ESTIMATION OF THE TREATMENT DIFFERENCE IN MULTICENTRE TRIALS

Valerii Fedorov[1], Byron Jones[1,*], Matthew Jones[2] and Anatoly Zhigljavsky[2]

[1]GlaxoSmithKline Pharmaceuticals

[2]University of Cardiff

ABSTRACT

The three fixed effects estimators of a treatment difference are compared under conditions of random enrolment in a multicentre clinical trial. These comparisons are done assuming five different enrolment schemes. The estimators are compared via simulation using their expected mean squared errors. Unlike previous discussions of these three estimators we take explicit account of the effect of centres that fail to enrol patients to one or both treatment arms.

Within each centre we assume enrolment follows a Poisson process and consider the two situations where the mean rate of this process is the same in every centre and where the mean rates are sampled from a gamma distribution. The effect of patient drop-out is studied as well as the effect of increasing the number of centres.

Simulations show that for many sound scenarios the simpler estimator corresponding to the simplest model works better, even for the cases when data are generated by more complex models.

Key Words: Multicentre trial; Fixed effects models; Random enrolment

*Corresponding author

1. Introduction

Clinical trials are run in more than one centre for a number of reasons: in order that the required number of patients can be recruited within the time limit of the trial, to obtain results from to a wider population of patients or to permit a more valid extrapolation of results. At the planning stage of the trial the required number of patients and centres are calculated. Each centre is then told the target number of patients to enrol, so that in total they will provide the required number. To maximize precision of estimation the optimal strategy is to have an equal number of patients in each centre and to allocate an equal number of patients to each treatment within a centre. While this is optimal from a statistical viewpoint, it assumes that the enrolment rate is the same for all centres. In reality, this assumption is never true and enrolment rates vary over the centres. To illustrate this, consider the histogram given in Figure 1 of the sizes of the 425 centres that participated in a recent trial completed by a large pharmaceutical company.

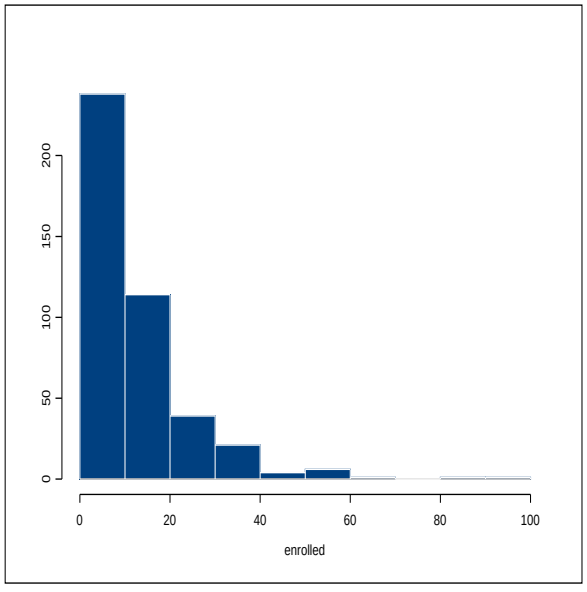

Figure 1. Centre sizes in a pharmaceutical trial

As is often the case in real trials, some centres in this trial enrolled more patients than others. A reason for this might be that some centres found it easier to enrol than others and consequently, perhaps, these centres were allowed to enrol more than their target. Such observed variation in the number of patients enrolled per centre is often consistent with a relatively simple probability distribution, the negative binomial. This distribution appears when (i) patients are enrolled independently to a centre at random and at a fixed rate, according to a Poisson distribution and (ii) the rates vary over the centres in way consistent with a gamma distribution.

In this report, to simplify the presentation, we will only consider trials with two treatments, although a number of results can be easily generalized.

In a multicentre trial each centre produces its own treatment effects and these must be combined in some way to produce a representative estimate of the "overall" treatment difference. Dragalin et al. (2001) defined the combined response to treatment (CRT) as an appropriate representative quantity and compared three alternative estimators of it. These estimators were defined in the "test-ground" situation where the fitted models are of the fixed-effects type and it is assumed that the inferences derived for the selected set of centres can be generalized to the whole population of centres. Fedorov et al. (2002) developed this work in the random-effects setting and proposed a model that contained random treatment effects and gave corresponding definitions of the CRT. This paper is concerned with comparing the three test-ground estimators in the context of random recruitment of patients to centres.

2. Models and estimators

Here we follow the notation used by Dragalin et al. (2001) and assume that N centres have been initiated, that there are two treatment arms in each centre ($j = 1, 2$) and that n_{ij} patients have been randomized to the j^{th} treatment in the i^{th} centre. Initially,

we assume that the n_{ij} are fixed, and hence that all randomness is attributable to the observational errors. Unlike previous work on this topic we will allow n_{i1} or n_{i2} or both to be zero, to allow for cases where one or both of the treatment arms have failed to enrol any patients. However, we will always assume that $n_{\cdot j} = \sum_{i=1}^{N} n_{ij} > 0$ for $j = 1, 2$, i.e., that overall at least one patient has been randomized to each of the treatments. For a reader striving for more mathematical rigour, we note that the probability that $n_{\cdot j} = 0$ is assumed to be very small.

Let variable Y_{ijk} represent the response observed on the k^{th} patient on the j^{th} treatment in the i^{th} centre, ($i = 1, 2, \ldots,\ N$, $j = 1,\ 2$ and $k = 1, 2, \ldots,\ n_{ij}$), let μ_{ij} represent the true mean response for treatment j at centre i, and let $\delta_i = \mu_{i2} - \mu_{i1}$ represent the true treatment difference at centre i.

In the linear case the CRT is defined as

$$CRT = \delta' = \sum_{i=1}^{N} \omega_i \delta_i \,,$$

where the ω_i are weights determined prior to the trial, and $\omega_i \geq 0$ with $\sum_{i=1}^{N} \omega_i = 1$.

Here we will concentrate on the simple version of the CRT:

$$CRT = \delta = \frac{1}{N} \sum_{i=1}^{N} \delta_i.$$

We consider models of increasing complexity for the responses observed in the trial:

$$\begin{cases} \text{Model I} & : \quad Y_{ijk} = \mu + (-1)^j \tau + \epsilon_{ijk} \,, \\ \text{Model II} & : \quad Y_{ijk} = \mu_i + (-1)^j \tau + \epsilon_{ijk} \,, \\ \text{Model III} & : \quad Y_{ijk} = \mu_i + (-1)^j \tau_i + \epsilon_{ijk} \,. \end{cases} \tag{1}$$

While our chosen parameterization is not standard, it is convenient for the two-arm case.

The true treatment difference in centre i is $\delta_i = 2\tau_i$. The measurement error for the k^{th} patient on treatment j in centre i is ϵ_{ijk}. These errors are assumed to be independent random variables with zero mean and the same variance $\sigma^2 > 0$.

In what follows, the standard notation for summing over a subscript is applied (that is, replacing the subscript with a dot), for example: $Y_{ij.} = \sum_{k=1}^{n_{ij}} Y_{ijk}$ and $n_{.j} = \sum_{i=1}^{N} n_{ij}$. The mean is denoted by the addition of a bar, e.g., $\bar{Y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$. If there are no observations on treatment arm j in centre i, then the corresponding $Y_{ij.}$ and $\bar{Y}_{ij.}$ are assumed to equal zero.

The least squares estimators of the CRT, obtained from these three models, respectively, are, for $J = I, II$ and III:

$$\widehat{\Delta}_J = \sum_{i=1}^{N} \left(W_{J,2i} \bar{Y}_{i2.} - W_{J,1i} \bar{Y}_{i1.} \right) \tag{2}$$

where

$$W_{I,ji} = \frac{n_{ij}}{n_{.j}}, \tag{3}$$

$$W_{II,1i} = W_{II,2i} = W_{II,i} = \frac{n_{i2} n_{i1} / (n_{i2} + n_{i1})}{\sum_{k=1}^{N} n_{k2} n_{k1} / (n_{k2} + n_{k1})}, \tag{4}$$

and

$$W_{III,ji} = \frac{1}{N}. \tag{5}$$

Note that in the presence of centres that have either $n_{i1} = 0$ or $n_{i2} = 0$ (a common situation in our experience), all three estimators are in general biased estimators of δ. If both $n_{i1} > 0$ and $n_{i2} > 0$ for all centres, and Model III is assumed to be the true model, then both Δ_I and Δ_{II} are biased (and of course Δ_{III} is unbiased).

When $n_{i1} = n_{i2} = n_i$, which we call *balanced randomization*, the weights for the first and second estimators coincide and $\Delta_I = \Delta_{II}$. If the n_i are the same for all centres, i.e.,

there is an equal number of patients per treatment per centre, a situation which we refer to as *balanced enrolment*, then $\Delta_I = \Delta_{II} = \Delta_{III}$.

Estimators Δ_I, Δ_{II} and Δ_{III} are equivalents of the estimators determined by the SAS® GLM procedure for the so-called "Type I", "Type II" and "Type III" models respectively (see, e.g., Senn, 1997 and Schwemer, 2000).

In order to compare these estimators we will use the Mean Squared Error (MSE).

