to achieve the same efficiency as that of the RCBD. The most widely used measure of relative efficiency is the relative precision defined as follows:

$$RE = \frac{\sigma^2_{\text{CRD}}}{\sigma^2_{\text{RCBD}}}. \tag{3}$$

The practical interpretation of RE is that $b{\cdot}RE$ replications per treatment are required for a CRD to attain the same degree of *precision* associated with the error variance as the RCBD with b blocks. A common estimate of RE is given by

$$ERE = \frac{MSWT^*}{MSE} = \frac{b(t-1)MSE + (b-1)MSBL}{(bt-1)MSE}. \tag{4}$$

It is easy to show that there exists a monotonic relationship between ERE and $H = MSBL/MSE$:

$$ERE = k + (1-k)H,$$

where $k = b(t-1)/(bt-1)$. Although H mimics the form of an F statistic, LENTNER *et al.* (1989) showed that it is not suitable for testing the blocking effects because of the manner in which randomization is conducted in the RCBD. This is often confused with testing the main effects in the two-factor model without interaction where each treatment combination is assigned at random to one experimental unit. The major difference revolves around the different randomization process, that is, restricted versus unrestricted randomization. The issue of conflicting results about the legitimacy of testing block effects is beyond the scope of this article, readers can refer to SAMUELS *et al.* (1991) and the discussion of their article for further details.

The measure *ERE* is appealing for providing a quick index for effective blocking. However, the value of *ERE* does not provide the complete information because it considers only the experimental errors of two designs. Note that the degrees of freedom for experimental error of an RCBD are not as great as those of a CRD, that is, $(t-1)(b-1)$ versus $t(b-1)$. Consequently, it is harder to reject the null hypothesis resulting in a loss of power. Therefore, the reduction in error may, however, be offset by the loss of $b-1$ degrees of freedom. This phenomenon is not properly shown by the *ERE*. As a concrete example, consider the following special case. The value $ERE = 1$ indicates that CRD is as efficient as RCBD when the sample sizes are identical. With the results in (1), (2) and (4), an expression imitating the form of F_{CRD} statistic is given by

$$F^*_{\mathrm{CRD}} = \frac{MSBT^*}{MSWT^*} = \frac{F_{\mathrm{RCBD}} + ERE - 1}{ERE}. \tag{5}$$

Thus, according to $ERE = 1$, it readily follows from (5) that the value F^*_{CRD} of the statistic F_{CRD} is identical to that of F_{RCBD}. Due to the loss in degrees of freedom for estimated error variance in RCBD, the observed significance level or p-value is decreased, and therefore RCBD is less sensitive than CRD. It is conceivable that the concepts of sensitivity and precision are fundamentally different, thus leading to different conclusions. However, it is questionable whether the *ERE* is better suited for evaluating the relative efficiency. Nevertheless, in order to take account of the different degrees of freedom in error variances, it is suggested in the literature, see COCHRAN and COX (1957, p. 34) and NETER *et al.* (1996, p. 1090), that the *ERE* can be adjusted by a factor denoted by *EREM*,

$$EREM = m \cdot ERE, \tag{6}$$

where $m = \dfrac{t(b-1)+3}{t(b-1)+1} \cdot \dfrac{(t-1)(b-1)+1}{(t-1)(b-1)+3}$.

The modification has little effect with moderate-sized degrees of freedom for experimental error variance estimates. More importantly, LENTNER *et al.* (1989) commented that the judgment of significant gains or losses due to blocking is subjective and must be considered from a practical standpoint. As mentioned above, *RE* has the simple interpretation in terms of relative sample sizes of two designs. In a certain sense, however, it may not be useful from the design planning point of view because the standard methods for sample size determination are more involved than just the error variance. In the next section, three alternative efficiency measures based on the notion of *p*-value, confidence interval and power will be considered. All these three measures evaluate different aspects of sensitivity of the experiments.

