Two-period, two-treatment crossover designs subject to non-ignorable missing data

JOHN N. S. MATTHEWS*, ROBIN HENDERSON

School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
john.matthews@ncl.ac.uk

SUMMARY

In common with most forms of designed experiment, crossover trials can be affected by missing data. Attempts to devise designs that can mitigate the possible effects of missing data, such as loss of efficiency, or even inestimability of certain contrasts, have been proposed. However, a potentially serious effect of missing data that has not been addressed in designs hitherto is that the treatment effects may be biased because of the nature of the missingness process. We investigate this problem in two-treatment, two-period crossover designs. In particular, we consider the robustness of the analysis under a missing at random assumption when, in fact, the data are non-ignorably missing. We show that the conventional AB/BA design still has good properties, although the design with sequences AB, BA, AA, and BB may be preferred if the chance of dropout depends primarily on the difference between the responses in the two periods.

Keywords: Crossover designs; Inverse probability weighting; Non-ignorable missing data.

1. INTRODUCTION

In many applications, designs are chosen to minimize the variance of the estimator of some key parameter. Investigators are also aware that in practice it is often not possible to collect all the observations specified by the design and that some data will be missing. It is widely appreciated that this will result in less precise estimates. Indeed, some experimental proposals, such as clinical trial protocols, will often contain statements such as ‘...a dropout rate of 10% is anticipated, so we aim to increase recruitment to allow for this’. While this is not unreasonable, it overlooks the fact that the data that are not collected may be systematically different from those that are, so missing data can induce bias of a magnitude that is difficult to predict. The extent to which some designs are more robust than others to the biasing effects of non-ignorable missing data does not seem to have been considered: we report an investigation in a biostatistical context.

We consider the design of two-treatment, two-period crossover designs when it is anticipated that some data will be missing. In doing so, we will adopt the missing data classification of Rubin (1976). Some attention has been paid to the issues of missing data in the design (Bose and Bagchi, 2008; Godolphin, 2004; Majumdar and others, 2008; Low and others, 1999) and analysis (Patel, 1985; Huang and Carrière, 1999) and others 2008; Low and others, 1999) and analysis (Patel, 1985; Huang and Carrière, 1999).
Crossover designs with missing data

Perhaps the most widely used crossover design is that which compares two treatments, $A$ and $B$, over two periods, using the sequences $AB$ and $BA$—the $AB/BA$ design. It has long been recognized that the $AB/BA$ design cannot provide an unbiased estimate of the treatment effect based on within-subject information if there is a carryover effect of treatment; see Jones and Kenward (2003, Chapter 2). A design-based approach to this problem was suggested by Balaam (1968), who pointed out that if the $AB/BA$ design were extended to include the sequences $AA$ and $BB$, then within-subject estimates of a direct treatment effect could be obtained even in the presence a carryover effect. However, Balaam’s design was able to achieve this because it assumed a particular model for the carryover effect which no longer enjoys wide support (Senn, 2002, Chapter 10). Indeed, it is now generally accepted that carryover effects should be excluded a priori, e.g. by using washout periods of sufficient duration. If this is possible, then the $AB/BA$ design is not only usable but also is the most efficient two-period, two-treatment design, and so Balaam’s design is seldom contemplated (but see Candel (2012) for a recent contribution).

Ho and others (2012), using the inverse-probability weighting (IPW) method of Robins and colleagues (Robins and others, 1995; Robins and Rotnitzky, 1995; Rotnitzky and others, 1998), considered, inter alia, the properties of an analysis of the $AB/BA$ design which assumed that data were missing at random (MAR) when, in fact, they were MNAR. If the data are MNAR, then the MAR analysis will be biased, but Ho and others (2012) found that the bias was considerably lower in certain parts of the parameter space of the dropout model than in others. Matthews and others (2012) found similar properties when they investigated the complete case analysis of the $AB/BA$ design. They also considered a two-period longitudinal study, which used only sequences $AA$ and $BB$, where they found that the bias of a MAR analysis was minimized in parts of the dropout parameter space where bias was large for the $AB/BA$ design.