The MSEs for the three estimators are as follows:

$$MSE(\Delta_I) = \sigma^2 \left(\frac{1}{n_{.2}} + \frac{1}{n_{.1}} \right) + \left[\sum_{i=1}^{N} \left((W_{I2i} - W_{I1i})\, \mu_i + \left(W_{I2i} + W_{I1i} - \frac{2}{N} \right) \tau_i \right) \right]^2, \qquad (6)$$

$$MSE(\Delta_{II}) = \sigma^2 \sum_{i=1}^{N} W_{IIi}^2 \left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}} \right) + 4 \left[\sum_{i=1}^{N} \left(W_{IIi} - \frac{1}{N} \right) \tau_i \right]^2, \qquad (7)$$

and, if $n_{ij} > 0$ for all i, j then

$$MSE(\Delta_{III}) = \frac{\sigma^2}{N^2} \sum_{i=1}^{N} \left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}} \right).$$

Note that in the case of normally distributed data, the second terms in equations (6) and (7) after dividing by σ^2, coincide with the noncentrality parameters of the χ^2-distributed sum of squared residuals (cf. Chinchilli and Bortey, 1991).

If $n_{ij} = 0$ for any i and j then the formulae for the MSEs of the first two estimators remain unchanged (we assume that $W_{IIi}^2/n_{ij} = 0$ if $n_{ij} = 0$). However, this is not the case with the Type III estimator, as Δ_{III} is no longer unbiased. Consequently, the expression for $MSE(\Delta_{III})$ needs to be modified as follows, where $\hat{\delta}_i = \bar{Y}_{i2.} - \bar{Y}_{i1.}$

Let $\psi = \{i : n_{i1} > 0 \text{ and } n_{i2} > 0\}$ represent the set of centres where there are patients on both treatment arms. Then:

$$\begin{aligned} MSE(\Delta_{III}) &= E[\Delta_{III} - \delta]^2 = E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\delta}_i - \delta_i\right)\right]^2 \\ &= \frac{1}{N^2}\left[\sigma^2\sum_{i\epsilon\psi}\left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}}\right) + 4\left(\sum_{i\notin\psi}\tau_i\right)^2\right]. \end{aligned} \tag{8}$$

Comparisons of the MSEs in equations (6) to (8) are difficult due to the large numbers of parameters involved. Simplification is possible by assuming particular scenarios for the values of $\{\mu_i\}$ and $\{\tau_i\}$. Let $\bar{\mu} = N^{-1}\sum_{i=1}^{N}\mu_i$ and $\bar{\tau} = N^{-1}\sum_{i=1}^{N}\tau_i$. A popular choice is $|\mu_i - \bar{\mu}| \leq \zeta$ and $|\tau_i - \bar{\tau}| \leq \xi$. We follow Dragalin et al. (2002) and assume that τ_i and μ_i are independently sampled from populations with means $E(\tau_i) = 1$ and $E(\mu_i) = \mu$, and variances σ_τ^2 and σ_μ^2, respectively. As shown in Dragalin et al. (2002) the expected MSEs of the first two estimators (over the distributions of μ_i and τ_i) are then:

$$\overline{MSE}(\Delta_I) = \sigma^2\left(\frac{1}{n_{\cdot 2}} + \frac{1}{n_{\cdot 1}}\right) + \sigma_\mu^2\sum_{i=1}^{N}(W_{I2i} - W_{I1i})^2 + \sigma_\tau^2\sum_{i=1}^{N}\left(W_{I2i} + W_{I1i} - \frac{2}{N}\right)^2 \tag{9}$$

and

$$\overline{MSE}(\Delta_{II}) = \sigma^2\sum_{i=1}^{N}W_{IIi}^2\left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}}\right) + 4\sigma_\tau^2\sum_{i=1}^{N}\left(W_{IIi} - \frac{1}{N}\right)^2. \tag{10}$$

These formulae hold irrespective of whether the values of n_{ij} are all positive. As stated previously, we only need to assume that $n_{\cdot 1} > 0$ and $n_{\cdot 2} > 0$.

We emphasize that $\{\mu_i\}$ and $\{\tau_i\}$ are considered as samples from populations only in order to produce simple and useful scenarios for these $2N$ effects. This should not be taken to imply that we are fitting random effect models to the treatment and centre effects.

Equation (8) implies that the expected MSE of the third estimator is

$$\overline{MSE}(\Delta_{III}) = \frac{1}{N^2}\left[\sigma^2\sum_{i\epsilon I}\left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}}\right) + 4L\left(\sigma_\tau^2 + 1\right)\right], \tag{11}$$

where L represents the number of centres that have one or more treatment arms with no patients. For the case where all $n_{ij} > 0$ formula (11) simplifies to

$$\overline{MSE}(\Delta_{III}) = \frac{\sigma^2}{N^2} \sum_{i=1}^{N} \left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}} \right) ,$$

which coincides with that of Dragalin et al. (2002).

3. Enrolment schemes

Here we will compare the average MSEs ($\overline{MSE}$s) of the estimators under conditions of random enrolment of patients to the centres. We first introduce some notation.

$Poisson(\lambda)$ will denote a Poisson random variable with mean rate λ, $Binomial(n, \pi)$ will denote a binomial random variable with n trials and success probability π and $Gamma(\alpha, \beta)$ will denote a gamma random variable with parameters α and β. (See Johnson, Kotz and Balakrishnan, 1994, and Johnson, Kotz and Kemp, 1993, for example for details.)

We will consider four scenarios that assume different types of random enrolment to the centres and a fifth that is included for the sake of comparison and corresponds to non-random enrolment. Within each centre the enrolment process for Scenarios I to IV follows the Poisson distribution. In Scenario III, patient drop-out is assumed to be at random and is independent of centre and treatment.

- Scenario I: n_{i1} and n_{i2} are independent; $n_{i1} \sim Poisson(\lambda_i)$ and $n_{i2} \sim Poisson(\lambda_i)$.
- Scenario II: $n_{i1} = n_{i2} = n_i$; $n_i \sim Poisson(\lambda_i)$.
- Scenario III: $n_i \sim Poisson(\lambda_i)$; $n_{i1} \sim Binomial(n_i, (1-p))$, $n_{i2} \sim Binomial(n_i, (1-p))$; for p a pre-specified constant, $0 \leq p \leq 1$.
- Scenario IV: n_{i1} and n_{i2} are independent; $n_{i1} \sim Poisson(\lambda_i)$ and $n_{i2} \sim Poisson(\lambda_i)$; with $\lambda_i \sim Gamma(\alpha, \beta)$.

- SCENARIO V: n_{i1} and n_{i2} are fixed under assumptions of balanced randomization.

Scenario III allows for patient drop-outs after enrolment, with the constant p representing the proportion of patients withdrawing from the trial. Here n_{i1} and n_{i2} each have a Poisson distribution with parameter $\lambda_i(1-p)$ (Johnson, Kotz and Kemp, 1993). In Scenario IV, the true recruitment rates vary over the centres.

To obtain an understanding of the behaviour of the estimators in different environments, three contrasting enrolment situations were generated, namely:

- CASE I: the number of centres is small, but a large number of patients have been enrolled at each centre. In our example we take $N = 10$, $n_{i1} \sim Poisson\ (\lambda_i)$ and $n_{i2} \sim\ Poisson\ (\lambda_i)$, where $\lambda_i = 100$.
- CASE II: the number of centres is large, but only a small number of patients have been enrolled at each centre. In our example we take $N = 100$, $n_{i1} \sim\ Poisson\ (\lambda_i)$ and $n_{i2} \sim\ Poisson\ (\lambda_i)$, where $\lambda_i = 10$.
- CASE III: the number of centres is extremely large, with a very low enrolment rate. In our example we take $N = 250$, $n_{i1} \sim\ Poisson\ (\lambda_i)$ and $n_{i2} \sim Poisson\ (\lambda_i)$ where $\lambda_i = 4$.

It is assumed that enrolment is stopped after a preselected time. More specifically in our case it means that each of the cases has on average 2,000 patients, with an average of 1,000 on each treatment arm.

In Section 4 we define the model validity range as introduced by Dragalin et al. (2002). This can be used to determine which of the three fixed-effects estimators is preferred depending on the variability of the treatment effects over the centres and the number of patients in each centre.

Sections 5 to 8 report results obtained using simulation. In Section 5 we compare the three estimators in terms of their expected MSEs for Scenarios I and II. In Section 6 the

effect of increasing the number of centres is studied under both Scenarios I and II. This is achieved by setting $\lambda_i = 1000/N$, for all i. In Section 7, we study the effects of patient drop-out (Scenario III) and consider the influence on the estimators of an increasing proportion, p, of patients withdrawing from the trial. In Section 8 we consider Scenario IV, where the enrolment rates are sampled from a gamma distribution, and study the behaviour of the estimators under conditions of increasing variability in the enrolment rates.

For comparison purposes, we include deterministic enrolment (Scenario V) in some of the simulation studies. The purpose of this is to compare the results of random enrolment with those for a pre-fixed balanced randomization of patients to centres.

We end by giving an overall review of the results and some conclusions in Section 9.

4. Model validity range

Although the purpose of the simulations is to compare the properties of the three types of estimator under different enrolment scenarios, we need to put this into the context of the assumptions made about the centre and treatment effects μ_i and τ_i. This can be done using the *model validity range*, introduced by Fedorov et al. (1998) and used in the multicentre context by Dragalin et al. (2002). In this section we will consider the n_{ij} as fixed (i.e., not random) in order to illustrate the usefulness of the model validity range. In Section 5.2 we will use the model validity range in the context of random enrolment.