3 The relative efficiency of RCBD to CRD

In this section, the *p*-value, estimation and power approaches are described for evaluating the relative efficiency of RCBD as compared with that of CRD. The *p*-value approach takes up the observed significance level of the *F* test statistic for treatment effects. The estimation approach employs the squared half width of confidence intervals of the Scheffé simultaneous estimation procedure, while the power approach calculates the probability of rejecting given real treatment differences. As presented in NETER *et al.* (1996, Sections 26.4 and 26.5), the power and Scheffé confidence interval procedures are the major methods for planning the sample sizes of RCBD. Therefore, the appealing attribute of the last two measures is that they conform to the practical interpretation in terms of sample size of the relative efficiency described previously.

To lay the basis for developing a simplified view and providing a concise visualization of the relation of relative efficiency measures, the following ratio is defined as

$$\Gamma = \frac{SSBL}{SSBL + SSE}.$$

Three important features of the measure Γ deserve attention. First, it is a one-to-one function of H. Second, it is within the range of [0, 1]; and lastly, it can be viewed as the coefficient of partial determination between responses and block effects, given that treatment effects are present in the models. For a meaningful discussion, it is assumed that the term $E(MSBL)$ is finite for all b. Furthermore, it can be readily established from the aforementioned definition in (4) that

$$ERE = k + (1-k)\frac{(t-1)\Gamma}{1-\Gamma},$$

which is a weighted average of 1 and $(t-1)\Gamma/(1-\Gamma)$ for $0 < k = b(t-1)/b(t-1) < 1$. Therefore,

$$ERE < 1 \text{ if and only if } \Gamma < 1/t,$$
$$ERE = 1 \text{ if and only if } \Gamma = 1/t,$$
$$ERE > 1 \text{ if and only if } \Gamma > 1/t.$$

The relation between ERE and Γ serves as a basis for the following comparison of relative measures. In general, all of the estimated relative efficiency measures are increasing with Γ when all other factors are fixed. Furthermore, it is found from the numerical assessment that the differences between ERE and $EREM$ are marginal for small and moderate values of Γ regardless of the numbers of treatments and blocks. However, their differences can be sizeable when Γ is close to 1.

3.1 *The p-value approach*

Let $p_{\mathrm{RCBD}} = P\{F^{t-1}_{(t-1)(b-1)} > E(F_{\mathrm{RCBD}})\}$ and $p_{\mathrm{CRD}} = P\{F^{t-1}_{t(b-1)} > E(F_{\mathrm{CRD}})\}$ denote the p-value associated with the F statistic at the value of $E(F_{\mathrm{RCBD}})$ and $E(F_{\mathrm{CRD}})$ for RCBD and CRD, respectively. The following measure evaluates the relative efficiency of RCBD to CRD in terms of the p-value for the expected value of the F statistic:

$$REV = \frac{p_{\mathrm{CRD}}}{p_{\mathrm{RCBD}}}.$$

When F_{RCBD} is available, one can estimate the REV with $EREV$ defined as

$$EREV = \frac{\hat{p}_{\mathrm{CRD}}}{\hat{p}_{\mathrm{RCBD}}}, \tag{7}$$

where $\hat{p}_{\mathrm{RCBD}} = P\{F^{t-1}_{(t-1)(b-1)} > F_{\mathrm{RCBD}}\}$ is the p-value of F_{RCBD} and $\hat{p}_{\mathrm{CRD}} = P\{F^{t-1}_{t(b-1)} > F^{*}_{\mathrm{CRD}}\}$ denotes the p-value for F_{CRD} evaluated at the value of F^{*}_{CRD} as defined in (5). Note that F_{CRD} cannot be observed in the setting of an RCBD. As a visual supplement, Figure 1 presents a plot of $EREV$ against Γ for