To be more specific, suppose that the observations on subject $i$ in periods 1 and 2 are, respectively, $Y_i = (Y_{i1}, Y_{i2})^T$. Let the indicator $R_i = 1$ if $Y_i$ is fully observed and $R_i = 0$ otherwise. In this paper, we restrict attention to one pattern of missingness, namely when the observation in period 1 is available but that in period 2 is not—some comment on this restriction can be found in Section 5. Consequently, $R_i = 0$ corresponds to the case when only $Y_{i1}$ is observed. Suppose that the continuation probability is modeled as

$$\Pr(R_i = 1 \mid Y_{i1}, Y_{i2}) = p_i(\theta_0, \theta_2) = \expit(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2}),$$

(1.1)

where we use $\expit(u)$ to denote $e^u/(1 + e^u)$ and $\theta = (\theta_0, \theta_1)^T$. Data are MAR or MNAR as $\theta_2 = 0$ or $\theta_2 \neq 0$, respectively. Using this model, Matthews and others (2012) showed that the bias in the complete case analysis of the $AB/BA$ design was minimized when $\theta_1 - \theta_2 = 0$, whereas for the longitudinal $AA/BB$ design this occurred when $\theta_1 + \theta_2 = 0$. For the $AB/BA$ design, it was shown by Ho and others (2012) that the bias in a MAR analysis is small not only in the region close to $\theta_2 = 0$, but also in a region close to $\theta_1 - \theta_2 = 0$. The formulation in (1.1) corresponds to a selection model factorization, but it is possible to get an accurate approximation to $\Pr(R_i = 1)$ if a pattern mixture formulation is preferred: details are provided in supplementary material available at Biostatistics online.

2006; Grieve, 1995; Lee and others, 2005) of crossover trials. The papers on design have assumed that the data will be missing completely at random (MCAR). The possibility that data are missing not at random (MNAR) has not been considered in work on design and has only recently been addressed in the context of analysis (Basu and Santra, 2010; Ho and others, 2012). There has also been consideration of the properties of analyzing only the complete cases, i.e. those patients who provide observations in both periods (Liu, 2011; Matthews and others, 2012). Nason and Follmann (2010) and Makubate and Senn (2010) describe applications of crossover designs where incomplete data arise for different reasons from those considered here.
There may often be good reason to believe that data missing from a crossover design will be MNAR—perhaps because patients who have a poorer experience in the second treatment period are more inclined to leave the study. Analyses taking this into account are challenging because some parameters, such as $\theta_2$, are generally inestimable. Although Ho and others (2012) made a proposal to address this difficulty, a MAR analysis remains important and the circumstances under which it would provide an acceptable approach warrant investigation. This paper considers a general two-treatment, two-period crossover design in which all sequences, $AA$, $BB$, $AB$, and $BA$ might be used. Although we assume throughout that there is no carryover effect, comments in the previous paragraph indicate that the use of the sequences $AA$ and $BB$ may have a beneficial role in providing a design that is more robust, at least in terms of bias, to departures from the MAR model. However, attention also needs to be paid to the efficiency of any design that is proposed, as good bias properties will be vitiated if too high a penalty is paid in terms of efficiency.

In the next section, the approach to the analysis of a two-treatment, two-period crossover trial is outlined and the key formula for the estimator of the treatment effect is derived. In Section 3, some asymptotic limits are presented which allow the properties of the MAR analysis to be elucidated when the data are MNAR. In Section 4, the bias and efficiency properties of a range of designs are given and compared, and some general remarks are made in Section 5. Some technical results are available in supplementary material available at Biostatistics online.

2. OUTLINE OF ANALYSIS

We suppose that there are $N$ patients in total, allocated between the four two-period sequences $AB$, $BA$, $AA$, and $BB$. The treatment allocated to patient $i = 1, \ldots, N$ in period $j = 1, 2$ is $t(i, j) = A$ or $B$. We further suppose that

$$E(Y_{ij}) = \mu + \pi_j + \tau_{t(i,j)},$$

(2.1)

where $\mu$ is a general mean, the $\pi_j$ denote period effects, and the $\tau_t$ denote treatment effects. To ensure identifiability, we assume $\pi_1 = -\pi_2 = \pi$ and $\tau_A = -\tau_B = \tau$. We also assume that carryover effects of treatment can be excluded.

A flexible method for the analysis of a crossover trial, which requires no distributional assumptions and which fits well with the IPW approach to missing data, is to use a generalized estimating equation (GEE) (Liang and Zeger, 1986). If we write $\beta$ for the vector $(\mu, \pi, \tau)^T$, and $X_i$ for the rows of the design matrix implied by (2.1) for patient $i$, then a GEE that is valid, provided the data are MAR, is

$$\sum_{i=1}^N \frac{R_i}{p_i(\theta)} \left[ X_i^T (S^{1/2} C S^{1/2})^{-1} (Y_i - X_i \beta) \right] = 0,$$