The model validity range for Model II with respect to Model III, is the set Γ_{II} of all values for the standardized deviations of centre specific treatment effects from their mean, $(\tau_i - \bar{\tau})/\sigma$, such that

$$MSE(\Delta_{II}) \leq MSE(\Delta_{III}).$$

If (for the sake of simplicity) it is assumed that all $n_{ij} > 0$:

$$\Gamma_{II} = \left\{ \frac{\tau_i - \bar{\tau}}{\sigma} : \ \Big[\sum_{i=1}^{N} \left(W_{IIi} - \frac{1}{N} \right) \frac{(\tau_i - \bar{\tau})}{\sigma} \Big]^2 \leq \frac{1}{4} \Big[\sum_{i=1}^{N} \Big(\frac{1}{N^2} - W_{IIi}^2 \Big) \Big(\frac{1}{n_{i2}} + \frac{1}{n_{i1}} \Big) \Big] \right\}.$$

Therefore, the model validity range is the set of all possible values of the standardized centre specific treatment effects for which one can use a more parsimonious model to get a better estimate of the CRT. This set is completely defined by the achieved enrolment $\{n_{ij}\}$. It is clear that under a situation that is close to balance, both for enrolment and randomization, i.e., $W_{IIi} = 1/N$, the model validity range can be very large, especially for low enrolment. Dependence here is more complex, because if $W_{IIi} \simeq 1/N$ then $W_{IIi}^2 \simeq 1/N^2$ and we have zeros on both sides.

Similarly, we can determine the model validity range for Model I with respect to Model III:

$$\Gamma_I = \left\{ \Big(\frac{\mu_i - \bar{\mu}}{\sigma}, \frac{\tau_i - \bar{\tau}}{\sigma} \Big) : \Big[\sum_{i=1}^{N} \Big\{ \Big(\frac{n_{i2}}{n_{\cdot 2}} - \frac{n_{i1}}{n_{\cdot 1}} \Big) \frac{(\mu_i - \bar{\mu})}{\sigma} + \Big(\frac{n_{i2}}{n_{\cdot 2}} + \frac{n_{i1}}{n_{\cdot 1}} - \frac{2}{N} \Big) \frac{(\tau_i - \bar{\tau})}{\sigma} \Big\} \Big]^2 \right.$$
$$\left. \leq \sum_{i=1}^{N} \frac{1}{N^2} \Big(\frac{1}{n_{i2}} + \frac{1}{n_{i1}} \Big) - \Big(\frac{1}{n_{\cdot 2}} + \frac{1}{n_{\cdot 1}} \Big) \right\}. \qquad (12)$$

Under balanced randomization, i.e., when $n_{i1} = n_{i2} = n_i$, estimators Δ_I and Δ_{II} coincide. Hence

$$\Gamma_I = \Gamma_{II} = \left\{ \Big[\sum_{i=1}^{N} \Big(\frac{n_i}{n'} - \frac{1}{N} \Big) \frac{(\tau_i - \bar{\tau})}{\sigma} \Big]^2 \leq \frac{1}{2} \Big[\frac{1}{N^2} \sum_{i=1}^{N} \frac{1}{n_i} - \frac{1}{n'} \Big] \right\},$$

where here we use $n' = n_{\cdot 1} = n_{\cdot 2}$. Notice that for n_i proportional to $\frac{1}{N}$, both parts of the inequality are zero, meaning that, in this special case, the model validity range is empty: the use of a more parsimonious model doesn't improve the properties of the estimator. However, remember that in this special case all three estimators coincide.

As in the previous section, it is convenient to assume that the μ_i and τ_i are independently sampled from populations with means $\bar{\mu}$ and $\bar{\tau}$, respectively, and variances σ_μ^2 and σ_τ^2, respectively. Making these assumptions, we take expectations with respect to the distributions of the μ_i and τ_i. We emphasize that this is done to give simple and useful scenarios values of μ_i and τ_i and not because we are fitting random effect models to the treatment and centre effects.

Straightforward calculations yield the expected model validity ranges (over the distributions of μ_i and τ_i):

$$\overline{\Gamma}_I = \left\{ \left(\frac{\sigma_\mu^2}{\sigma^2}, \frac{\sigma_\tau^2}{\sigma^2}\right) : \quad \frac{\sigma_\mu^2}{\sigma^2} \sum_{i=1}^{N} \left(\frac{n_{i2}}{n_{\cdot 2}} - \frac{n_{i1}}{n_{\cdot 1}}\right)^2 + \frac{\sigma_\tau^2}{\sigma^2} \sum_{i=1}^{N} \left(\frac{n_{i2}}{n_{\cdot 2}} + \frac{n_{i1}}{n_{\cdot 1}} - \frac{2}{N}\right)^2 \right.$$
$$\left. \leq \sum_{i=1}^{N} \frac{1}{N^2} \left[\left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}}\right) - \left(\frac{1}{n_{\cdot 2}} + \frac{1}{n_{\cdot 1}}\right) \right] \right\}$$

and

$$\overline{\Gamma}_{II} = \left\{ \frac{\sigma_\tau^2}{\sigma^2} : \quad \frac{\sigma_\tau^2}{\sigma^2} \sum_{i=1}^{N} \left(W_{IIi} - \frac{1}{N}\right)^2 \leq \frac{1}{4} \Big[\sum_{i=1}^{N} \left(\frac{1}{N^2} - W_{IIi}^2\right) \left(\frac{1}{n_{i2}} + \frac{1}{n_{i1}}\right) \Big] \right\}. \tag{13}$$

The purpose of the validity range is to identify situations where Model I or Model II would give smaller MSEs than Model III even in the situation where treatment effects vary over the centres. Given fixed σ^2, the validity range depends on the σ_τ^2 and the n_{ij}. To illustrate the use of the validity range we consider the special case $n_{i1} = n_{i2} = n_i$, so that $\overline{\Gamma}_I = \overline{\Gamma}_{II}$.

In order to provide a situation where the values of n_{ij} are unequal, we will consider a special simple case where $n_i = \lambda$, for $i = 1, 2, \ldots, \frac{N}{2}$ and $n_i = (\lambda - d)$ for $i = \frac{N}{2} + 1, \frac{N}{2} + 2, \ldots, N$. We will refer to the parameter d as the *disbalance*. Note that the parameter λ as used in this illustrative example is a fixed scalar and not a random variable.

Corresponding to the three enrolment situations introduced in the previous section we consider the following: Case I, $N = 10$ and $\lambda = 100$; Case II, $N = 100$ and $\lambda = 10$ and Case III, $N = 250$ and $\lambda = 4$. In all cases $\sigma^2 = 1$.

Replacing the inequality in (13) with an equality, and given values of λ and d, we can determine the value of σ_τ for which the MSEs of the three models are equal. If this is done for Case I, for example, for the full range of d, then the values of σ_τ define a boundary line between two regions as illustrated in the left-hand plot of Figure 4. The region below the line corresponds to combinations of (d, σ_τ) for which Model II (and Model I for this special case) is preferred. The middle and right-hand plots in Figure 4 correspond to Cases II and III, respectively. It is clear from the plots that (i) in the presence of severe disbalance Model II may be preferred even in the presence of large treatment variability and (ii) larger treatment variability can be tolerated and Model II preferred if the number of centres is large and the number of patients per centre is small.

5. Simulation study for Scenarios I and II

A simulation study was conducted with fixed values for the standard deviations as used by Dragalin et al. (2002): $\sigma = 1, \sigma_\tau = 0.25$ and $\sigma_\mu = 0.25$. Each simulation consisted of 500,000 runs.

All the results obtained from the simulation study are normalised. That is, the $\overline{MSE}$ values are multiplied by the expected total number of patients (in our case $E(n_{.1}) + E(n_{.2}) = 2\sum_{i=1}^{N} \lambda_i = 2,000$).

The results are mainly presented in the form of histograms, which give in each case, the frequency distribution of $\overline{MSE}$ for a particular estimator under the two different scenarios and three enrolment cases.

Looking at Figures 3 to 5 it is evident that the distributions of $\overline{MSE}$ for Case I are unimodal with some slight asymmetry. This was also true for Cases II and III (not shown).