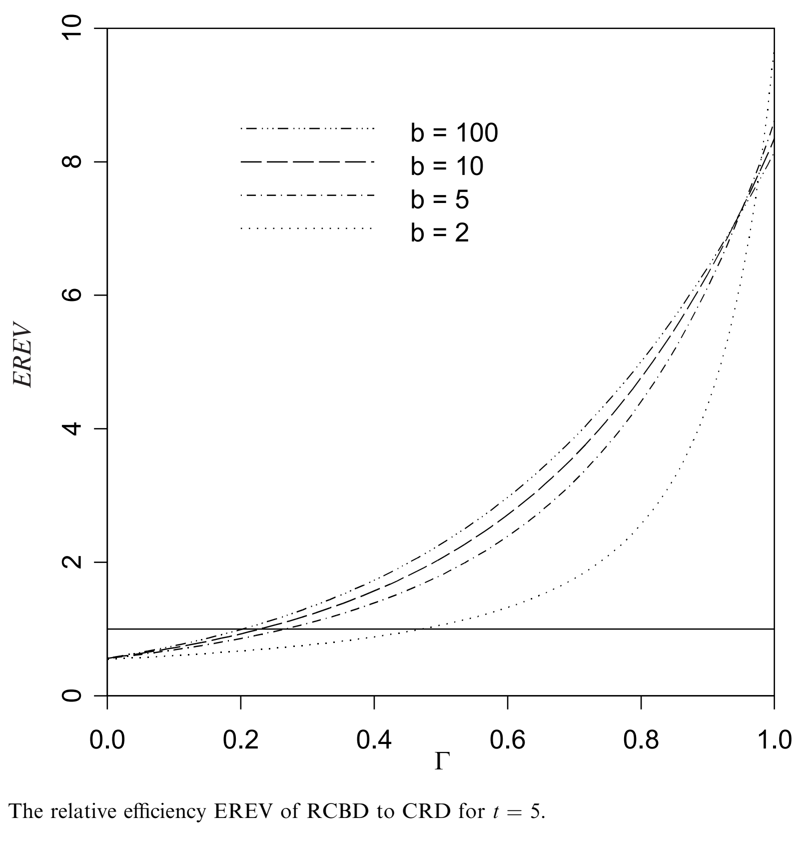

Fig. 1. The relative efficiency EREV of RCBD to CRD for $t = 5$.

$t = 5$ and four different values $b = 2$, 5, 10 and 100 by setting the p-value of F_{RCBD} at $\hat{p}_{\mathrm{RCBD}} = \alpha = 0.05$.

3.2 *The estimation approach*

In order to encompass all possible contrasts among the treatment effects and control the overall family confidence level, the estimation approach compares the expected squared half widths of Scheffé confidence intervals for both designs. Assume the family of interest is the set of all possible contrasts among treatment effects of the form:

$$L = \sum_{i=1}^{t} c_i \tau_i, \quad \text{where } \sum_{i=1}^{t} c_i = 0.$$

It follows that an unbiased estimator of L is $\hat{L} = \sum_{i=1}^{t} c_i \bar{Y}_i$, and the expected squared half widths of Scheffé confidence limits for L can be written as

$$S^2_{\mathrm{RCBD}} = (t-1)F[t-1,(t-1)(b-1),\alpha]\frac{\sigma^2_{\mathrm{RCBD}}}{b}\sum_{i=1}^{t} c_i^2$$

and

$$S^2_{\mathrm{CRD}} = (t-1)F[t-1,t(b-1),\alpha]\frac{\sigma^2_{\mathrm{CRD}}}{b}\sum_{i=1}^{t} c_i^2$$

for RCBD and CRD, respectively, where $F[df1, df2,\alpha]$ denotes the $(1-\alpha)$th percentile of the F distribution with $df1$ and $df2$ degrees of freedom. The relative efficiency REW in terms of the expected squared half width of the Scheffé confidence interval is defined as

$$REW = \frac{S^2_{\mathrm{CRD}}}{S^2_{\mathrm{RCBD}}} = w \cdot RE,$$

where $w = F[t-1, t(b-1), \alpha]/F[t-1, (t-1)(b-1), \alpha]$ and RE is defined in (3). Along the same line of estimation, a natural estimator of REW is

$$EREW = w \cdot ERE, \tag{8}$$

where ERE is given in (4). It is worth noting that

$$F[t-1,t(b-1),\alpha] < F[t-1,(t-1)(b-1),\alpha] \quad \text{for all } t \text{ and } b.$$

Therefore, $w < 1$ and $EREW$ provides another modification of the ERE for deflating the overstatement of ERE as the adjustment $EREM$ described in (6). Figure 2 shows the plot of $EREW$ against Γ for $t = 5$ and four different values $b = 2$, 5, 10 and 100 with $\alpha = 0.05$.

