

Contrasting principal stratum and hypothetical strategy estimands in multi-period crossover trials with incomplete data

John N.S. Matthews^{1,2}  | Sofia Bazakou¹ | Robin Henderson¹ | Linda D. Sharples³

¹School of Mathematics, Statistics & Physics, Newcastle University, Newcastle upon Tyne, UK

²Public Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

³Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

Correspondence

John N.S. Matthews, School of Mathematics, Statistics, and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.

Email: john.matthews@ncl.ac.uk

Funding information

National Institute for Health Research, Grant/Award Number: RMOFS 2012/05

Abstract

Complete case analyses of complete crossover designs provide an opportunity to make comparisons based on patients who can tolerate all treatments. It is argued that this provides a means of estimating a principal stratum strategy estimand, something which is difficult to do in parallel group trials. While some trial users will consider this a relevant aim, others may be interested in hypothetical strategy estimands, that is, the effect that would be found if all patients completed the trial. Whether these estimands differ importantly is a question of interest to the different users of the trial results. This paper derives the difference between principal stratum strategy and hypothetical strategy estimands, where the former is estimated by a complete-case analysis of the crossover design, and a model for the dropout process is assumed. Complete crossover designs, that is, those where all treatments appear in all sequences, and which compare t treatments over p periods with respect to a continuous outcome are considered. Numerical results are presented for Williams designs with four and six periods. Results from a trial of obstructive sleep apnoea-hypopnoea (TOMADO) are also used for illustration. The results demonstrate that the percentage difference between the estimands is modest, exceeding 5% only when the trial has been severely affected by dropouts or if the within-subject correlation is low.

KEYWORDS

closed skew-normal distribution, complete-case analysis, crossover design, hypothetical strategy estimand, principal stratum strategy estimand, Williams design

1 | INTRODUCTION

Missing data are a common problem in clinical trials and this has led to the development of a range of analytical methods to address the issue: see, for example, Molenberghs and Kenward (2007). Further, there has been discussion of how the design and conduct of studies can

minimize the effects of missing data (National Research Council, 2010). A related issue is the importance of specifying the quantity the trial aims to estimate much more precisely than might have been the case in the past. Specifying the quantity to be estimated, the *estimand*, which most pertinently reflects the aims of the study has received substantial attention over the last decade. Much

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Biometrics* published by Wiley Periodicals LLC on behalf of International Biometric Society.

of this thinking appears in an addendum ICH E9(R1), (International Council for Harmonisation, 2019), to the ICH E9 guidelines on statistical principles in clinical trials.

Identifying the most pertinent estimand requires careful consideration of many features that may arise in the conduct of the study, such as inter-current events or patient withdrawal. Once an estimand has been identified, the design, data collection and analysis need to be chosen accordingly. The ICH E9(R1) addendum discusses several estimands, such as the Treatment Policy Strategy, which aligns closely with the intention to treat (ITT) principle or the hypothetical strategy, in which it is assumed that all patients adhere to the trial as planned.

An estimand that might have particular clinical relevance is the one which compares the effect of treatments among those patients who can tolerate them—a principal stratum strategy (PSS). In a trial comparing a new treatment, A, with the standard, S, this approach would compare those who receive, and tolerate, A with those receiving S who would have tolerated A had they received it (and vice versa). In a parallel group trial those allocated to S do not also receive A, and vice versa, so the relevant subgroups cannot be identified. However, crossover designs, allocating as they do several trial treatments to each patient, provide greater opportunities for the implementation of a PSS, because comparisons are within patients. For example, inference might be based on the population of patients who can tolerate all treatments being studied. This naturally leads to the use of complete crossover designs, where all treatments are present in all treatment sequences, and the use of a *complete case* (CC) analysis, which excludes all patients who do not provide a response in every treatment period.

This contrasts with a common analysis which uses restricted maximum likelihood (REML) to fit a linear mixed-effects model to all available data. With the assumption that data are missing at random (MAR) (Rubin, 1976), this will provide an unbiased estimate of a hypothetical strategy (HS) estimand.

While specification of a target estimand can clarify the aim of the study and guide the design and analysis, trials will often provide data of interest to a range of stakeholders, not all of whom may wish to focus on the same estimand. The designation of *primary* and *secondary* estimands has been discussed (Leuchs et al., 2015; Mallinckrodt et al., 2017): for example, clinicians and regulators may focus on a PSS estimand, while drug developers may favor an HS estimand. This paper addresses the difference between PSS and HS estimands for treatment effects in complete crossover trials, where the former is estimated using a CC analysis. This is done by computing the expectation of the treatment effect in a CC analysis in terms

of the treatment parameter for the HS estimand. The latter will be the expectation from the standard linear mixed effects analysis, if the missing data are MAR. The two would coincide if no patients dropped out, so the difference will inevitably depend on a model for the dropout probabilities. This is essentially the analysis presented for the AB/BA design in Matthews et al. (2014).

Other contributions to the issues that arise in crossover designs with missing data include the effect of data missing not at random (Ho et al., 2012), joint modeling of dropout and response (Basu & Santra, 2010; Wang & Chinchilli, 2021) and missingness being related to threshold exceedance (Liu, 2011). A different strand of research focuses on the fact that while the planned design, D , is assumed to be connected, that is, all treatment contrasts are estimable, the design that results after missing data are taken into account, \tilde{D} , may be disconnected (Bate et al., 2008; Godolphin, 2006; Godolphin & Godolphin, 2019; Low et al., 1999; Majumdar et al., 2008; Prescott & Mansson, 2001; Varghese et al., 2002). These papers take a different approach to the one in the present paper, as they assume that data are missing completely at random (MCAR) and analyze all available data. Nevertheless, our calculations allow us to comment on the issue of connectedness as it applies to the approach presented here.

In Section 2, the proposed analysis and associated notation is introduced, along with the form of the missingness process. In Section 3, the expectation of the CC estimate of the treatment effect, the PSS estimand, is evaluated in terms of the HS estimand. Section 4 presents numerical results for selected four- and six-period complete crossover designs based on Williams' squares. Results on disconnection in this context are given in Section 5 and discussion and possible extensions are in Section 6.