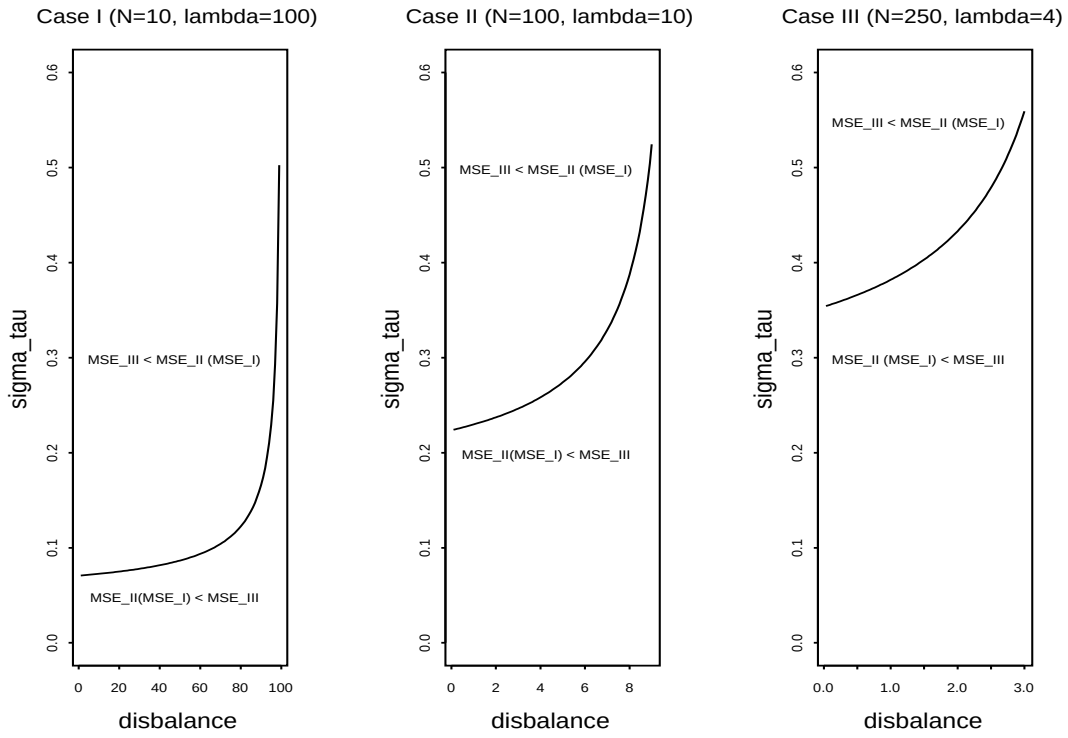

FIGURE 2. Validity ranges for varying amounts of disbalance

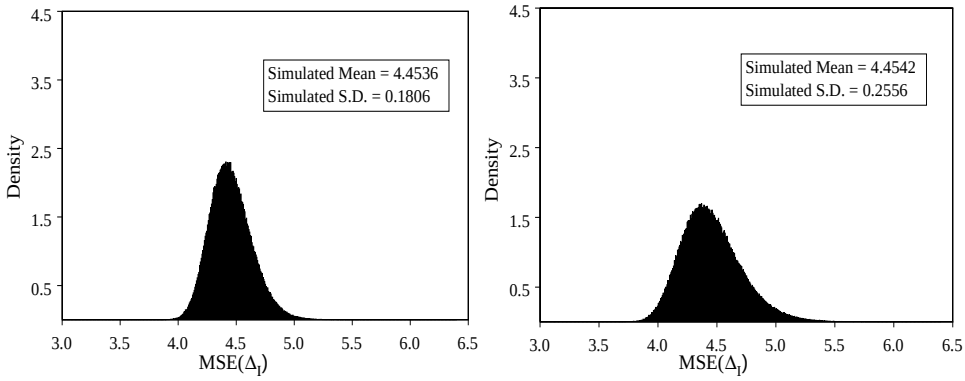

FIGURE 3. (a) Histogram of $\overline{MSE}(\Delta_I)$ for Case I, Scenario I; (b) Histogram of $\overline{MSE}(\Delta_I)$ for Case I, Scenario II.

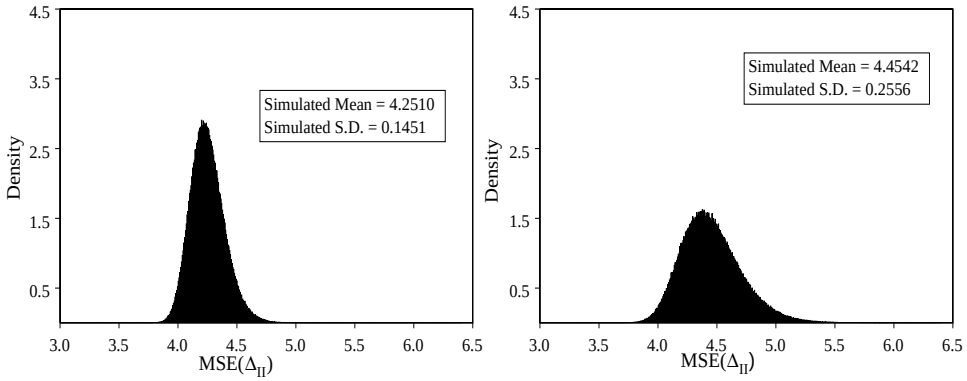

FIGURE 4. (a) Histogram of $\overline{MSE}(\Delta_{II})$ for Case I, Scenario I ; (b) Histogram of $\overline{MSE}(\Delta_{II})$ for Case I, Scenario II.

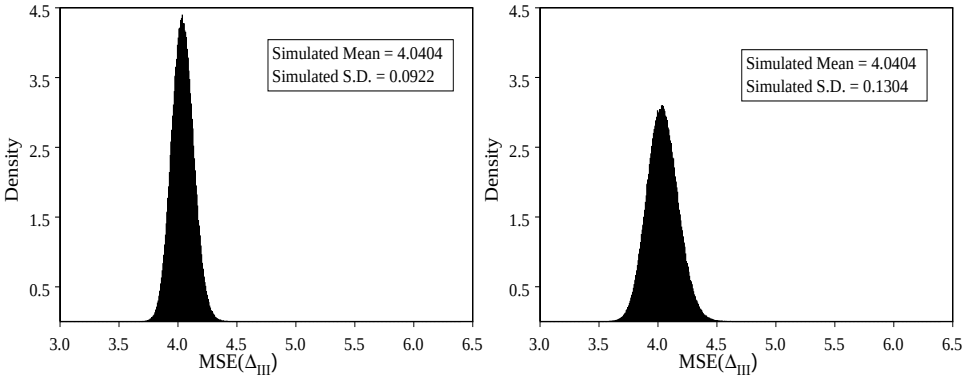

FIGURE 5. (a) Histogram of $\overline{MSE}(\Delta_{III})$ for Case I, Scenario I; (b) Histogram of $\overline{MSE}(\Delta_{III})$ for Case I, Scenario II.

It is also true that, in the majority of the histograms, the distributions for Scenario II were noticeably wider than Scenario I, implying that the $\overline{MSE}$s for Scenario II were more variable. This may be due to the potential in Scenario II for more between-centre variation in the number of patients enrolled to a centre as compared to Scenario I. The independent sampling to each treatment arm in Scenario I gives an opportunity for a small(large) enrolment on one arm to be augmented by a large(small) enrolment on the other arm. In the ith centre under Scenario II the variance of the total number of patients

is $\mathrm{Var}(2\lambda_i) = 4\lambda_i$, whereas under Scenario I it is $2\,\mathrm{Var}(\lambda_i) = 2\lambda_i$, giving a ratio of the respective standard deviations of $\sqrt{1/2} = 0.71$

Statistic		CASE I		CASE II		CASE III	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
$\overline{MSE}\,(\Delta_I)$	Scen. I	4.45357	0.18063	4.49001	0.11229	4.50246	0.10571
	Scen. II	4.45424	0.25559	4.49033	0.15905	4.50252	0.14949
$\overline{MSE}\,(\Delta_{II})$	Scen. I	4.25096	0.14510	4.50541	0.12282	4.96985	0.15227
	Scen. II	4.45424	0.25559	4.49033	0.15905	4.50252	0.14949
$\overline{MSE}\,(\Delta_{III})$	Scen. I	4.04044	0.09216	4.52061	0.16279	6.31610	0.37304
	Scen. II	4.04036	0.13039	4.51687	0.20983	5.80069	0.32927

TABLE 1. Summary statistics for $\overline{MSE}\,(\Delta_I)$ – $\overline{MSE}\,(\Delta_{III})$ for Scenarios I and II, in all three cases.

Some summary statistics are given in Table 1. Looking first at results for Cases I and II there is a clear difference between the results for estimators Δ_I and Δ_{II} versus Δ_{III}. For the first two estimators the variation is larger for Case I under both scenarios. A theoretical explanation for this is given in Appendix A.

Looking at Case III, we see that the results for $\overline{MSE}\,(\Delta_{III})$ are much more variable than the corresponding results for the other two estimators. This is most likely due to the presence of centres with no patients on one or both treatment arms causing increased variability. The mean of $\overline{MSE}$ is considerably larger for Δ_{III} compared with those of the other two estimators. This fact confirms one of the points made by Dragalin et al. (2002) based on the model validity range: that the simpler estimator gets relatively better with relatively more imbalance.

A property of all three estimators under both scenarios is that their $\overline{MSE}$ value increases from Case I to Case III, as the patients are distributed over an increasing number of centres. However, the rate at which $\overline{MSE}$ increases differs greatly over the three estimators. Prominent changes occur with the Type II and Type III estimators (particularly the latter), with less noticeable increases associated with the Type I estimator. It seems that the Type I estimator is more robust to the dispersal of patients over an increasing number of centres.

The changes associated with $\overline{MSE}$ for the Type III estimator are greater as the treatment effects are unweighted. This estimator is poor when centre enrolment rates are low due to the relatively large variances of the estimated within-centre treatment differences.

The order of precedence for the estimators depends on the type of enrolment. The rankings below apply to both the expectation and variance of the $\overline{MSE}$ under Scenario I:

- Case I: Type III estimator is the best estimator, followed by the Type II, and finally the Type I estimator.
- Cases II and III: Type I estimator is the best, followed by Type II, with the Type III estimator being (notably) the worst.

Again these results confirm that the validity range of Model I is expanding together with the increase of variability in enrolment. The latter causes a significant imbalance in the number of patients per centre per arm.

5.1. $\overline{MSE}$ **versus** σ_τ**, with** $\sigma_\tau = \sigma_\mu$**.** To gain some understanding of the influence that centre and treatment-by-centre interaction effects have on the estimators we will consider the behaviour of $\overline{MSE}$ under increasing values of σ_τ and for Scenario I. Everywhere in this section we assume that σ_τ=σ_μ.

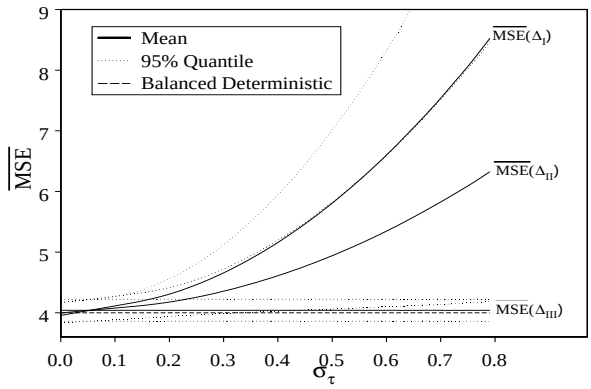

FIGURE 6. Plot of mean $\overline{MSE}$ versus σ_τ^2 for Case I, Scenario I.