(2.2)

where we have written $p_i(\theta)$ for $p_i(\theta, 0)$, the continuation probability if we assume that the data are MAR. Here $S$ is a diagonal $2 \times 2$ matrix of the variances of $Y_{i1}$ and $Y_{i2}$ and $C$ is a $2 \times 2$ working correlation matrix with $C_{12} = w$. In the interests of simplicity, and in keeping with the models usually used for crossover data, we assume that $\text{var}(Y_{ij}) = \sigma^2$ for all $i$ and $j$, so we can, without loss of generality, take $S$ to be the identity matrix. With these assumptions, (2.2) can be solved for $\beta$, giving

$$\hat{\beta} = \left( \sum_{i=1}^N \frac{R_i}{p_i(\hat{\theta})} X_i^T C^{-1} X_i \right)^{-1} \left( \sum_{i=1}^N \frac{R_i}{p_i(\hat{\theta})} X_i^T C^{-1} Y_i \right),$$

(2.3)

where the estimate $\hat{\theta}$ is obtained from a logistic regression of the $R_i$ on $Y_{i1}$.
2.1 Explicit estimator of $\tau$ for a two-period crossover design

Progress can be made with the analytic evaluation of (2.3) because subjects allocated to the same sequence share the same $X_i$. Denote the $2 \times 3$ design matrix for sequence $rs$, where $r, s \in \{A, B\}$, by $X_{rs}$ and write $\omega_{rs} = \sum_{i \in rs} R_i / p_i(\hat{\theta})$ (with the convention that $\omega_{rs} = 0$ if no subjects are allocated to sequence $rs$). We also need to define

$$\tilde{S}_{rs} = \omega_{rs}^{-1} \sum_{i \in rs} \frac{R_i}{p_i(\hat{\theta})} (Y_{i1} + Y_{i2}), \quad \tilde{D}_{rs} = \omega_{rs}^{-1} \sum_{i \in rs} \frac{R_i}{p_i(\hat{\theta})} (Y_{i1} - Y_{i2}).$$

With these definitions, we can evaluate (2.3), although our interest is focussed on the final element, $\hat{\tau}$. This can be written in the form

$$\hat{\tau} = \sum_{rs} \frac{g_{rs}^+ \omega_{rs}}{G} \tilde{S}_{rs} + \sum_{rs} \frac{g_{rs}^- \omega_{rs}}{G} \tilde{D}_{rs},$$

(2.4)

where $G = 2[(1 - w)(P^2 + PQ - E^2) + (1 + w)(Q^2 + PQ - F^2)]$ and where we have written $P = \omega_{AA} + \omega_{BB}$, $Q = \omega_{AB} + \omega_{BA}$, $E = \omega_{AA} - \omega_{BB}$, and $F = \omega_{AB} - \omega_{BA}$. The coefficients $g_{rs}^+$ and $g_{rs}^-$ are given in Table 1. Similar formulae are given by Carri`ere and Reinsel (1992) when there are no missingness can be based on $G$ and equal numbers of patients are allocated to $AA$ and $BB$ and equal numbers to $AB$ and $BA$. Note that, for the $AB/BA$ design, $E = 0$, and so only the $\tilde{D}_{rs}$ feature in the estimator, something which applies to all designs if $w = 1$. For designs with equal weight on $AA/BB$ and $AB/BA$, i.e. $\omega_{AA} = \omega_{BB}$ and $\omega_{AB} = \omega_{BA}$, $E = F = 0$ so the estimator uses only the $\tilde{S}_{rs}$ from the $AA$ and $BB$ sequences and the $\tilde{D}_{rs}$ from $AB$ and $BA$. Of course, even if the intention was to balance sequences in this way, the missingness process means that in general all the $\tilde{S}_{rs}$ and $\tilde{D}_{rs}$ will feature in the estimator of $\tau$.

The variance of $\hat{\tau}$ cannot be computed exactly from (2.4) for a number of reasons, principally that the coefficients $G^{-1} g_{rs}^+ \omega_{rs}$ are random variables. However, simulation studies confirm that a good approximation can be based on

$$\text{var}(\hat{\tau}) \approx \sum_{rs} \frac{g_{rs}^+ \omega_{rs}^2}{G^2} \text{var}(\tilde{S}_{rs}) + \sum_{rs} \frac{g_{rs}^- \omega_{rs}^2}{G^2} \text{var}(\tilde{D}_{rs}) + 2 \sum_{rs} \frac{g_{rs}^+ g_{rs}^- \omega_{rs}^2}{G^2} \text{cov}(\tilde{S}_{rs}, \tilde{D}_{rs}),$$