2 | ESTIMANDS, ESTIMATION, AND MISSING DATA

2.1 | Estimands and estimation

It is assumed that the design compares t treatments over p periods and that the outcome variable is continuous. In particular, if the response of patient i in period j is $Y_{i,j}$, then the model usually adopted for a crossover study is some variant of

$$Y_{i,j} = \pi_j + \tau_{d(i,j)} + \xi_i + \epsilon_{ij} \quad i = 1, \dots, n; j = 1, \dots, p. \quad (1)$$

Here, ξ_i and ϵ_{ij} are normally distributed random patient effects and residuals, with mean 0 and variances σ_B^2 and σ_W^2 , respectively. The effect of period j is π_j , the effect of treatment k , $k = 1, \dots, t$ is τ_k (with $\tau_1 = 0$ for identifiability)

and $d(i, j)$ denotes the treatment allocated to patient i in period j . We use π, τ to denote the vectors of the period and treatment effects, although the latter will usually require further qualification.

We also assume that there is no treatment by period interaction. In a crossover design, such a term is closely related to treatment carryover effects (in the AB/BA design the two are wholly confounded), so some practitioners may include a first-order carryover term in (1): indeed, much of the work on connectedness cited in Section 1 does do this. However, the adequacy of the usual form of carryover to account for any realistic mechanism by which the effect of a treatment might persist has been questioned, perhaps most strongly in Senn (2002, Chapter 10). It is now common to use (1) and justify the absence of carryover on non-statistical grounds, such as requiring observations in successive periods to be sufficiently separated in time: this is the stance taken in the present paper.

It is also assumed that each patient provides a single outcome in each period. Although patients may be observed on a number of occasions during a period, this assumption is probably less restrictive than might be imagined. The principal focus may be on the final observation in a period, at least partly to minimize the chance of a carryover effect, or the analysis may use a summary of the observations made on the patient during the period. If some of the components of this summary are missing, then whether the summary is deemed to be missing, or it is calculated using the available observations, is likely to depend on the context and the judgment of the analyst.

The residual terms in (1) give rise to an equicorrelation, or compound symmetry, structure for the dispersion of the vector of observations on patient i , \mathbf{Y}_i , where $\text{var}(Y_{i,j}) = \sigma^2 = \sigma_B^2 + \sigma_W^2$ and, for $j \neq j'$, $\text{cov}(Y_{i,j}, Y_{i,j'}) = \sigma_B^2 = \rho\sigma^2$, with ρ the intra-class correlation coefficient. The $p \times p$ matrix Ω is defined to have ones on the diagonal, with all other elements being ρ , so the dispersion matrix of a completely observed patient would be $\sigma^2\Omega$.

If model (1) was applied to the population where it is supposed that all patients provide results in every period, then τ would be an HS estimand, and to distinguish from other estimands is denoted by τ_{HS} . In practice, some patients may provide fewer than p observations and a widely used method of estimation (see, e.g., Jones and Kenward 2014, Chap. 5) is to apply REML to all available data. If the data were MAR, this method would give an unbiased estimate of τ_{HS} (Kackar & Harville, 1984).

If REML and Equation (1) are applied only to the population of patients who provide observation in all periods, that is, a CC analysis, then τ constitutes a PSS estimand, which we denote by τ_{PSS} . The expectation of the treatment estimator from this analysis, $\hat{\tau}_{\text{PSS}}$, will be τ_{PSS} , but this can

also be expressed in terms of τ_{HS} if we postulate a model for the missingness process which gives rise to the difference between the populations. The first step is to derive an expression for $\hat{\tau}_{\text{PSS}}$ found from a CC analysis.

The vector of observations, \mathbf{Y}_i , on any patient i included in the CC analysis will have dimension p . For this population, the $np \times (p + t - 1)$ design matrix for the model in Equation (1) is $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T, \dots, \mathbf{X}_n^T)^T$ and \mathbf{X}_i is the $p \times (p + t - 1)$ design matrix for patient i . It will be assumed that the design is composed of S distinct treatment sequences: the design matrix will be the same for all patients allocated to a given sequence, so we write $\mathbf{X}_s, s = 1, \dots, S$ for the design matrices corresponding to the S sequences. If patient i is allocated to sequence s , \mathbf{X}_i is identified with \mathbf{X}_s . While the design D allocates M_s patients to sequence s , only N_s of these are CCs. If the responses in \mathbf{Y}_i are ordered by time, then each \mathbf{X}_s will be of the form $(\mathbf{I}_p | \mathbf{T}_s)$, where \mathbf{I}_p , the $p \times p$ identity matrix, is associated with the period effects and \mathbf{T}_s is a $p \times (t - 1)$ matrix describing the treatment allocation in sequence s .

The responses on different patients are assumed to be independent, so the treatment estimator for the CC analysis, $\hat{\tau}_{\text{PSS}}$, can be expressed as

$$\hat{\tau}_{\text{PSS}} = \mathbf{C}^{-1} \sum_s N_s (\mathbf{T}_s - \bar{\mathbf{T}})^T \hat{\Omega}^{-1} \bar{\mathbf{Y}}_s, \quad (2)$$

where

$$\mathbf{C} = \sum_s N_s (\mathbf{T}_s - \bar{\mathbf{T}})^T \hat{\Omega}^{-1} (\mathbf{T}_s - \bar{\mathbf{T}}), \quad \bar{\mathbf{T}} = \sum_s N_s \mathbf{T}_s / \sum_s N_s. \quad (3)$$

Here, $\bar{\mathbf{Y}}_s$ is the mean of all the complete \mathbf{Y}_i allocated to sequence s , and $\hat{\Omega}$ is Ω evaluated at $\rho = \hat{\rho}$, the REML estimator of ρ . An important observation, demonstrated in Web Appendix A, is that for the CC analysis of complete crossover designs and Ω having compound symmetry, the matrices $\mathbf{C}^{-1} (\mathbf{T}_s - \bar{\mathbf{T}})^T \hat{\Omega}^{-1}, s = 1, \dots, S$ are invariant with respect to $\hat{\rho}$, and hence so is $\hat{\tau}_{\text{PSS}}$. Consequently, Equation (2) can be greatly simplified by setting $\hat{\rho}$ equal to 0, so that $\hat{\Omega}$ is replaced by \mathbf{I}_p .

In order to make progress with evaluating $E(\hat{\tau}_{\text{PSS}})$ in terms of τ_{HS} , it is necessary to specify more precisely how missing data arise.