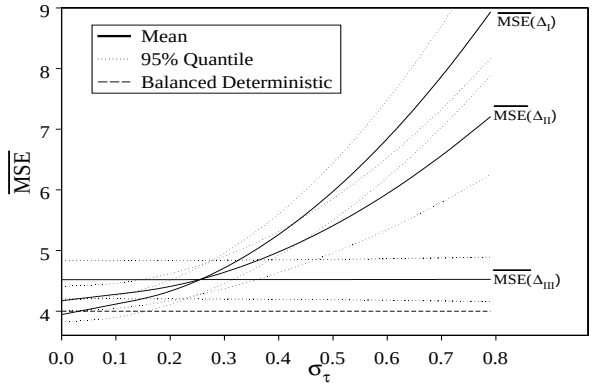

FIGURE 7. Plot of mean $\overline{MSE}$ versus σ_τ^2 for Case II, Scenario I.

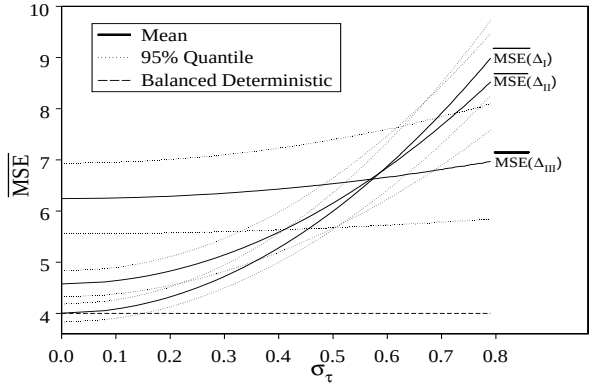

FIGURE 8. Plot of mean $\overline{MSE}$ versus σ_τ^2 for Case III, Scenario I.

Although σ_τ will vary, the within-centre variance will not and σ will be fixed and set equal to 1. The constant $\overline{MSE}$ values of the deterministic balanced randomization scenario (Scenario V) will also be included in the plots (here $n_{i1} = n_{i2} = \lambda_i$).

Each of Figures 6 to 8 show a plot of the mean $\overline{MSE}$ over the simulations. They demonstrate the marked effect that increasing σ_τ has on $\overline{MSE}(\Delta_I)$ and $\overline{MSE}(\Delta_{II})$. The increased variation in the system has caused these to compare unfavourably with $\overline{MSE}(\Delta_{III})$. Moreover, it is also evident that as patients are distributed over a broader base of centres (that is, as we move from Case I to Case III), larger values of σ_τ are required to make $\overline{MSE}(\Delta_I)$ and $\overline{MSE}(\Delta_{II})$ greater than $\overline{MSE}(\Delta_{III})$. The reason for this is that the $\overline{MSE}$ of the Type III estimator is far greater when patients are spread out over more centres, as seen in Table 1. Thus (see Figures 6 to 8) the model validity range for Model I has expanded from a minuscule value of 0.05 for Case I, to 0.25 for Case II to 0.55 for Case III.

Through studying Figure 8 we discover that, when the value of N is large, the mean value of $\overline{MSE}(\Delta_{III})$ is no longer a constant for increasing σ_τ. This can be explained by the term in (11), that accounts for centres with no patients on one or both treatment arms, becoming more prominent. To demonstrate this we calculate and compare the probabilities that one or both of the treatment arms have no patients in each of the three cases. For Case III

$$\text{Prob}\,(\text{one or both arms empty}|\lambda = 4) \;=\; 2 \times e^{-4} - \left(e^{-4}\right)^2 = 0.036296,$$

compared with 9.1 10^{-5} for Case II ($\lambda = 10$), and 7.4 10^{-44} for Case I ($\lambda = 100$). Thus, it is clearly evident that the probability that one or both arms empty is far greater for Case III.

Similar behaviour is evident in the results for Scenario II (figures not shown).

5.2. **The effect of enrolment rate disbalance on $\overline{MSE}$.** Here we again use the concept of disbalance introduced in the previous section and consider an extreme situation where half of the centres have an enrolment rate of $\lambda_i = \lambda$, and the other half have an enrolment rate of $\lambda_i = \lambda - d$, where d is the *disbalance* and λ is a constant.

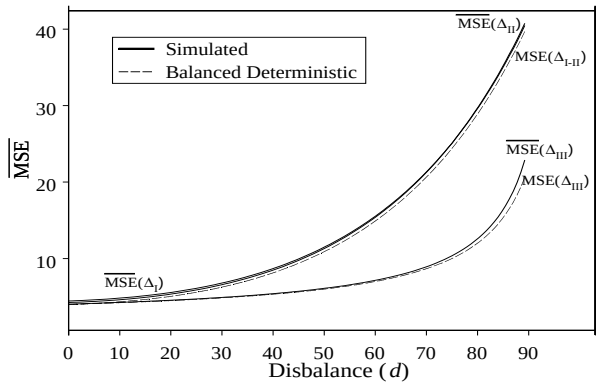

FIGURE 9. Plot of mean $\overline{MSE}$ versus disbalance for all estimators for Case I, Scenario I.

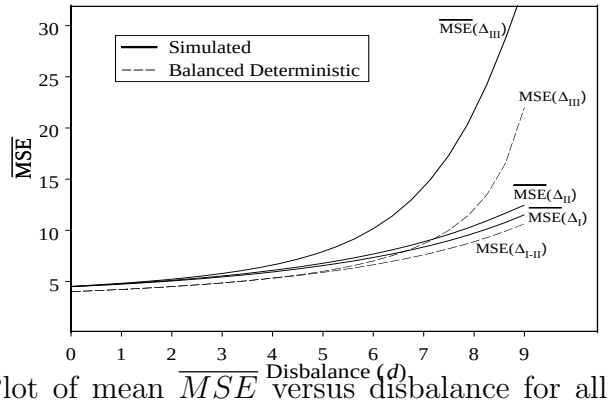

FIGURE 10. Plot of mean $\overline{MSE}$ versus disbalance for all estimators for Case II, Scenario I.

To study such a situation and gain an understanding of how this phenomenon effects the estimators, simulations were performed for each of the three cases for a range of values of the disbalance parameter d. In each of Figures 9, 10 and 11 the mean of $\overline{MSE}$ over the simulations is plotted. The figures show the effects of increasing the value of d from 0 to 90% of the mean enrolment rate assumed for each Case. It can be seen that, with the exception of $\overline{MSE}(\Delta_{II})$ and $MSE(\Delta_{I-II})$ in Case I, the $\overline{MSE}$ values all three estimators are relatively unaffected by a minor amount of disbalance. However, further increases in d lead to more significant increases in the $\overline{MSE}$s. This is particularly the case for values

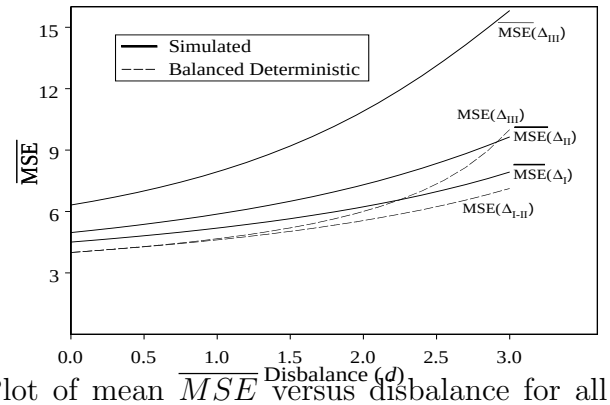

FIGURE 11. Plot of mean $\overline{MSE}$ versus disbalance for all estimators for Case III, Scenario I.

of d approximately greater than $\lambda/2$, in each of the three cases. We note that estimator Δ_{III} is relatively unaffected by disbalance in Case I, but is most affected by disbalance in Cases II and III. It can also be seen that the simulated values depart further from the deterministic value as d increases in each of the three Cases.

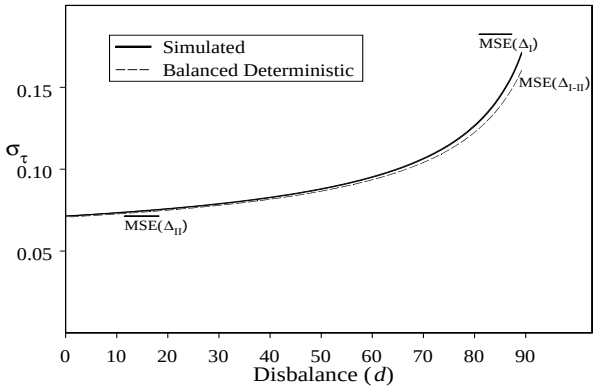

FIGURE 12. Plot of mean σ_τ against disbalance such that $\overline{MSE}(\Delta_I)$ and $\overline{MSE}(\Delta_{II})$ are equal to $\overline{MSE}(\Delta_{III})$ for Case I, Scenario I.

To determine the influence of σ_τ^2 on the effects of disbalance, values of σ_τ vs d were plotted for each of the three cases in Figures 12, 13 and 14, respectively. The dotted line in each plot is the boundary line for the case of deterministic(fixed) values of n_{ij} and exactly corresponds to the model validity ranges plotted earlier in Figure 4 for the case

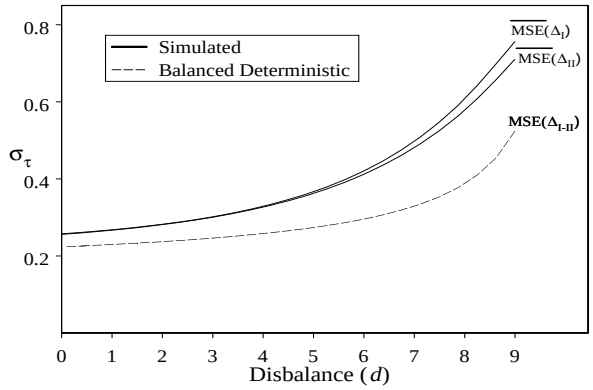

FIGURE 13. Plot of σ_τ against disbalance such that $\overline{MSE}(\Delta_I)$ and $\overline{MSE}(\Delta_{II})$ are equal to $\overline{MSE}(\Delta_{III})$ for Case II, Scenario I.