(2.5)

in which the variability in the coefficients and in $\hat{\theta}$ are neglected. The approximation, based on the delta method and given in more detail in supplementary material available at Biostatistics online, works well, provided that the proportion of missing data expected on any sequence does not exceed about 75% and $N \geq 25$.
It follows that the coefficients, $a_{rs}$, can be written as

$$a_{rs} = \lambda_{rs} \omega_{rs},$$

where $\omega_{rs}$ is the function of the weak law of large numbers says that, in probability

$$\frac{\omega_{rs}}{n_{rs}} = \frac{1}{n_{rs}} \sum_{i \in r,s} \frac{R_i}{p_i(\theta^*)} \to \mathbb{E} \left( \frac{R_i}{p_i(\theta^*)} \mid i \in r,s \right) = \chi_{rs}, \quad \text{say},$$

where $\theta^*$ is the least false value of $\hat{\theta}$, the fitted value assuming a MAR model, when the data are MNAR. It follows that the coefficients, $G^{-1} g_{rs}^\tau \omega_{rs}$ in (2.4) tend to $\lambda_{rs} \chi_{rs} a_{rs}^\tau$, where $a_{rs}^+$ and $a_{rs}^-$ are the functions of the $\lambda_{rs}$ and $\chi_{rs}$ given in Table 2.

The calculation of the bias in (2.4) is facilitated by writing $Z_{ij} = Y_{ij} - \mathbb{E}(Y_{ij})$ and $\mu_{rsj}$ for the mean response in period $j$ of sequence $rs$. Hence,

$$\tilde{S}_{rs} = (\mu_{rs1} + \mu_{rs2}) + \frac{n_{rs}^{-1} \sum_{i \in r,s} U_i^+}{n_{rs}^{-1} \omega_{rs}},$$

$$\tilde{D}_{rs} = (\mu_{rs1} - \mu_{rs2}) + \frac{n_{rs}^{-1} \sum_{i \in r,s} U_i^-}{n_{rs}^{-1} \omega_{rs}},$$

where $U_i^\pm = (Z_{i1} \pm Z_{i2}) R_i / p_i(\hat{\theta})$. Substituting the above into (2.4) and noting that $G^{-1}[\sum_{rs} g_{rs}^+ \omega_{rs}(\mu_{rs1} + \mu_{rs2}) + \sum_{rs} g_{rs}^- \omega_{rs}(\mu_{rs1} - \mu_{rs2})] = \tau$, we find that $\hat{\tau} - \tau$ is

$$\sum_{rs} \frac{g_{rs}^+ \omega_{rs}}{G} \frac{\tilde{U}_{rs}^+}{n_{rs}^{-1} \omega_{rs}} + \sum_{rs} \frac{g_{rs}^- \omega_{rs}}{G} \frac{\tilde{U}_{rs}^-}{n_{rs}^{-1} \omega_{rs}},$$

where $\tilde{U}_{rs}^\pm$ are the means of the $U_i^\pm$ for $i \in r,s$. A further application of the weak law of large numbers shows that in probability $\tilde{U}_{rs}^\pm \to \mathbb{E}(U_i^\pm \mid i \in r,s)$ and therefore that the asymptotic bias in $\hat{\tau}$ is

$$B = \sum_{rs} \lambda_{rs} a_{rs}^+ \mathbb{E}(U_i^+ \mid i \in r,s) + \sum_{rs} \lambda_{rs} a_{rs}^- \mathbb{E}(U_i^- \mid i \in r,s).$$

### Table 2. The values of $Aa_{rs}$

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$Aa_{rs}^+$</th>
<th>$Aa_{rs}^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AA$</td>
<td>$(1 - w)(2\lambda_{BB} \chi_{BB} + \lambda_{AB} \chi_{AB} + \lambda_{BA} \chi_{BA})$</td>
<td>$- \lambda_{BB} + \lambda_{AB} \chi_{BA} + \lambda_{BA} \chi_{BB}$</td>
</tr>
<tr>
<td>$BB$</td>
<td>$-(1 - w)(\lambda_{AA} \chi_{AA} + \lambda_{AB} \chi_{AB} + \lambda_{BA} \chi_{BA})$</td>
<td>$-(1 - w)(\lambda_{AB} \chi_{BA} + \lambda_{BA} \chi_{BB})$</td>
</tr>
<tr>
<td>$AB$</td>
<td>$-(1 - w)(\lambda_{AA} \chi_{AA} - \lambda_{BB} \chi_{BB})$</td>
<td>$(1 + w)(\lambda_{AB} \chi_{BA} + \lambda_{AA} \chi_{AA} + \lambda_{BA} \chi_{BB})$</td>
</tr>
<tr>
<td>$BA$</td>
<td>$-(1 - w)(\lambda_{AA} \chi_{AA} - \lambda_{BB} \chi_{BB})$</td>
<td>$-(1 - w)(\lambda_{AB} \chi_{BA} + \lambda_{AA} \chi_{AA} + \lambda_{BB} \chi_{BB})$</td>
</tr>
</tbody>
</table>