2.2 | Missingness process

Attention will be restricted to monotonic missing patterns, so if a patient is missing in period i , they are also missing in period j for all $j > i$. Patients with no responses at all cannot be accommodated, so patients are assumed to drop out in any period after period 1 up to period $p - 1$, or not to drop out at all. If patient i drops out immediately after

period $j \geq 1$, we define the dropout indicator R_i to take the value j , so $R_i = p$ denotes a CC: for a study with $p = 4$ the possibilities are illustrated below, where o denotes an observed value and x is a missing value.

Period 1	Period 2	Period 3	Period 4	Dropout indicator
o	x	x	x	$R = 1$
o	o	x	x	$R = 2$
o	o	o	x	$R = 3$
o	o	o	o	$R = 4$

A model for the missingness process is required and this is often specified in terms of suitable logistic regressions on the outcomes, but subsequent calculations will be expedited if a probit link is assumed. We will use a model of the form

$$\Pr(R_i = j | R_i \geq j, \mathbf{Y}_i) = \Phi(\theta_{0j} + \theta_{1j}Y_{i,j+1} + \theta_{2j}Y_{i,j} + \dots + \theta_{j+1,j}Y_{i,1}), \quad j = 1, \dots, p - 1, \quad (4)$$

where $\Phi(\cdot)$ is the distribution function of a standard normal variable: note that $\Pr(R_i = p | R_i \geq p, \mathbf{Y}_i) = 1$. The models allow the probability of dropout to depend on all values previously observed on the patient and, when $\theta_{1j} \neq 0$, on the response that would have been obtained in the period that the patient dropped out, which corresponds to missing not at random (MNAR). The case when $\theta_{1j} = 0$ for all j corresponds to MAR, unless $\theta_{kj} = 0$ for all $k \geq 1$, which corresponds to MCAR. In practice, sub-models with many fewer parameters may be sufficient but the level of generality in Equation (4) will be maintained for now.

Calculations used in the investigation of expectation in Section 3 require an expression for the probability that patient i provides a CC, i.e. $\Pr(R_i = p | \mathbf{Y}_i)$. This can be built up sequentially from $\Pr(R_i = 1 | R_i \geq 1, \mathbf{Y}_i) = \Pr(R_i = 1 | \mathbf{Y}_i)$ and, for $j > 1$,

$$\Pr(R_i = j | \mathbf{Y}_i) = \Pr(R_i = j | R_i \geq j, \mathbf{Y}_i) \times \prod_{u=1}^{j-1} \{1 - \Pr(R_i = u | R_i \geq u, \mathbf{Y}_i)\}. \quad (5)$$

It follows from this and Equation (4) that the probability of a CC is

$$\Pr(R_i = p | \mathbf{Y}_i) = \prod_{j=1}^{p-1} \Phi(-\theta_{0j} - \theta_{1j}Y_{i,j+1} - \theta_{2j}Y_{i,j} - \dots - \theta_{j+1,j}Y_{i,1}). \quad (6)$$

In the next section, it will be convenient to rewrite this product using p -dimensional vectors θ_j , $j = 1, \dots, p - 1$,

defined using an appropriate pattern of zeros, so that Equation (6) becomes

$$\Pr(R_i = p | \mathbf{Y}_i) = \prod_{j=1}^{p-1} \Phi(-\theta_{0j} - \theta_j^T \mathbf{Y}_i). \quad (7)$$

3 | EXPECTATION OF THE TREATMENT EFFECT FROM A COMPLETE CASE ANALYSIS IN TERMS OF τ_{HS}

3.1 | Conditional expectation

Expression (2) for the CC estimator, $\hat{\tau}_{PSS}$, depends on the random variables N_s as well as \bar{Y}_s , so the first step is to evaluate $E(\hat{\tau}_{PSS} | N_1, \dots, N_S)$, which in turn requires the evaluation of $E(\bar{Y}_s | N_1, \dots, N_S)$ for each sequence s . This is shown to be $E(\mathbf{Y}_i(s) | R_i(s) = p)$ in Web Appendix B in the Supporting information, where we write $\mathbf{Y}_i(s)$ for \mathbf{Y}_i to emphasize that here we consider only those patients i allocated to sequence s : $R_i(s)$ is defined analogously. The unconditional mean of the observations on an individual allocated to sequence s is written as μ_s , and as the unconditional expectation is taken over the hypothetical population with no missing values, $\mu_s = \pi + T_s \tau_{HS}$. Consequently, $E(\mathbf{Y}_i(s) | R_i(s) = p)$ can be found as

$$\frac{\int \mathbf{y} \Pr(R_i(s) = p | \mathbf{y}) \phi_p(\mathbf{y}; \mu_s, \sigma^2 \Omega) d\mathbf{y}}{\int \Pr(R_i(s) = p | \mathbf{y}) \phi_p(\mathbf{y}; \mu_s, \sigma^2 \Omega) d\mathbf{y}}, \quad (8)$$

where $\phi_p(\cdot; \mu, \Sigma)$ is the density of a p -dimensional multivariate normal distribution. If Equation (7) is substituted in the above then the integrals can be evaluated using the closed skew-normal distribution (González-Farías et al., 2004).

To see this, we first recall that a random variable $\mathbf{U} \in \mathbb{R}^p$ has a closed skew-normal distribution, $CSN_{p,q}(\mu, \Sigma, \mathbf{G}, \nu, \Delta)$, if its density is

$$g(\mathbf{u}) = \frac{\phi_p(\mathbf{u}; \mu, \Sigma) \Phi_q(\mathbf{G}(\mathbf{u} - \mu); \nu, \Delta)}{\Phi_q(0; \nu, \Delta + \mathbf{G}\Sigma\mathbf{G}^T)}, \quad (9)$$

where (μ, Σ) and (ν, Δ) are, respectively, p - and q -dimensional means and dispersion matrices and \mathbf{G} is a $q \times p$ matrix. The distribution function of a q -dimensional multivariate normal distribution with mean ν and dispersion Δ is denoted by $\Phi_q(\cdot; \nu, \Delta)$. In order to evaluate (8) we need to identify the numerator in Equation (9) with $\Pr(R_i(s) = p | \mathbf{y}) \phi_p(\mathbf{y}; \mu_s, \sigma^2 \Omega)$: it will then follow that the denominator in Equation (8) must coincide with the denominator in Equation (9). This is achieved by setting $\mu = \mu_s$, $\Sigma = \sigma^2 \Omega$ and choosing \mathbf{G} , ν and Δ so that the

expression in Equation (7) can be identified with $\Phi_q(\mathbf{G}(\mathbf{u} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta})$.