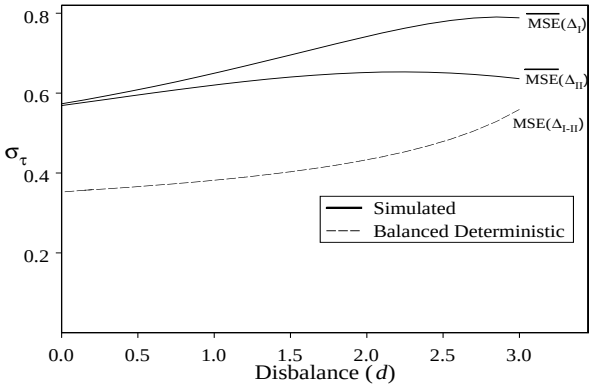

FIGURE 14. Plot of σ_τ against disbalance such that $\overline{MSE}(\Delta_I)$ and $\overline{MSE}(\Delta_{II})$ are equal to $\overline{MSE}(\Delta_{III})$ for Case III, Scenario I.

of fixed patient numbers. Compared to the validity ranges for deterministic enrolment, as displayed in Figure 4, the corresponding ranges correspond to larger values of σ_τ. In other words, Model II is preferred to Model III for even larger amounts of variability in the treatment effects (especially for Cases II and III) when enrolment is random. In addition, we see that the values of σ_τ required to make each of $\overline{MSE}(\Delta_I)$ and $\overline{MSE}(\Delta_{II})$ equal to $\overline{MSE}(\Delta_{III})$, increase with disbalance, particularly in Cases I and II. However, from Figure 14 we can see that the value of σ_τ^2 increases at a far slower rate for large

disbalance in Case III. Moreover, the value of σ_τ^2 that makes $\overline{MSE}(\Delta_{II}) \simeq \overline{MSE}(\Delta_{III})$ begins to decrease for $d > 2$.

5.3. **Imbalance measure.** A further understanding of the behaviour of the $\overline{MSE}$ of the estimators can be gained through its decomposition into individual parts attributable to different sources of variation. A source of variation that is of particular interest is the part relating to the treatment-by-centre effects. Several papers in the field of multicentre trials have concentrated on this aspect (see for example Källén, 1997 and Snapinn, 1998). We shall refer to this term as the *imbalance measure*. This measure will only be considered for the Type I and Type II [†] estimators. The Type III estimator is not considered as it is unbiased when n_{i1} and $n_{i2} > 0$ and has negligible bias otherwise.

The imbalance measure for the two estimators are as follows, (see equations (9) and (10):

$$Imb(\Delta_I) = \sigma_\tau^2 \sum_{i=1}^{N} \left(W_{I2i} + W_{I1i} - \frac{2}{N} \right)^2$$

and

$$Imb(\Delta_{II}) = 4\sigma_\tau^2 \sum_{i=1}^{N} \left(W_{IIi} - \frac{1}{N} \right)^2 .$$

If the design of the trial was fully balanced the imbalance measure would be zero and the MSE minimized. The deviation from balance can be measured, for example, by the amount of disbalance present in the design. To see the relationship between the degree of disbalance and the resulting amount of imbalance, the values of the imbalance measure for increasing amounts of disbalance are plotted in Figures 15 and 16. Clearly the imbalance measure is strongly related to d and the type of enrolment assumed. Imbalance contributes greatly to the $\overline{MSE}$ when there is a large amount of disbalance. This can be seen by comparing the current plots to Figure 9. The contribution becomes less when we have fewer patients spread over a larger number of centres, as in Cases II and III (compare

[†]It is equal to the $Bias^2$ for the Type II estimator.

with Figures 10 and 11). Secondly, the contribution of the imbalance term is greater for the Type II estimator, particularly for larger d. The difference in magnitude between the imbalance measures for the Type II and Type I estimators increases as we go from Case I to Case III.

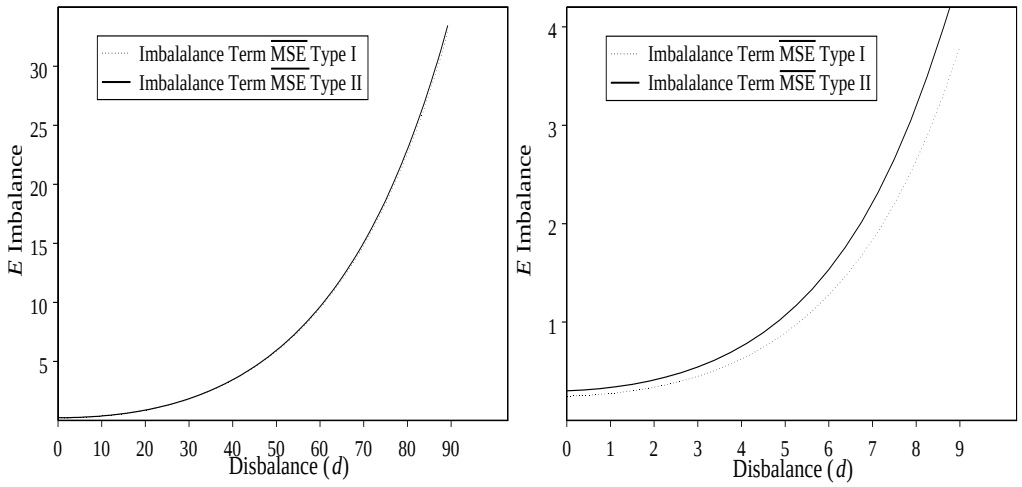

FIGURE 15. (a) Mean imbalance measure against disbalance for Case I, Scenario I; (b) Mean imbalance measure against disbalance for Case II, Scenario I.

The imbalance measure increases from Case I to Case III for increasing d, due to the fact that the treatment-by-centre effect is more prominent when there is disbalance and only a few centres. The non-constant enrolment rates across the centres significantly inflate the term relating to treatment-by-centre interaction when only a few centres are involved in the trial. However, the increase is not as prominent in Cases II to III, as the larger number of centres cause a dampening of this effect.

This increase in the imbalance measure compensates for the decrease in the other components of $\overline{MSE}$ (e.g., the second component of $\sigma_\tau^2 \sum_{i=1}^{N} \left(W_{I2i} - W_{I1i}\right)^2$ of $\overline{MSE}(\Delta_I)$ is zero when $n_{i1} = n_{i2}$, for all i). In Appendix A we prove that the imbalance measure always increases when we go from Scenario I to Scenario II. The second term of $\overline{MSE}\,(\Delta_I)$

in (9) disappears under Scenario II. However, for the values of the parameters that we have used in the simulations, it is evident that increased values of the imbalance term under Scenario II introduce increased variability in the system. This provides an explanation for the output in Table 1.

The imbalance measure for the Type II estimator is greater than that of the Type I estimator, particularly in Cases II and III, due to the increasing instability of the weights W_{IIi}. As d increases, the weights become less stable, and this is particularly prevalent in cases where the enrolment rate is already low.

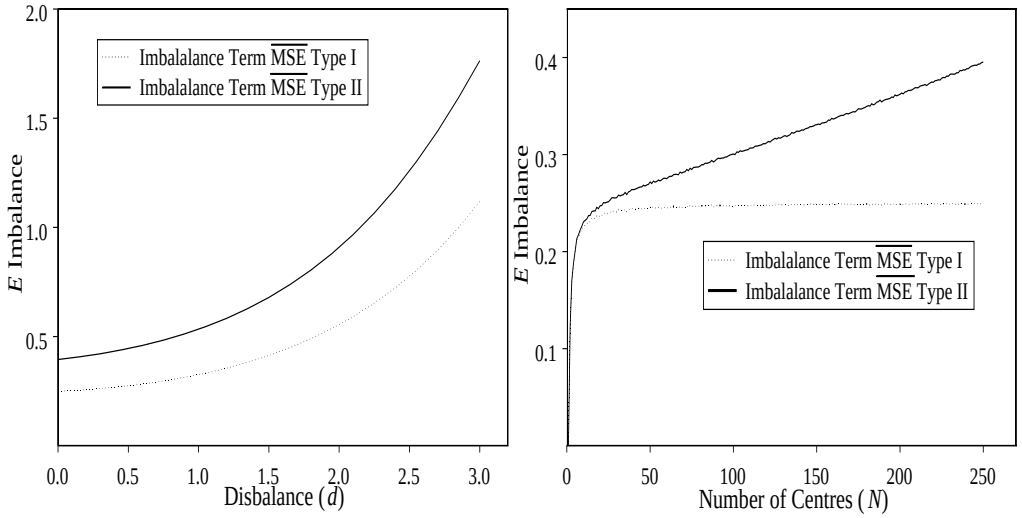

FIGURE 16. (a) Mean imbalance measure against disbalance for Case III, Scenario I; (b) Mean imbalance measure against different number of centres N, for Scenario I.

6. Effect of increasing N on $\overline{MSE}$

As specified earlier (see Section 1), one of the purpose of multicentre trials is to ensure faster patient enrolment. Lin (2000) argues that there is a limited statistical penalty to

pay if we use as many centres as logistically required. He believes that exposing the drug to more centres and physicians is worth the additional cost.