3.3 *The power approach*

Let $F^{df1}_{df2}(\lambda)$ denote a noncentral F distribution with noncentrality λ and degrees of freedom $df1$ and $df2$. It follows from the model formulation that

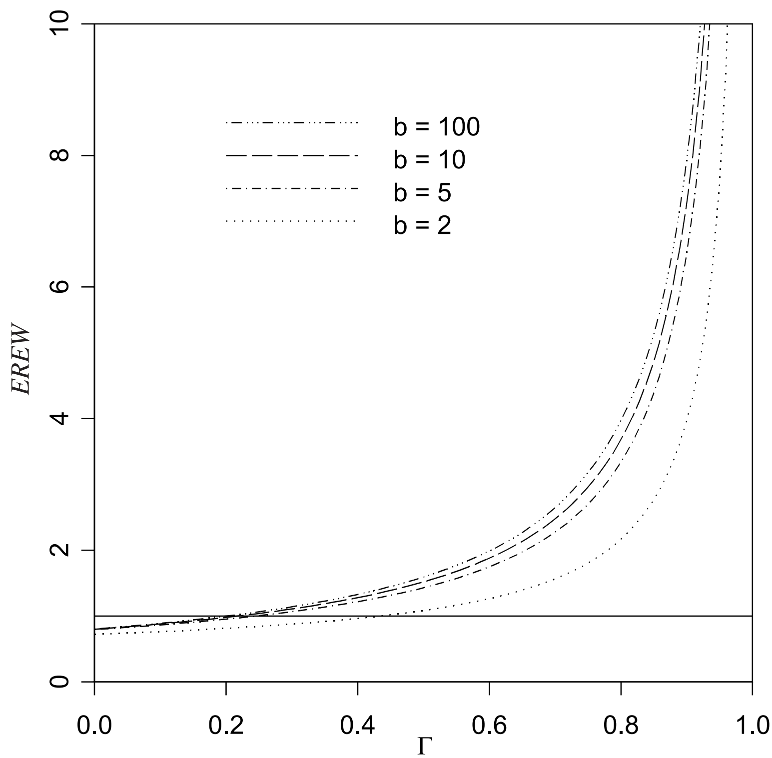

Fig. 2. The relative efficiency EREW of RCBD to CRD for $t = 5$.

$$F_{\text{RCBD}} \sim F^{t-1}_{(t-1)(b-1)}(\delta_{\text{RCBD}}) \quad \text{and} \quad F_{\text{CRD}} \sim F^{t-1}_{t(b-1)}(\delta_{\text{CRD}}),$$

where $\delta_{\text{RCBD}} = b\sum_{i=1}^{t} \tau_i^2/(2\sigma^2_{\text{RCBD}})$ and $\delta_{\text{CRD}} = b\sum_{i=1}^{t} \tau_i^2/(2\sigma^2_{\text{CRD}})$. The power-based relative efficiency measure of RCBD to CRD is

$$REP = \frac{P_{\text{RCBD}}\{\delta_{\text{RCBD}}\}}{P_{\text{CRD}}\{\delta_{\text{CRD}}\}},$$

where $P_{\text{RCBD}}\{\delta_{\text{RCBD}}\} = P\{F^{t-1}_{(t-1)(b-1)}(\delta_{\text{RCBD}}) > F[t-1, (t-1)(b-1), \alpha]\}$ and $P_{\text{CRD}}\{\delta_{\text{CRD}}\} = P\{F^{\,t-1}_{t(b-1)}(\delta_{\text{CRD}}) > F[t-1, t(b-1), \alpha]\}$ are the associated power of F_{RCBD} and F_{CRD} respectively, at the significance level of α. It is obvious that $\delta_{\text{CRD}} = \delta_{\text{RCBD}}/RE$ according to the definitions given above. Consequently, an estimate of δ_{CRD} can be obtained from the estimate of δ_{RCBD} with the multiplier 1/*ERE*. Thus, a practical estimate of *REP* is given by