If we choose $q = p - 1$ and $\boldsymbol{\Delta} = \mathbf{I}_{p-1}$ then $\Phi_q(\cdot; \boldsymbol{\nu}, \boldsymbol{\Delta})$ will become a product of $p - 1$ scalar normal distribution functions as in Equation (7). To reproduce the arguments in the factors in Equation (7) we need to define the j th row of \mathbf{G} to be $-\theta_j$. For example, in a four-period design where missingness depends only on the current and immediately preceding value (i.e., $\theta_{kj} = 0, k > 2$), then

$$\mathbf{G} = \begin{pmatrix} -\theta_{21} & -\theta_{11} & 0 & 0 \\ 0 & -\theta_{22} & -\theta_{12} & 0 \\ 0 & 0 & -\theta_{23} & -\theta_{13} \end{pmatrix}. \quad (10)$$

The identification is completed by noting that the j th element of $\boldsymbol{\nu}$, $j = 1, \dots, p - 1$, needs to be the unconditional expectation of $\theta_{0j} + \theta_j^T \mathbf{Y}_i(s)$: this is $\theta_{0j} + \theta_j^T \boldsymbol{\mu}_s = \Theta_{sj}$, say. It is convenient to write $\boldsymbol{\Theta}_s$ for the $p - 1$ dimensional vector of these quantities.

With these definitions, it follows that for sequence s

$$\begin{aligned} E(\mathbf{Y}_i(s) | R_i(s) = p) \\ = \int \mathbf{y} \frac{\phi_p(\mathbf{y}; \boldsymbol{\mu}_s, \sigma^2 \boldsymbol{\Omega}) \Phi_{p-1}(\mathbf{G}(\mathbf{y} - \boldsymbol{\mu}_s); \boldsymbol{\Theta}_s, \mathbf{I}_{p-1})}{\Phi_{p-1}(\mathbf{0}; \boldsymbol{\Theta}_s, \mathbf{I}_{p-1} + \sigma^2 \mathbf{G} \boldsymbol{\Omega} \mathbf{G}^T)} d\mathbf{y}, \end{aligned} \quad (11)$$

which is the expectation of a $\text{CSN}_{p,p-1}(\boldsymbol{\mu}_s, \sigma^2 \boldsymbol{\Omega}, \mathbf{G}, \boldsymbol{\Theta}_s, \mathbf{I}_{p-1})$ distribution. The moment generating function of the closed skew-normal distribution is available in González-Farías et al. (2004), and using this an expression for the mean is derived in Web Appendix C.

Provided the N_s values permit the inversion of \mathbf{C} , then

$$\begin{aligned} E(\hat{\boldsymbol{\tau}}_{\text{PSS}} | N_1, \dots, N_S) &= \mathbf{C}^{-1} \sum_s N_s (\mathbf{T}_s - \bar{\mathbf{T}})^T E(\bar{\mathbf{Y}}_s | N_1, \dots, N_S) \\ &= \boldsymbol{\tau}_{\text{HS}} + \sigma^2 \mathbf{C}^{-1} \sum_s N_s (\mathbf{T}_s - \bar{\mathbf{T}})^T \boldsymbol{\Omega} \mathbf{G}^T \boldsymbol{\Gamma}_s, \end{aligned} \quad (12)$$

where it should be remembered that \mathbf{C} is now $\sum N_s (\mathbf{T}_s - \bar{\mathbf{T}})^T (\mathbf{T}_s - \bar{\mathbf{T}})$. The expression for the j th element of the $p - 1$ dimensional vector $\boldsymbol{\Gamma}_s$ is derived in Web Appendix C, and can be written as follows, where for clarity we omit the subscript s from $\boldsymbol{\Theta}_s$ and note that $\boldsymbol{\Lambda} = \mathbf{I}_{p-1} + \sigma^2 \mathbf{G} \boldsymbol{\Omega} \mathbf{G}^T$ does not depend on s ,

$$\begin{aligned} (\boldsymbol{\Gamma}_s)_j &= \Phi_{p-2}(\mathbf{0}; \boldsymbol{\Theta}_{-j} - \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_{jj}^{-1} \Theta_j, \boldsymbol{\Lambda}_{-jj} \\ &\quad - \boldsymbol{\Lambda}_{jj}^{-1} \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^T) \phi(\mathbf{0}; \Theta_j, \boldsymbol{\Lambda}_{jj}) / \Phi_{p-1}(\mathbf{0}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}). \end{aligned} \quad (13)$$

Here, Θ_j is the j th element of $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_{-j}$ is $\boldsymbol{\Theta}$ omitting Θ_j . The j th diagonal element of $\boldsymbol{\Lambda}$ is $\boldsymbol{\Lambda}_{jj}$, $\boldsymbol{\Lambda}_{-jj}$ is $\boldsymbol{\Lambda}$ omitting the j th row and j th column and $\boldsymbol{\Lambda}_j$ is the j th column of $\boldsymbol{\Lambda}$, omitting $\boldsymbol{\Lambda}_{jj}$.

Equation (12) expresses a key result: it shows that the difference between the expectation of $\hat{\boldsymbol{\tau}}_{\text{PSS}}$ (conditional on the numbers of completely observed sequences) and $\boldsymbol{\tau}_{\text{HS}}$ is

a suitable combination of the elements of $\boldsymbol{\Gamma}_s$, which measures the effect of including only CCs in the estimation. Taking the expectation over the N_i allows the difference between $\boldsymbol{\tau}_{\text{PSS}}$ and $\boldsymbol{\tau}_{\text{HS}}$ to be evaluated.