To obtain the coefficients $a_{rs}^\pm$, the above entries must be divided by $A = 4[2(1 - w)\chi_{AA} \chi_{BB} \lambda_{AB} \lambda_{BB} + 2(1 + w)\chi_{AB} \chi_{BA} \lambda_{AB} \lambda_{BA} + (\chi_{AA} \lambda_{AA} + \chi_{BB} \lambda_{BB})(\chi_{AB} \lambda_{AB} + \chi_{BA} \lambda_{BA})].$
A similar exercise can be undertaken to evaluate the limiting value of the variance given in (2.5). Some algebra leads to

$$\text{var}(\hat{t}) \rightarrow \frac{1}{N} \sum_{rs} \lambda_{rs} \left( a_{rs}^2 V_{rs}^+ + a_{rs}^2 V_{rs}^- + 2a_{rs}^+ C_{rs} \right), \quad (3.5)$$

where, for $i \in rs$,

$$V_{rs}^+ = \text{var}(U_i^+) + \frac{E(U_i^+)^2 \text{var}(R_i/p_i)}{\chi_{rs}^2} - 2 \frac{E(U_i^+ \text{cov}(U_i^+, R_i/p_i)}{\chi_{rs}}$$

$$V_{rs}^- = \text{var}(U_i^-) + \frac{E(U_i^-)^2 \text{var}(R_i/p_i)}{\chi_{rs}^2} - 2 \frac{E(U_i^- \text{cov}(U_i^-, R_i/p_i)}{\chi_{rs}}$$

$$C_{rs} = \text{cov}(U_i^+, U_i^-) + \frac{E(U_i^+ E(U_i^-) \text{var}(R_i/p_i)}{\chi_{rs}^2} - \frac{E(U_i^- \text{cov}(U_i^+, R_i/p_i)}{\chi_{rs}} - \frac{E(U_i^+ \text{cov}(U_i^-, R_i/p_i)}{\chi_{rs}}.$$

The above have been obtained using the delta method to approximate the variances and covariance of $\tilde{S}_{rs}$ and $\tilde{D}_{rs}$.

In order to proceed, we need to evaluate the moments of $U_i^\pm$ and $R_i/p_i$ in the above expressions. While our proposed method of analysis makes no distributional assumptions, we can obtain substantial insight into the properties of these designs using an analytic approach in which we assume that the $Y_i$ are from bivariate normal distributions. We will assume that the distributions share the same dispersion $\Omega$, which has common variance $\sigma^2$ and correlation $\rho$. The moments are evaluated at $\theta^*$ as defined following (3.1). Equations for $\theta^*$ are in Section 3.2 and the further calculations needed for (3.4) and (3.5) are in supplementary material available at Biostatistics online.

### 3.2 Least false value of $\hat{\theta}$

The least false value $\theta^*$ minimizes the Kullback–Liebler divergence between the binomial distribution for $R_i$ with fitted success probability $\expit(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2})$ from that with the true success probability $\expit(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2})$ (Claeskens and Hjort, 2008, p. 25). This leads to the following equations for $\theta^*$:

$$\sum_{rs} \lambda_{rs} E[\expit(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2}) | i \in rs] = \sum_{rs} \lambda_{rs} E[\expit(\theta_0^* + \theta_1^* Y_{i1}) | i \in rs],$$

$$\sum_{rs} \lambda_{rs} E[Y_{i1} \expit(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2}) | i \in rs] = \sum_{rs} \lambda_{rs} E[Y_{i1} \expit(\theta_0^* + \theta_1^* Y_{i1}) | i \in rs].$$

Approximate analytic expressions for these expectations can be obtained by using the extended skew-normal distribution and the approximation $\expit(u) \approx \Phi(cu)$, where $c = (16 \sqrt{3})/(15 \pi)$ (Johnson and others, 1995, p. 119) and $\Phi(c)$ denotes the distribution function of a standard normal variable: details are given in Ho and others (2012). The approximation to the first of the above equations is

$$\sum_{rs} \lambda_{rs} \Phi \left( \frac{c \Theta_{rs}}{k_0} \right) \approx \sum_{rs} \lambda_{rs} \Phi \left( \frac{c \Theta_{rs}^*}{k_0^*} \right), \quad (3.6)$$