$$EREP = \frac{P_{\text{RCBD}}\{\hat{\delta}_{\text{RCBD}}\}}{P_{\text{CRD}}\{\hat{\delta}_{\text{CRD}}\}}, \tag{9}$$

where $\hat{\delta}_{\text{RCBD}} = (t-1)\{[(t-1)(b-1)-2]F_{\text{RCBD}}/(t-1)(b-1)-1\}/2$ represents an unbiased estimate of the noncentrality δ_{RCBD} and $\hat{\delta}_{\text{CRD}} = \hat{\delta}_{\text{RCBD}}/ERE$. To

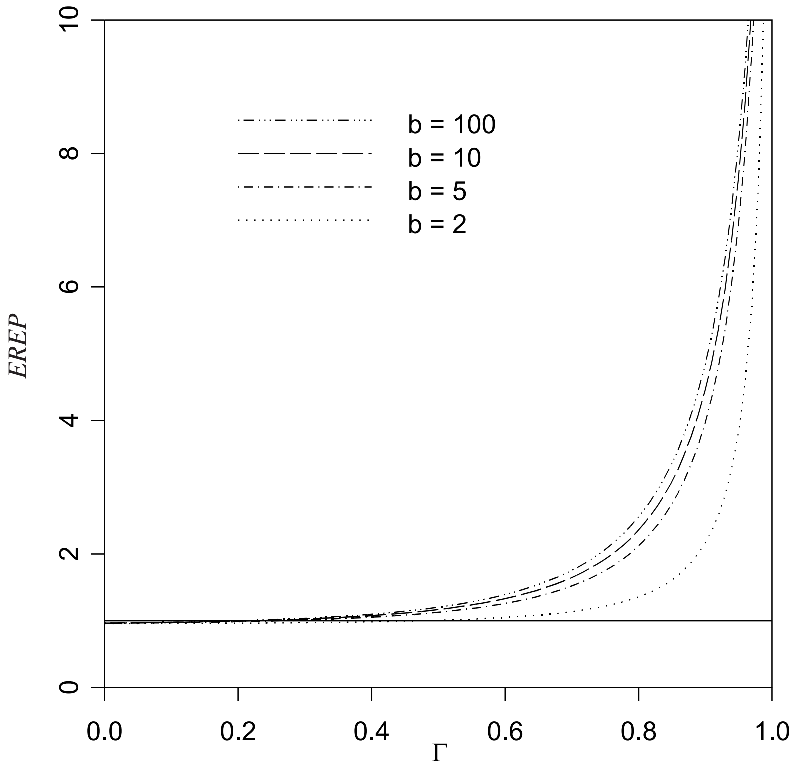

Fig. 3. The relative efficiency EREP of RCBD to CRD for $t = 5$.

visualize the magnitude of *EREP*, Figure 3 shows the plot of *EREP* against Γ for $t = 5$ and four different values $b = 2$, 5, 10 and 100 with $\alpha = 0.05$ and $\hat{\delta}_{\text{RCBD}}$ is chosen such that $P_{\text{RCBD}}\{\hat{\delta}_{\text{RCBD}}\} = 1 - \beta$ with $\beta = 0.05$.

3.4 *The t = 2 case*

For the comparison between two treatments $t = 2$, it is well known that the F test statistics for RCBD and CRD reduce to the squares of paired-sample and two-sample t statistic, respectively. As compared with the previous case of $t = 5$, each estimated relative efficiency measure decreases consistently for given values of b and Γ.