In this section the effect of increasing N on the $\overline{MSE}$ of the estimators will be studied, again using simulation. This is achieved by fixing the rate of enrolment at each centre to be $\lambda_i = \frac{1\,000}{N}$. In other words, we keep the mean number of patients per treatment arm fixed, and equal to 1,000. The values of the parameters σ, σ_τ, and σ_μ remain the same as stated in Section 5, that is $\sigma = 1$, $\sigma_\tau = 0.25$, and $\sigma_\mu = 0.25$. The simulation results are plotted in Figure 17.

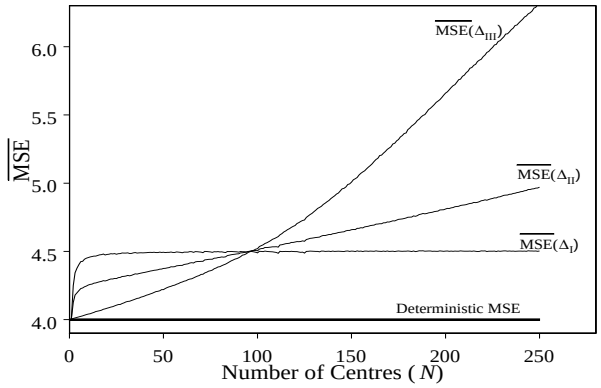

FIGURE 17. Plot of mean $\overline{MSE}$ for the estimators against different values of N, in Scenario I.

It is evident that the effect of increasing N leads to an increase in the $\overline{MSE}$ of the Type II and Type III estimators. However the $\overline{MSE}$ of the Type I estimator remains fairly constant when the number of centres exceeds approximately 10. Looking back to Figure 16 (b) we see that the imbalance term for the Type II estimator behaves in a same manner, initially increasing rapidly with N, but then stabilising for $N > 10$.

The behaviour of the Type I estimator as N increases is due to the stability of the estimator, with the variation attributable to centre and treatment-by-centre effects becoming settled for values of N greater than 10. However, this is not so with $\overline{MSE}(\Delta_{II})$ and $\overline{MSE}(\Delta_{III})$. The former increases due to the weights becoming unstable as a result

of fewer patients attending each centre, and the latter due to greater variation in the estimators of treatment effect δ_i. Similar behaviour is evident in Scenario II (figure not shown).

7. Drop-out distribution (Scenario III)

It is common in multicentre clinical trials for some patients to drop out of the trial for various reasons, which may or may not be related to treatment. We shall refer to the resulting distribution of patient numbers as the drop-out distribution as defined earlier in Section 3.

In the following, and again using simulation, the $\overline{MSE}$ of the three estimators will be studied in the presence of different values of p for Cases I – III, to determine the effect that an increasing proportion of patients withdrawing from a trial has on $\overline{MSE}(\Delta_I)$, $\overline{MSE}(\Delta_{II})$ and $\overline{MSE}(\Delta_{III})$.

Figures 18 to 20 show plots of $\overline{MSE}$ for increasing values of p, for the three types of estimator under Scenario III. As before, these figures were obtained using simulation.

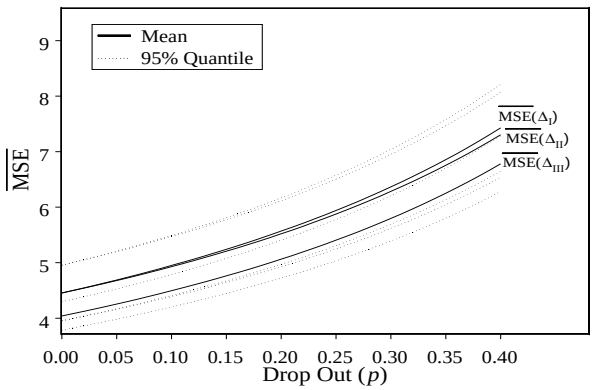

Figure 18. Plot of mean $\overline{MSE}$ for the estimators against different values of p in Case I, Scenario III.

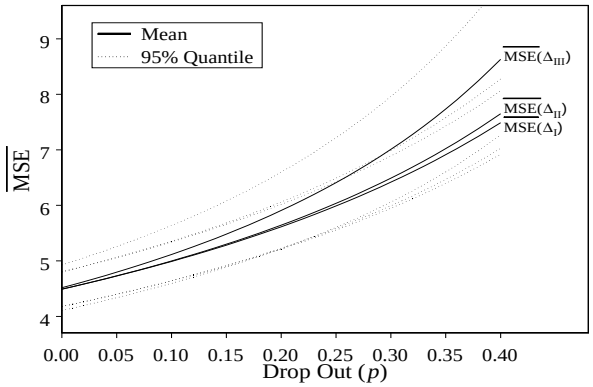

FIGURE 19. Plot of mean $\overline{MSE}$ for the estimators against different values of p in Case II, Scenario III.

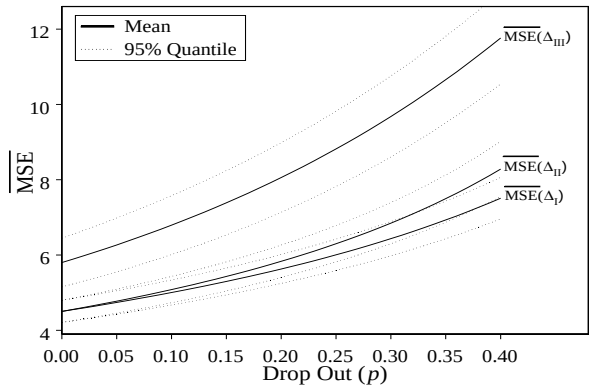

FIGURE 20. Plot of mean $\overline{MSE}$ for the estimators against different values of p in Case III, Scenario III.

Looking at Figure 18, we see that an increasing drop-out rate in Case I inflates the $\overline{MSE}$s of the estimators. However, the effect appears to be similar for each of the estimators. This is a consequence of the centres having "stability in numbers" prior to patients dropping out; Case I has an average of 100 patients per treatment arm.

Looking at Figure 19 we see fairly similar results for Case II, except that the increasing drop-out rate is far more detrimental to the Type III estimator. This detrimental effect is also seen in Figure 20 for Case III, to an even greater extent.

8. Gamma distributed rates of enrolment (Scenario IV)

The enrolment processes studied so far have assumed that (in the absence of disbalance) the enrolment rate λ is the same for all centres. However, as noted in the Introduction, it is most likely that the enrolment rate varies over the centres, and, therefore it is more realistic to treat λ as a random variable. A reasonable distribution for λ is the gamma distribution. This type of enrolment was described in Section 3 and was referred to as Scenario IV.

In this scenario, the gamma distribution is regarded as being the *mixing distribution* for the Poisson. The unconditional distribution of this mixed Poisson process is well-known and is commonly referred to as the *negative binomial* distribution.

An interesting situation is to fix the mean of the gamma distribution and compare the values of the $\overline{MSE}$ for the three estimators (in each of the Cases), against different values of the variance of λ. This will give information on the robustness of the estimators in the presence of increasingly variable enrolment rates. This is done in Figures Figures 21 to 23.

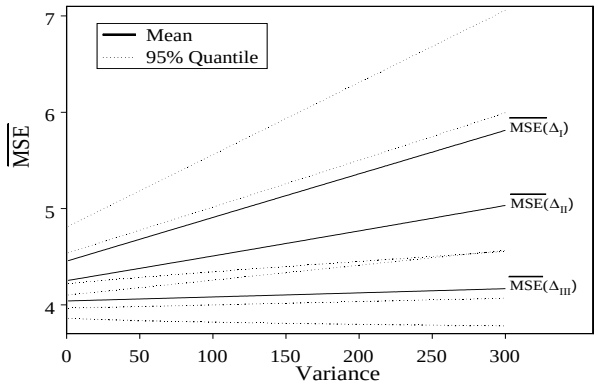

Figure 21. Plot of mean $\overline{MSE}$ for the estimators against different values of the variance for λ_i in Case I, Scenario IV.

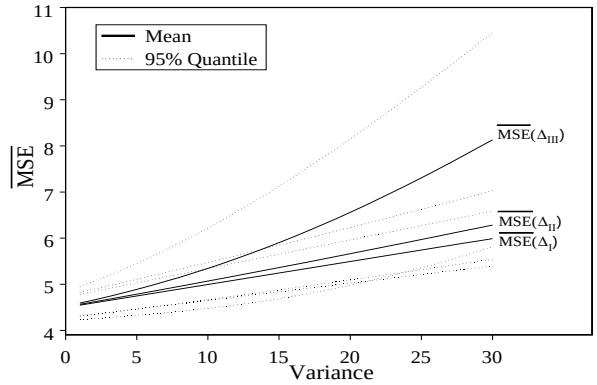

FIGURE 22. Plot of mean $\overline{MSE}$ for the estimators against different values of the variance for λ_i in Case II, Scenario IV.

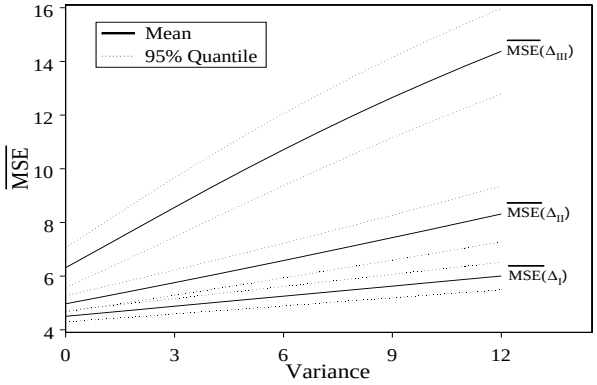

FIGURE 23. Plot of mean $\overline{MSE}$ for the estimators against different values of the variance for λ_i in Case III, Scenario IV.