3.2 | Unconditional expectation

The estimand $\boldsymbol{\tau}_{\text{PSS}}$ is the unconditional expectation of the CC treatment estimator, and this is found by taking the expectation of Equation (12) with respect to $N_s, s = 1, \dots, S$. These random variables are independent with $N_s \sim \text{binomial}(M_s, \pi_s)$, where π_s denotes the probability that an individual allocated to sequence s is a CC. This probability can be evaluated as

$$\begin{aligned} \pi_s &= \Pr(R_i(s) = p) = \int \Pr(R_i(s) = p | \mathbf{y}) \phi_p(\mathbf{y}; \boldsymbol{\mu}_s, \sigma^2 \boldsymbol{\Omega}) d\mathbf{y} \\ &= \Phi_{p-1}(\mathbf{0}; \boldsymbol{\Theta}_s, \mathbf{I}_{p-1} + \sigma^2 \mathbf{G} \boldsymbol{\Omega} \mathbf{G}^T). \end{aligned} \quad (14)$$

The expectation of Equation (12) can be approximated using a Taylor expansion about the mean of N_1, \dots, N_S , that is, about $M_1 \pi_1, \dots, M_S \pi_S$. A first-order expansion gives $\boldsymbol{\tau}_{\text{PSS}} = E(\hat{\boldsymbol{\tau}}_{\text{PSS}}) \approx E(\hat{\boldsymbol{\tau}}_{\text{PSS}} | M_1 \pi_1, \dots, M_S \pi_S)$, whereas a second-order expansion is

$$\boldsymbol{\tau}_{\text{PSS}} \approx E(\hat{\boldsymbol{\tau}}_{\text{PSS}} | M_1 \pi_1, \dots, M_S \pi_S) + \sum_{s=1}^S M_s \pi_s (1 - \pi_s) \mathbf{D}_{ss}, \quad (15)$$

where \mathbf{D}_{ss} is defined at the end of Web Appendix D.

4 | ILLUSTRATIONS FOR FOUR- AND SIX-PERIOD DESIGNS

The difference between the estimands, $\boldsymbol{\tau}_{\text{PSS}} - \boldsymbol{\tau}_{\text{HS}}$, derived in Section 3.2, will be illustrated using two crossover designs, which are shown in Table 1. The designs are based on Williams squares (Williams, 1949), which are forms of Latin squares, so have equal numbers of treatments and periods, and are balanced so that each treatment follows every other treatment equally often. One design has four periods and comprises two Williams squares: this will be referred to as the Tomado design, as it was the design used by the investigators in the TOMADO trial of treatments for sleep apnoea-hypopnoea (Quinnell et al., 2014). It is reasonable to be concerned that problems with patients dropping out may be more severe in longer trials, so this is investigated using a design comprising two six-period Williams squares, which can be found as an example of a perpetually connected design in Godolphin and Godolphin (2019). Williams squares have optimal or near optimal properties for a model including a carryover effect of

TABLE 1 The four- and six-period designs used in Section 4: columns are sequences and rows are periods, cell entries are treatments

Tomado design							
sequence							
1	2	3	4	5	6	7	8
1	2	3	4	1	2	3	4
3	4	2	1	4	3	1	2
4	3	1	2	3	4	2	1
2	1	4	3	2	1	4	3

Six-period design											
sequence											
1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	1	2	3	4	5	6
2	3	4	5	6	1	3	4	5	6	1	2
6	1	2	3	4	5	2	3	4	5	6	1
3	4	5	6	1	2	5	6	1	2	3	4
5	6	1	2	3	4	6	1	2	3	4	5
4	5	6	1	2	3	4	5	6	1	2	3

Notes: Designs each comprise two Williams Squares: the Tomado design, comparing four treatments, was used in Quinnell et al. (2014) and the six-period design, comparing six treatments, is Design 3E in Godolphin and Godolphin (2019).

treatment. Although many trials now eschew such a term in favor of washout periods, these designs have been chosen because many practitioners still use them as their balance seems inherently attractive.

4.1 | Estimand difference and level of missingness

The difference in estimands is illustrated numerically with an example contrast c for each design. These are shown for the Tomado and six-period designs in Figures 1 and 2, respectively, with the values assumed for the parameters in Equation (1) given in the captions. Whether an observation is missing is assumed to depend only on the observation itself and the immediately preceding observation, that is, $\theta_{kj} = 0, k > 2$ and, for simplicity, that $\theta_{0j} = \theta_0, \theta_{1j} = \theta_1, \theta_{2j} = \theta_2$. This model allows the effects of MNAR ($\theta_1 \neq 0$), MAR ($\theta_1 = 0, \theta_2 \neq 0$) and MCAR ($\theta_1 = \theta_2 = 0$) processes to be assessed. The percentage difference between the estimands for the contrast is $100c^T(\tau_{PSS} - \tau_{HS})/c^T\tau_{HS} = ED$, say, with the numerator found from Equation (15): this is plotted in the figures against θ_1 and θ_2 , each in the range $[-1,1]$ and $\theta_0 = -2.5$. The corresponding plots when $\theta_0 = -1.5$ are shown in Web Appendix E of the Supporting information. The percentage of CCs expected for each combination of θ_1, θ_2 is also plotted in these figures as $100 \sum M_s \pi_s / \sum M_s$: in these illustrations

TABLE 2 ED of contrast c and expected number of complete cases as a percentage of the number allocated to the design

	ρ	Mean	Min	Q1	Median	Q3	Max
Tomado: $c = \tau_4 - \tau_2$							
ED	0.3	-10	-38	-13	-9	-5	0
	0.8	-3	-17	-4	-2	-1	0
Complete cases	0.3	78	51	69	77	86	98
	0.8	83	59	76	84	91	98
Six-period Williams design: $c = \tau_5 - \tau_2$							
ED	0.3	-11	-35	-14	-9	-5	0
	0.8	-3	-15	-4	-3	-1	0
Complete cases	0.3	68	33	57	66	79	97
	0.8	78	54	70	77	86	97

Note: Tabulated values are the mean and five-number summary over the range $-1 < \theta_1, \theta_2 < 1$, with $\theta_{0j} = -2.5$ and are given for $\rho = 0.3, 0.8$. Model parameters as described in Figures 1 and 2.

$M_s = 2$ for all s . These values for the θ s have been chosen to ensure that the missingness probabilities (4), given the values assumed for the parameters in (1), are realistic and lead to plausible percentages of CCs, cf. Figures 1 and 2.

Summaries of the ED and CCs are shown in Table 2 for $\rho = 0.3, 0.8$ over the region $-1 < \theta_1, \theta_2 < 1$. In Figure 1 the ED for the contrast $\tau_4 - \tau_2$ from Tomado with a within-unit correlation $\rho = 0.3$ is largely less than 12%, and is close to zero in the vicinity of $\theta_1 = \theta_2 = 0$, corresponding to MCAR. The ED values are smaller when the within-unit correlation is higher, being generally less than 5% when $\rho = 0.8$, cf. Table 2. The upper plots in Figures 1 and 2 show a broadly elliptical pattern, with bias being least along the line $\theta_2 = \theta_1$ and largest along $\theta_2 = -\theta_1$, which is consistent with the result for the AB/BA design (Matthews et al., 2014). The largest differences between the estimands seen in Table 2 arise in small parts of the (θ_1, θ_2) region, close to the extremes where $\theta_1 = -\theta_2$: this is where the missingness probabilities depend on the change in response between successive periods, rather than their general level.