where $k_0 = \sqrt{1 + c^2 \sigma^2 (\theta_1^2 + \theta_2^2 + 2 \rho \theta_1 \theta_2)}$ and $k_0^* = \sqrt{1 + c^2 \sigma^2 \theta_1^{*2}}$. The quantities $\Theta_{rs}$ and $\Theta_{rs}^*$ are $\theta_0 + \theta_1 \mu_{rs1} + \theta_2 \mu_{rs2}$ and $\theta_0^* + \theta_1^* \mu_{rs1}$, respectively.
The same approach applied to the second equation, together with (3.6), gives
\[
\tau \sum_{rs} \lambda_{rs} \Phi \left( \frac{c(\Theta_{rs}^* k_0)}{\lambda_{rs} \Phi} \right) (-1)^{[rs]} + \frac{c}{\lambda_{rs} \Phi} \sum_{rs} \lambda_{rs} \Phi \left( \frac{c(\Theta_{rs}^* k_0)}{\lambda_{rs} \Phi} \right) = \tau \sum_{rs} \lambda_{rs} \Phi \left( \frac{c(\Theta_{rs}^* k_0)}{\lambda_{rs} \Phi} \right) (-1)^{[rs]} + \frac{c}{\lambda_{rs} \Phi} \sum_{rs} \lambda_{rs} \Phi \left( \frac{c(\Theta_{rs}^* k_0)}{\lambda_{rs} \Phi} \right),
\]
(3.7)
where \((-1)^{[rs]}\) is 1 if \(rs\) is \(AA\) or \(AB\) and -1 if \(rs\) is \(BB\) or \(BA\) and \(\phi(\cdot)\) is the standard normal density. Equations (3.6) and (3.7) can be solved numerically for \(\theta^*\).

4. Robust low-bias and efficient designs

4.1 Variation of bias with \(\theta_1\) and \(\theta_2\)

The variation of the bias of \(\hat{\theta}\) with the dropout parameters can be illustrated by plotting (3.4) against \(\theta_1\) and \(\theta_2\). This was done for both the \(AB/BA\) design and the Balaam design. We consider \(\pi = 0, \mu = 0.5, \sigma^2 = 1\) and \(\tau = 0.5, 1, \rho = \frac{1}{5}, \frac{2}{5}\). For the dropout model, we restrict attention to \(-1 < \theta_1, \theta_2 < 1\) and \(\theta_0 = 0.5\); we also considered \(\theta_0 = 1\), which gave similar results to those for \(\theta_0 = 0.5\) (results not shown). With these values for the parameters, the probability of continuation, \(Pr(R_i = 1)\), varies between 0.42 and 0.80 across the above region for \((\theta_1, \theta_2)\).

The bias (and also variance) for the \(AB/BA\) design does not depend on the working correlation, \(\rho\), as \(\lambda_{AA} = \lambda_{BB} = 0\); cf. Table 2. For the Balaam design, we consider \(w = \frac{1}{2}\), to reflect the belief, common in the context of crossover designs, that observations on the same individual will exhibit substantial positive correlation.

In Figure 1, we see that for all designs the bias is zero or close to zero when \(\theta_2\) is close to zero, as would be anticipated, and is also low for the region near to the line \(\theta_1 = \theta_2\) (as was demonstrated for the \(AB/BA\) design by Ho and others, 2012). However, away from these areas there can be noticeable bias, with that for the \(AB/BA\) design being larger than that for the Balaam design. This can also be seen in Table 3, where values for the maximum and minimum bias are further from zero for the \(AB/BA\) design than for the Balaam design.

For most purposes comparisons between the designs cannot be made solely in terms of bias, as some account must be taken of the resulting \(\text{var}(\hat{\theta})\). As the estimates \(\hat{\theta}\) are biased, comparisons between designs are done in terms of the RMSE. In Figure 2, we show the ratios of the RMSEs between the Balaam design and the \(AB/BA\) design for \(N = 50\) and \(N = 100\) and for \(\rho = \frac{1}{3}\) and \(\rho = \frac{2}{3}\), where \(\theta_0 = 0.5\) and \(\tau = 1\). In Figure 3, we compare the \(AB/BA\) design with a design Inter, intermediate between the Balaam design and the \(AB/BA\) design, namely one where \(\lambda_{AA} = \lambda_{BB} = \frac{1}{5}\) and \(\lambda_{AB} = \lambda_{BA} = \frac{3}{8}\). This design is chosen because: (i) it is dual-balanced, i.e. it is invariant under relabeling of the treatments, a property known to be desirable for crossover designs (Matthews, 1987) and (ii) \(\lambda_{AA} = \frac{1}{5}\) is intermediate between the value for the \(AB/BA\) design, \(\lambda_{AA} = 0\), and for the Balaam design, \(\lambda_{AA} = \frac{1}{4}\).