Looking at these we see that increasing the variance of λ does not change the relative behaviour of the estimators (that is, the order of precedence remains the same), it simply causes the $\overline{MSE}$ to become inflated. For Case I it is $\overline{MSE}(\Delta_{III})$ that is the smallest. However, the reverse pattern is seen for Cases II and III, with the Type III estimator having the largest $\overline{MSE}$ over the whole range for the variance; it also increases at the fastest rate. It can also be seen in Figures 22 and 23 that $\overline{MSE}(\Delta_{II})$ increases at a faster rate than $\overline{MSE}(\Delta_I)$ under Cases II and III.

9. Conclusions

The results in Section 5 indicated that the distribution of the $\overline{MSE}$ was wider for Scenario II (equal numbers of patients on each arm) compared to Scenario I (independent sampling to each treatment arm). There was a clear distinction between the results for estimators Δ_I and Δ_{II} compared to Δ_{III}. In terms of minimizing the $\overline{MSE}$: (1) for Case I ($N = 10$, $\lambda = 100$) Δ_{III} was the best estimator, followed by Δ_{II} and then Δ_I; there was not a lot of difference in mean and standard deviation values for Δ_I and Δ_{II}, (2) for Case II ($N = 100$, $\lambda = 10$) Δ_I was the best, followed closely by Δ_{II}, with Δ_{III}, the worst.

When the value of σ_τ was increased under Scenario I the mean $\overline{MSE}$ values for both Δ_I and Δ_{II} increased, as might be expected, with Δ_I having the higher rate of increase. Interestingly, for Case III ($N = 250$, $\lambda = 4$), there was a notable dependency between the mean $\overline{MSE}$ values and σ_τ for estimators of Types I and II, which was caused by the more frequent occurrence of centres with no patients on one or both treatment arms.

In Section 4, and for deterministic enrolment, we saw that the model validity range increased in the presence of growing disbalance as we moved from Case I to Case III. In other words the use of Models I or II, even in the presence of treatment effects that vary over the centres, becomes more attractive as the number of centres increases (subject to a fixed total number of patients) and in the presence of increased variability in centre enrolment rates. In Section 5.2 we saw that when enrolment is random, and in the presence of disbalance, the model validity ranges were increased. That is, the boundary between the regions in the (σ_τ, d) plane where Model III is preferred to Models I and II occurred at higher values of σ_τ. Under random enrolment, minor disbalance in the centre enrolment rates was seen to have little effect on the mean $\overline{MSE}$ values of Δ_I and Δ_{II}. In the presence of a large disbalance in Case I, the mean $\overline{MSE}s$ of Δ_I and Δ_{II}increased considerably (from about 5 to about 45) compared to the situation with no disbalance and was consistently higher than the corresponding values for Δ_{III}. For the other two

cases the mean $\overline{MSE}$ of Δ_{III} was consistently higher than the corresponding values for Δ_I and consistently higher than the corresponding values for Δ_{II}.

Imbalance was taken as a measure of treatment-by-centre interaction and considered in Section 5.3. The size of the imbalance was shown to be strongly and positively related to the size of the disbalance. The imbalance contributed greatly to the size of the mean $\overline{MSE}$ when the disbalance was large. This effect was most apparent for Case I (N=10).

The effect of increasing the number of centres (N) on mean $\overline{MSE}$ was investigated in Section 6. It was seen that for Δ_I the mean $\overline{MSE}$ did not increase with N once N was greater than about 10. However the mean $\overline{MSE}$ of Δ_{III} increased rapidly when N was increased. The mean $\overline{MSE}$ of Δ_{II} also increased with N, but at a slower rate. These conclusions applied to both Scenarios I and II.

The effect of patient drop-out was considered in Section 7. For Case I ($N = 10$, $\lambda = 100$), the size of mean $\overline{MSE}$ for all three estimators increased with an increasing amount of drop-out. However, the mean $\overline{MSE}$ values for Δ_{III} were always below those of the other two estimators. The values for Δ_I and Δ_{II} were very similar, except when the drop-out probability exceeded about 0.10 when mean $\overline{MSE}$ for Δ_I became (slightly) larger. For Case II ($N = 100$, $\lambda = 10$) the results were similar to those for Case I for Δ_I and Δ_{II}. However, for Case III the values of mean $\overline{MSE}$ for Δ_{III} were always larger than those of the other two estimators and especially so when the drop-out probability exceeded about 0.15.

Varying enrolment rates were considered in Section 8, where the rates were sampled from a gamma distribution. Of interest was the effect on mean $\overline{MSE}$ of increasing the variance of the chosen gamma distribution. Increasing the variance increased the values of mean $\overline{MSE}$ but in a way that kept the rank ordering of the three estimators similar to the ordering they had when the Poisson sampling rates were kept constant over the

centres. That is, the mean $\overline{MSE}$ values for Δ_{III} were the largest and those for Δ_I were the smallest, with those of Δ_{II} being intermediate.

As a quick, and oversimplified, summary of the results we may say that Δ_{III} has attractive properties when the number of centres is small and the number of patients per centre is large. The simplest estimator Δ_I works well when treatment imbalance is high, or enrolment rates vary considerably or when there are many centres.

An important practical conclusion from this study is that Δ_I, the estimator of the CRT derived from the simplest model, performs better in some situations than the corresponding estimators derived from the other two more complex models.

Appendix A. Comparison of $\overline{MSE}\,(\Delta_I)$ under Scenario I and Scenario II.

From the simulation results described in Section 5 it was evident that variance of $\overline{MSE}\,(\Delta_I)$ was greater under Scenario II (when $n_{i1} = n_{i2} = n_i$) than Scenario I, see Table 1. Here we give an explanation for that result.

Consider $\overline{MSE}\,(\Delta_I)$ and denote the imbalance term $Imb(\Delta_I)$ by J, where

$$J \;=\; \sigma_\tau^2 \sum_{i=1}^{N} \left(\frac{n_{i2}}{n_{.2}} + \frac{n_{i1}}{n_{.1}} - \frac{2}{N} \right)^2 .$$

For Scenario I we have

$$J = J_1 = \sigma_\tau^2 \left[\sum_{i=1}^{N} \left(\frac{n_{i2}}{n_{.2}} - \frac{1}{N} \right)^2 + \sum_{i=1}^{N} \left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right)^2 + 2 \sum_{i=1}^{N} \left(\frac{n_{i2}}{n_{.2}} - \frac{1}{N} \right) \left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right) \right] .$$

For the case where $n_{i1} = n_{i2}$ (Scenario II), it can be shown that

$$J = J_2 = \sigma_\tau^2 \sum_{i=1}^{N} \left(\frac{2n_{i1}}{n_{.1}} - \frac{2}{N} \right)^2 = 4\sigma_\tau^2 \sum_{i=1}^{N} \left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right)^2 .$$

Calculating the expectation of J_1 and J_2 with respect to the distribution of n_{ij} gives:

$$\begin{aligned} E(J_1) &= 2\sigma_\tau^2 \left(\sum_{i=1}^{N} E\left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right)^2 + \sum_{i=1}^{N} \left[E\left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right) \right]^2 \right) , \\ E(J_2) &= 4\sigma_\tau^2 \sum_{i=1}^{N} E\left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right)^2 . \end{aligned}$$

Thus,

$$E(J_2) - E(J_1) = 2\sigma_\tau^2 \sum_{i=1}^{N} \left[E\left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right)^2 - \left(E\left(\frac{n_{i1}}{n_{.1}} - \frac{1}{N} \right) \right)^2 \right] > 0.$$

Similar results hold for the imbalance measure of the second estimator $Imb(\Delta_{II}) = 4\sigma_\tau^2 \sum_{i=1}^{N} \left(W_{IIi} - \frac{1}{N} \right)^2$.

References

[1] Chinchilli, V.M. and Bortey, E.B. (1991)Testing for consistency in a single multicenter trial. *Journal of Biopharmaceutical Statistics*, **1**, 67–80.

[2] Dragalin, V., Fedorov, V., Jones, B. and Rockhold, F. (2002) Estimation of the combined response to treatment in multicentre trials,. *Journal of Biopharmaceutical Statistics*, **11**, 275–295.

[3] Fedorov, V., Jones B. and Rockhold F. (2002) The design and analysis of multicentre trials in the random effect setting. *GSK BDS Technical Report 2002-03.*

[4] Fedorov, V.V., Montepiedra, G. and Nachtsheim, C.J. (1998) Optimal design and the model validity range. *J. Statist. Planning and Inference*, 72:215–227.

[5] Johnson N.L., Kotz S. and Balakrishnan, N. (1994) *Continuous Univariate Distributions, Volume 1.* Second Edition. John Wiley & Sons, New York.

[6] Johnson N.L., Kotz S. and Kemp A.W. (1993) *Univariate Discrete Distributions.* Second Edition. John Wiley & Sons, New York.

[7] Källén, A. (1997). Treatment-by-center interaction: what is the issue? *Drug Information Journal*, 31: 927–936.

[8] Lin Z. (2000) The number of centres in a multicentre clinical study: effects on statistical power. *Drug Information Journal*, **34**, 379–386.

[9] Schwemer G. (2000) General linear models for multicenter trials. *Controlled Clinical Trials*, **21**, 21–29.

[10] Senn S. (1997) *Statistical Issues in Drug Development*. John Wiley & Sons, Chichester.

[11] Snapinn M. S. (1998) Interpreting interaction: The classical approach. *Drug Information Journal*, **32**, 433–438.