The corresponding results for the six-period Williams design are very similar for ED with, as might be anticipated, slightly lower values for percentage of CCs for the longer design (cf. Table 2).

From Table 2 and from the contour plots, it is seen that the ED is always negative, indicating that the contrast $c^T\tau_{PSS}$ is closer to zero than $c^T\tau_{HS}$. This is also consistent with the result in Matthews et al. (2014).

The missingness process corresponds to MAR when $\theta_1 = 0$. In this case, the standard REML analysis of all available data would provide an unbiased estimate of τ_{HS} . The ED and percentage CCs for the CC analysis when data

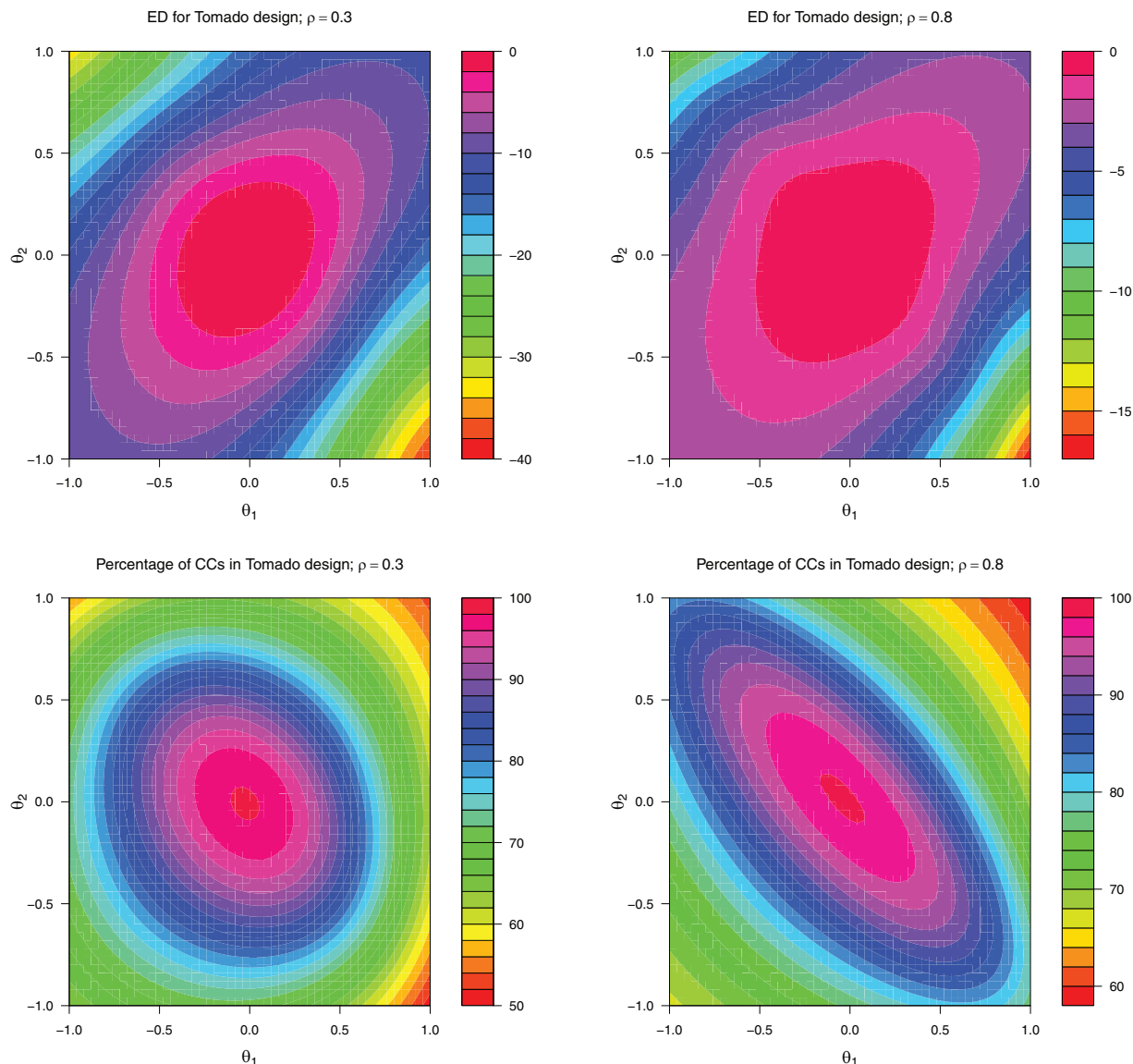


FIGURE 1 Contour plots of ED for $\tau_4 - \tau_2 = 0.4$ and of the percentage of complete cases, from the Tomado design for $\sigma^2 = 4$ and $\rho = 0.3, 0.8$. Model parameters are: $\pi_1, \dots, \pi_4 = 0, 0.2, -0.2, 0.4$, $\tau_2, \tau_3, \tau_4 = 0, 0.2, 0.4$. This figure appears in color in the electronic version of this paper.

are MAR are shown in the top row of Figure 3 for both designs and for $\rho = 0.3, 0.8$. The ED is less than 5% for all θ_2 when $\rho = 0.8$, and is less than 15% when $\rho = 0.3$. In the latter case, the ED is less than 5%, provided the percentage of CCs is above about 85%. The lower row in Figure 3 shows a very similar picture when θ_1 varies, while $\theta_2 = 0$, an instance of data MNAR. In all cases, the ED is zero when data are MCAR, that is, $\theta_1 = \theta_2 = 0$ but there is some loss of cases because of the role of $\theta_0 = -2.5$. From Equation (6), the probability of a CC in this instance is $\Phi(-\theta_0)^{(p-1)}$, which is 98% for the four-period designs and 97% for the six-period design, as confirmed in Table 2. These values change to 81% and 71% if $\theta_0 = -1.5$, as can be seen in Web Appendix E.