In this case, a useful expression for Γ is to rewrite it as $\Gamma = (1 + r{\cdot}g)/2$, where $r = S_{12}/(S_1S_2)$ and $g = (S_1S_2)/[(S_1^2 + S_2^2)/2]$ with

$$S_{12} = \sum_{j=1}^{b}(Y_{1j} - \bar{Y}_{1\cdot})(Y_{2j} - \bar{Y}_{2\cdot}) \quad \text{and} \quad S_i^2 = \sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i\cdot})^2, \quad i = 1 \text{ and } 2.$$

Note that r is the usual Pearson correlation coefficient between Y_{1j} and Y_{2j} and g is the ratio of geometric mean and arithmetic mean for S_1^2 and S_2^2, therefore, $-1 \leqslant r \leqslant 1$ and $0 < g \leqslant 1$. Consequently, r represents a simple efficiency measure of matching or blocking in a paired-sample experiment. When Y_{1j} and Y_{2j} are negatively correlated or uncorrelated with $r \leqslant 0$, which corresponds to $\Gamma \leqslant 1/2$, the matching is always ineffective for all four estimated relative efficiency measures and the design should be reconstructed. While Y_{1j} and Y_{2j} are positively correlated with $r > 0$ or $\Gamma > 1/2$, there are situations for the blocking to be effective.

3.5 *Cross-examination*

To examine further the similarities and differences between the estimated relative efficiency measures, the values of five estimated measures are plotted against Γ with $b = 5$ and $\alpha = \beta = 0.05$ in Figure 4 for $t = 5$. The three measures *ERE*, *EREM* and *EREW* are very much alike in the case of $t = 5$. Overall, their magnitudes are in the order of $ERE > EREM > EREW$ for the same b and Γ. However, the other two measures *EREV* and *EREP* behave substantially differently.

Instead of presenting extensive numerical tables for different combinations of t and b, Table 2 lists the minimum of b, denoted by b^*, such that the estimated measure is greater than or equal to 1 for $t = 5$, respectively, under different configurations of Γ, α and β. According to the definitions of *EREM*, *EREV*, *EREW* and *EREP* given earlier in (6)–(9), they are all less than 1 for $\Gamma \leqslant 1/t$, regardless of the value of b. Therefore, the number b^* is not available and is denoted by NA in the tables. For an RCBD experiment with $b \geqslant b^*$ at the specified Γ, the relative efficiency estimate is

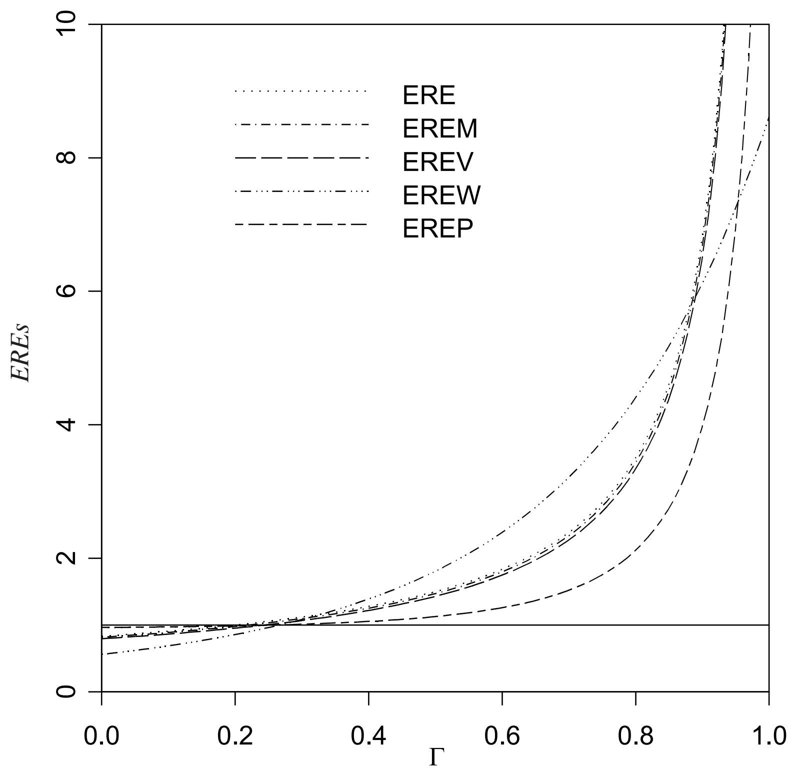

Fig. 4. The relative efficiency of RCBD to CRD for $t = 5$ and $b = 5$.