Figures 2 and 3 show similar patterns. The \(AB/BA\) design has lower RMSE for most parts of the \((\theta_1, \theta_2)\)-space but there are parts where the Balaam design is superior. These are the parts of the space where \(|\theta_1|\) and \(|\theta_2|\) are close to 1 and of opposite sign. The same applies to the comparison between the \(AB/BA\) design and the intermediate design, although in the areas where the \(AB/BA\) design is superior, its superiority is, as one might expect, less marked than in the comparison with the Balaam design. The Balaam and intermediate designs are superior over a larger part of the \((\theta_1, \theta_2)\)-space when \(\rho = \frac{1}{3}\) than when \(\rho = \frac{2}{3}\). This is probably because, for larger \(\rho\), \(\text{var}(\hat{\theta})\) will be smaller for the \(AB/BA\) design due to the greater weight this design gives to the \(D_i\) relative to the \(S_i\).
Fig. 1. Bias values for $\theta_0 = 0.5$, as percentage of $\tau$: (a) $AB/BA$, $\rho = \frac{1}{3}$, $\tau = 0.5$; (b) $AB/BA$, $\rho = \frac{2}{3}$, $\tau = 0.5$; (c) Balaam, $w = \frac{1}{2}$, $\rho = \frac{1}{3}$, $\tau = 0.5$; (d) Balaam, $w = \frac{1}{2}$, $\rho = \frac{2}{3}$, $\tau = 0.5$; (e) $AB/BA$, $\rho = \frac{1}{3}$, $\tau = 1$; (f) $AB/BA$, $\rho = \frac{2}{3}$, $\tau = 1$; (g) Balaam, $w = \frac{1}{2}$, $\rho = \frac{1}{3}$, $\tau = 1$; and (h) Balaam, $w = \frac{1}{2}$, $\rho = \frac{2}{3}$, $\tau = 1$. 
Table 3. The minimum and maximum bias over the region $-1 < \theta_1, \theta_2 < 1, \theta_0 = 0.5$, expressed as a percentage of $\tau$

<table>
<thead>
<tr>
<th>Design</th>
<th>$\tau = 0.5$</th>
<th>$\tau = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = \frac{1}{3}$</td>
<td>$\rho = \frac{2}{3}$</td>
</tr>
<tr>
<td>$AB/BA$</td>
<td>$\begin{array}{cc} -15.6 &amp; 6.4 \ -9.8 &amp; 5.0 \end{array}$</td>
<td>$\begin{array}{cc} -11.8 &amp; 2.7 \ -7.2 &amp; 1.8 \end{array}$</td>
</tr>
<tr>
<td>Balaam $w = \frac{1}{2}$</td>
<td>$\begin{array}{cc} -10.4 &amp; 2.2 \ -6.6 &amp; 1.4 \end{array}$</td>
<td>$\begin{array}{cc} -8.3 &amp; 1.4 \ -5.0 &amp; 0.3 \end{array}$</td>
</tr>
</tbody>
</table>

Table 4 presents the maximum RMSE values over the same $(\theta_1, \theta_2)$-region as is plotted in Figures 2 and 3, for $N = 100$ and for the $AB/BA$, Balaam and Inter designs (the last two with $w = \frac{1}{2}$). This shows that the design with the smallest maximum RMSE (a form of minimax design) occurs with either the Balaam or Inter design in two of the four cases shown. While such a criterion, depending as it does on poor properties that may obtain only in small regions of the $(\theta_1, \theta_2)$-space, is unlikely to be decisive, it illustrates that once the effects of data MNAR are taken into account, the choice of design is less clear-cut.

Fig. 2. Ratio of RMSE Balaam: $AB/BA$ ($w = 0.5, \tau = 1, \theta_0 = 0.5$): values greater than 1 favor $AB/BA$: (a) $N = 50$, $\rho = \frac{1}{3}$; (b) $N = 50$, $\rho = \frac{2}{3}$; (c) $N = 100$, $\rho = \frac{1}{3}$; and (d) $N = 100$, $\rho = \frac{2}{3}$.
Crossover designs with missing data

Fig. 3. Ratio of RMSE Inter: $AB/BA$ ($w = 0.5$, $\tau = 1$, $\theta_0 = 0.5$): values greater than 1 favor $AB/BA$: (a) $N = 50$, $\rho = \frac{1}{3}$; (b) $N = 50$, $\rho = \frac{2}{3}$; (c) $N = 100$, $\rho = \frac{1}{3}$; and (d) $N = 100$, $\rho = \frac{2}{3}$.

Table 4. The maximum RMSE ($\times 100$) for $N = 100$ over the region $-1 < \theta_1, \theta_2 < 1$, $\theta_0 = 0.5$

<table>
<thead>
<tr>
<th>Design</th>
<th>$\tau = 0.5$</th>
<th>$\tau = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = \frac{1}{3}$</td>
<td>$\rho = \frac{2}{3}$</td>
</tr>
<tr>
<td>$AB/BA$</td>
<td>12.7</td>
<td>8.7</td>
</tr>
<tr>
<td>Balaam $w = \frac{1}{2}$</td>
<td>17.4</td>
<td>13.0</td>
</tr>
<tr>
<td>Inter $w = \frac{1}{2}$</td>
<td>14.6</td>
<td>10.3</td>
</tr>
</tbody>
</table>