The results are presented in terms of the percentage difference in estimands, calculated as $100\mathbf{c}^T(\boldsymbol{\tau}_{\text{PSS}} - \boldsymbol{\tau}_{\text{HS}})/\mathbf{c}^T\boldsymbol{\tau}_{\text{HS}}$, but based on calculations using an assumed value for the contrast $\mathbf{c}^T\boldsymbol{\tau}_{\text{HS}}$. Results for other values of the contrast (not shown) illustrate that the percentage difference is largely unaffected by the size of the contrast, so percentage difference is a suitable summary. This accords with the approximation derived for the AB/BA design in Matthews et al. (2014), where the estimand difference is shown to be approximately proportional to the HS estimand. If the contrast $\mathbf{c}^T\boldsymbol{\tau}_{\text{HS}}$ is zero, then the CC estimator of the contrasts appears to be zero across the (θ_1, θ_2) plane—see Web Appendix E—suggesting that in this case $\mathbf{c}^T\boldsymbol{\tau}_{\text{PSS}}$ also vanishes.

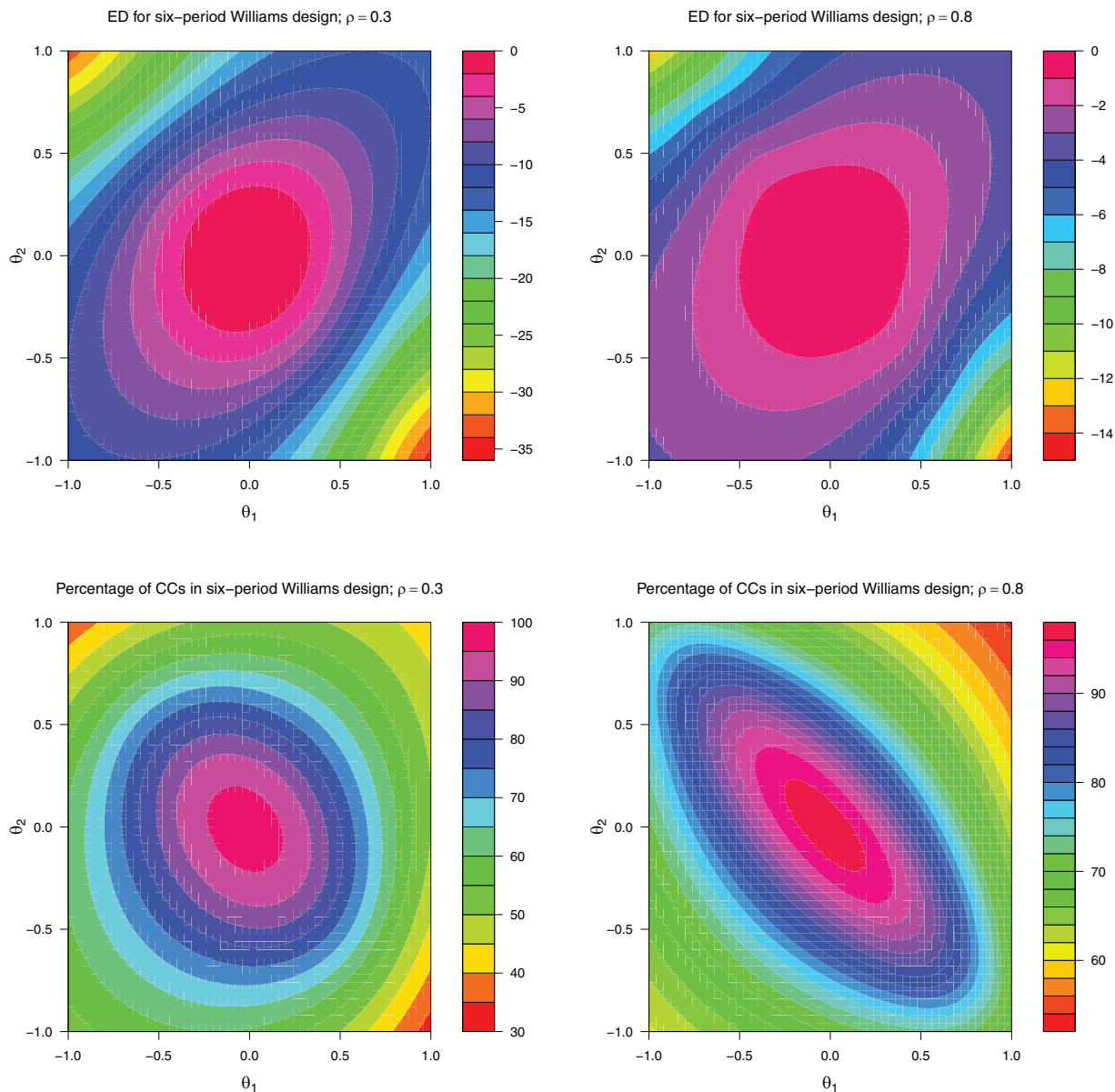


FIGURE 2 Contour plots of ED for $\tau_5 - \tau_2 = 0.4$ and of the percentage of complete cases, from the Six-period Williams design for $\sigma^2 = 4$ and $\rho = 0.3, 0.8$. Model parameters are: $\pi_1, \dots, \pi_6 = 0, 0.1, 0, 0.2, -0.2, 0.4$, $\tau_2, \dots, \tau_6 = 0, 0.2, 0.4, 0.4, -0.2$. This figure appears in color in the electronic version of this article.

4.2 | Illustrations from the TOMADO trial

The TOMADO trial (Quinnell et al., 2014) compared three mandibular advancement devices with no intervention for the treatment of mild to moderate obstructive sleep apnea-hypopnea, using the four-period design in Table 1: the outcome considered here is the Epworth Sleepiness Scale (ESS). Treatments 1–4 are, respectively, no intervention, self-fitted (SP1), semi-bespoke (SP2), and bespoke (bMAD).