Table 2. The minimum number of blocks b^* for effective blocking when $t = 5$.

	Γ						
	$\leqslant 0.2$	0.225	0.250	0.275	0.3	0.4	0.5
ERE[1]	NA	2	2	2	2	2	2
EREM	NA	5	3	2	2	2	2
EREV ($\alpha = 0.05$)	NA	12	7	5	4	3	2
EREW ($\alpha = 0.05$)	NA	8	5	4	4	3	2
EREP ($\alpha = 0.05$, $\beta = 0.05$)	NA	10	6	5	4	3	2
EREP ($\alpha = 0.05$, $\beta = 0.10$)	NA	10	6	5	4	3	2
EREV ($\alpha = 0.10$)	NA	11	6	5	4	3	2
EREW ($\alpha = 0.10$)	NA	7	4	4	3	2	2
EREP ($\alpha = 0.10$, $\beta = 0.05$)	NA	8	5	4	4	3	2
EREP ($\alpha = 0.10$, $\beta = 0.10$)	NA	8	5	4	3	3	2

[1] $b^* = 2$ if $\Gamma = 0.2$.

larger than 1, leading to the conclusion that the blocking is effective with respect to the chosen relative efficiency measure. Therefore, the values of b^* in these tables not only can be utilized to tell whether the blocking is effective, but also provide a comparison of the required efforts in terms of the magnitude of b for the relative efficiency measures. It is important to note that $b^* = 2$ for *ERE* throughout the table for $\Gamma > 1/t$. It implies that as long as $\Gamma > 1/t$, the blocking is effective according to the estimated measure *ERE*. The results for *EREW* and *EREP* are quite similar and give larger b^* than *EREM* for moderate Γ, while the largest b^* is associated with *EREV*. According to these numerical results, the proposed estimated relative efficiency measures *EREV*, *EREW* and *EREP* are considerably different from *ERE* and *EREM*. This phenomenon is more pronounced for small $\Gamma(>1/t)$.

3.6 *Multiple-objective efficiency measure*

As there are usually several goals and concerns in the experiment, it is desirable to have an adequate efficiency to reflect different aspects of interests and needs of the researchers. Any single efficiency measure can not take all these considerations into account. In order to achieve this, we adopt the idea of a multiple-objective strategy for the construction of an efficiency measure.

Suppose there are L distinct efficiency measures representing different objectives as implemented by the function Ψ_l, $l = 1,\ldots,L$. Specifically, let $\Psi_l(t, b, \Gamma)$ denote the efficiency measure of the RCBD relative to the CRD with respect to the function Ψ_l. We could consider a multiple-objective efficiency measure based on a weighted average of the functions Ψ_l, $l = 1,\ldots,L$, as follows

$$\Psi(t,b,\Gamma) = \sum_{l=1}^{L} w_l \Psi_l(t,b,\Gamma),$$

where $0 \leqslant w_l \leqslant 1$ are user-selected constants and $\sum_{l=1}^{L} w_l = 1$. To exemplify this approach, we consider the combination of efficiency measures *EREM*, *EREV*, *EREW*, and *EREP* described in the previous sections:

$$\Psi(t, b, \Gamma) = w_1 EREM + w_2 EREV + w_3 EREW + w_4 EREP,$$

wherethe weights w_l, $l = 1,\ldots,4$, are chosen to reflect the relative importance of these four distinct criteria and result in an optimal formulation for a profound evaluation of RCBD. This leads to finding an RCBD which can satisfy certain level of efficiency or guarantee higher efficiency under the compound criterion $\Psi(t, b, \Gamma)$. For example, it may be advisable to consider only an RCBD with a weighted efficiency $\Psi(t, b, \Gamma)$ larger than, say, 125% to be better than the comparable CRD.