5. Discussion

Traditional comparisons of the $AB/BA$ design and the Balaam design, for models with no carryover effect of treatment (Carrière and Reinsel, 1992), show that the former is the design of choice. With no missing data and using generalized least squares with $\rho$ known to be $\frac{2}{3}$, the variance of $\hat{\tau}$ from the Balaam design is approximately 1.6 times larger than for the $AB/BA$ design. However, in practice not only are there likely to be missing data, but the possibility that they will be MNAR needs to be entertained. Moreover, if data are MNAR, the choice of design depends on parameters whose values are not known a priori nor can they be estimated from the trial data. Consequently, traditional approaches to non-linear design, in which a prior distribution is sought for the unknown parameters, may be especially challenging in these circumstances.
When judged solely in terms of bias, the Balaam design has markedly better properties than the AB/BA design. This is usually not the case when designs are compared with respect to RMSE, when the AB/BA design has superior properties. However, there are circumstances when the Balaam design is superior, namely when \((\theta_1, \theta_2)\) is close to the line \(\theta_1 = -\theta_2\). Here the dropout model (1.1) becomes \(\Pr(R = 1 | Y_1, Y_2) = \exp(\theta_0 + \theta_1(Y_1 - Y_2))\), i.e. the probability of dropping out depends on the contrast between the outcomes in the two periods. As patients may well dropout if they have a disappointing experience in the second period compared with the first, this is certainly a plausible region of the \((\theta_1, \theta_2)\)-plane in the context of a crossover trial. However, where the Balaam design is superior, it is because of its better bias properties and these will be outweighed in the RMSE by the contribution of \(\text{var}(\hat{\tau})\) unless \(N\) is large. It is rare for a crossover trial to have \(N > 100\) and values below 50 are quite common. For values of \(N\) that are likely in crossover trials, and in the absence of very specific information about the dropout parameters, the AB/BA design remains the one to choose.

Taking account of the possibility that data are MNAR in the design of a crossover experiment has led to a problem depending on many unknown parameters. The results presented here are typical of those we have obtained using a wider variety of treatment and period effects and different dropout parameters. Simulations confirmed the accuracy of the approximations used to obtain (3.4) and (3.5), provided that the dropout probabilities stayed below about 75\%. This is not a serious restriction because it is very unlikely that an experiment would proceed if such high dropout rates were anticipated. We have compared dual-balanced designs in which the proportions on \(AB/BA\) and \(AA/BB\) are \((1,0), (\frac{1}{2}, \frac{1}{2})\), and \((\frac{3}{4}, \frac{1}{4})\). A more sophisticated approach might be to seek the values for the \(\lambda_{rs}\) which gave, for example, optimal RMSE. This would certainly be possible if a way could be found to accommodate the other parameters in the problem, for example if a suitable prior could be specified. We have not pursued this at present.

As mentioned, investigators often anticipate that some of the data prescribed by their design will not be obtained, and they try to allow for this by increasing recruitment. In general, the allowance does not get more sophisticated than this, and the form of the design is not changed in any way. For crossover designs, there is some acknowledgement that subjects may be more likely to drop out from later periods, and some designs may be preferred with this in mind (Bose and Bagchi, 2008; Low and others, 1999). However, in many circumstances investigators will be mindful that it may not be realistic to assume that the missing data are MCAR or MAR. Our work shows that while the possibility of non-ignorable data can, in principle, mean that non-traditional design choices might be warranted, the reassuring result is that the \(AB/BA\) design remains the design of choice in most practical circumstances.

This paper has presented an initial investigation of the effects of non-ignorable missing data on the properties of two-period, two-treatment crossover designs. It has made a rather restrictive assumption that data are missing only in the second period. It is certainly the case that instances of data being available only in period 2 and not period 1 can arise and the extension to this case is of interest. However, this would require elaboration of the dropout model (1.1) and, in particular, care would need to be taken over the role of \(Y_{i2}\) in the model for the probability of being missing in the first period: relevant contributions can be found in Robins and others (1995). We have also based our investigation on an IPW approach implemented via a GEE, rather than on the widely used mixed-model analysis. The latter will have similar bias properties to those presented here but will lead to smaller values of the RMSE because additional information will be obtained from those cases where only \(Y_{i1}\) is observed. This could lead to changes in the relative merits of the Balaam and \(AB/BA\) designs with respect to the RMSE, at least for some values of \(N\): we hope to investigate this aspect in future work.

**Supplementary material**

Crossover designs with missing data

Acknowledgments

We thank the reviewers and Editors for their helpful comments. Conflict of Interest: None declared.

References


[Received June 4, 2012; revised February 4, 2013; accepted for publication February 20, 2013]