As an alternative to the multiple-objective efficiency measure, multiple constrained criteria can be applied simultaneously to the efficiency measures. In this case, the researcher needs to prioritize the relative importance of the efficiency measures. For a general discussion and recent advances in multiple-objective design strategies, see Wong (1999) for details.

4 Summary

The completely randomized design and randomized complete block design are the most fundamental and useful in the analysis of variance models. The major advantage of using a randomized complete block design instead of a completely randomized design is the reduction in error variance. The widely accepted relative precision measure is purported to evaluate the relative efficiency in terms of the ratio of error variances of both designs. However, this relative precision measure does not take account of the loss in error degrees of freedom in a randomized complete block design as compared with that in a completely randomized design. Although the decrease in probability of obtaining a significance result for a randomized complete block design is well recognized in the literature, it seems that less attention has been paid to the discrepancy between the issues of precision and sensitivity. More importantly, no specific guidance for constructing a practical efficiency measure can be found in the literature. In this article, we consider three relative efficiency measures for evaluating different aspects of the sensitivity of both designs. Unlike other research that examine parameter values, we focus on the estimates of the relative efficiency measure that possess immediate applicability and practical importance. The characteristics of the estimated relative efficiency measures are presented both graphically and numerically. The results showed that they are substantially different, especially when the block effects are small or moderate in terms of the coefficient of partial determination between responses and block effects, given that treatment effects are present in the models. Naturally, these should be the cases more likely to occur in practice. Recognition of the similarities and differences between the relative measures of precision and sensitivity helps to clarify the issue of evaluating effective blocking and to choose an appropriate measure of relative efficiency in a randomized complete block experiment.

We conclude that the proposed efficiency measures provide feasible solutions for the evaluation of efficiency regarding sensitivity in the context of RCBD relative to CRD. As a practical guideline, the efficiency evaluation of RCBD relative to CRD can be approached in terms of (1) comparing the precisions, (2) comparing the observed significance levels, (3) comparing the widths of Scheffé confidence intervals, or (4) comparing the powers. Additionally, when there are various goals in the RCBD experiment, a multiple-objective approach to the evaluation of efficiency can be employed to reflect the specific needs of the researcher more adequately. Finally, we note that the *p*-value, width of confidence interval and power level considered in this article are essential to the two approaches of interval estimation and hypothesis testing in statistical inference. The proposed three efficiency measures could be extended for use in more general block designs and models. It would be of practical importance to develop these extensions.

Acknowledgements

The authors are grateful to two referees for several suggestions that have improved the exposition.

References

Cochran, W. G. and G. M. Cox (1957), *Experimental designs* (2nd edn), John Wiley & Sons, New York.

Hinkelmann, K. and O. Kempthorne (1994), *Design and analysis of experiment: volume I: introduction to experimental design*, John Wiley & Sons, New York.

Jensen, D. R. (1980), Efficiencies in multivariate paired experiments, *Biometrical Journal* **22**, 399–405.

Jensen, D. R. (1982), Efficiency and robustness in the use of repeated measurements, *Biometrics* **38**, 813–825.

Lentner, M., J. C. Arnold and K. Hinkelmann (1989), The efficiency of blocking: how to use MS(Blocks)/MS(Error) correctly, *The American Statistician* **43**, 106–108.

Morrison, D. F. (1972), The analysis of a single sample of repeated measurements, *Biometrics* **28**, 55–71.

Neter, J., M. H. Kutner, C. J. Nachtsheim and W. Wasserman (1996), *Applied linear statistical models* (4th edn), Irwin, Homewood, IL.

Samuels, M. L., G. Casella and G. P. McCabe (1991), Interpreting blocks and random factors, *Journal of the American Statistical Association* **86**, 798–808.

Vonesh, E. F. (1983), Efficiency of repeated measures designs versus completely randomized designs based on multiple comparisons, *Communications in Statistics–Theory and Methods* **12**, 289–301.

Wong, W. K. (1999), Recent advances in multiple-objective design strategies, *Statistica Neerlandica* **53**, 257–276.

Received: April 2002. Revised: August 2003.