Ninety patients were randomized in the trial, but seven did not complete even the first period and provided no data: these patients play no part in our analysis. Seventy-

four patients provided complete data and some data were available on 83 patients, providing 314 observations. Of the 83 patients, 10 were allocated to sequences 2, 5, and 6, with 9 allocated to sequence 8 and 11 to each of the other sequences. Analysis of all available data using a linear mixed effects model and REML gave the estimates of $\pi_1, \dots, \pi_4, \tau_2, \dots, \tau_4$ as 10.65, 9.88, 9.69, 10.10, $-1.51, -2.15, -2.37$ with estimates for ρ, σ^2 of 0.61 and 16.63, respectively. The corresponding estimates from an analysis of the 74 CCs were very similar: the period and treatment parameters were 10.58, 9.93, 9.67, 10.09, $-1.47, -2.30, -2.41$, with ρ and σ^2 being 0.62 and 16.43, respectively. In both analyses, the standard errors of these

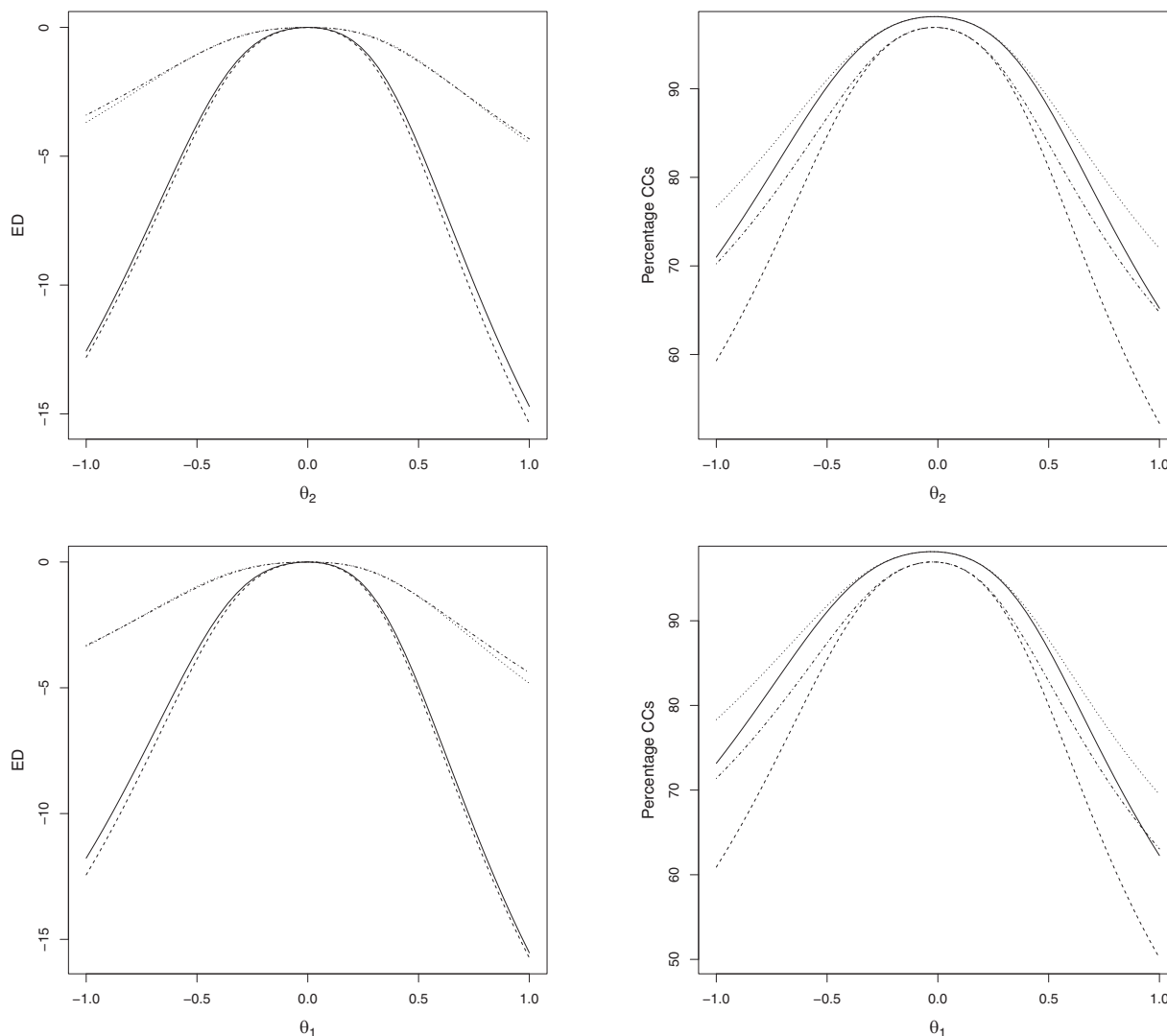


FIGURE 3 ED and expected percentage of complete cases for the Tomado (solid line) and six-period Williams (dashed line) designs for $\rho = 0.3$ and for Tomado (dotted line) and six-period Williams (dot-dashed line) designs for $\rho = 0.8$: for all cases $M_s = 2$ for all s . Selected contrast is $\tau_4 - \tau_2$ for the four-period design and $\tau_5 - \tau_2$ for the six-period design. Top row of plots shows ED and percentage expected number of CCs as θ_2 varies with $\theta_1 = 0$, that is, corresponding to missing at random (MAR). Bottom row of plots shows corresponding plots as θ_1 varies with $\theta_2 = 0$, so cases here are MNAR. In all cases, $\theta_0 = -2.5$. All plots show the MCAR case when the abscissa is zero.

estimates of period parameters are approximately 0.52, and approximately 0.40 for the treatment parameters.

Figure 4A,B corresponds to the contour plots in Figure 1 but assume the parameter values from the analysis of all available data (the plots are indistinguishable from those based on the CC analysis). The scale of the outcome in TOMADO, ESS, differs from that used in Figure 1, with larger means and variances. As such the region $-1 < \theta_1, \theta_2 < 1$ would include very extreme missingness probabilities, so a range $-0.2 < \theta_1, \theta_2 < 0.2$ is more appropriate.

Values for $\theta_0, \theta_1, \theta_2$ cannot be estimated from the data. However, an illustration of the effect on the difference in the PSS and HS estimands of different values of these parameters can be provided by locating triples

$(\theta_0, \theta_1, \theta_2)$, where the expected proportion of CCs matches the observed proportion, 74/83. Starting from the MCAR case, where $\theta_1 = \theta_2 = 0$, $\theta_0 = -\Phi^{-1}(\sqrt[3]{74/83}) = -1.78$, the solid line in Figure 4C is the locus in the (θ_1, θ_2) -plane, where the expected proportion of CCs is within 0.01 of 74/83. The other lines give the corresponding locus when θ_0 takes the values given in the caption to Figure 4. The curves in Figure 4D show the ED for the SP1-bMAD contrast plotted against θ_1 , as (θ_1, θ_2) track along the corresponding locus in Figure 4C. The ED is zero for the MCAR case, and departs from zero by up to -4.5% at the extremes of the loci shown in Figure 4C. This variation in ED occurs because of the changes in the values of θ_1 and θ_2 , and not because of any change in the level of missingness, which is fixed at 74/83.

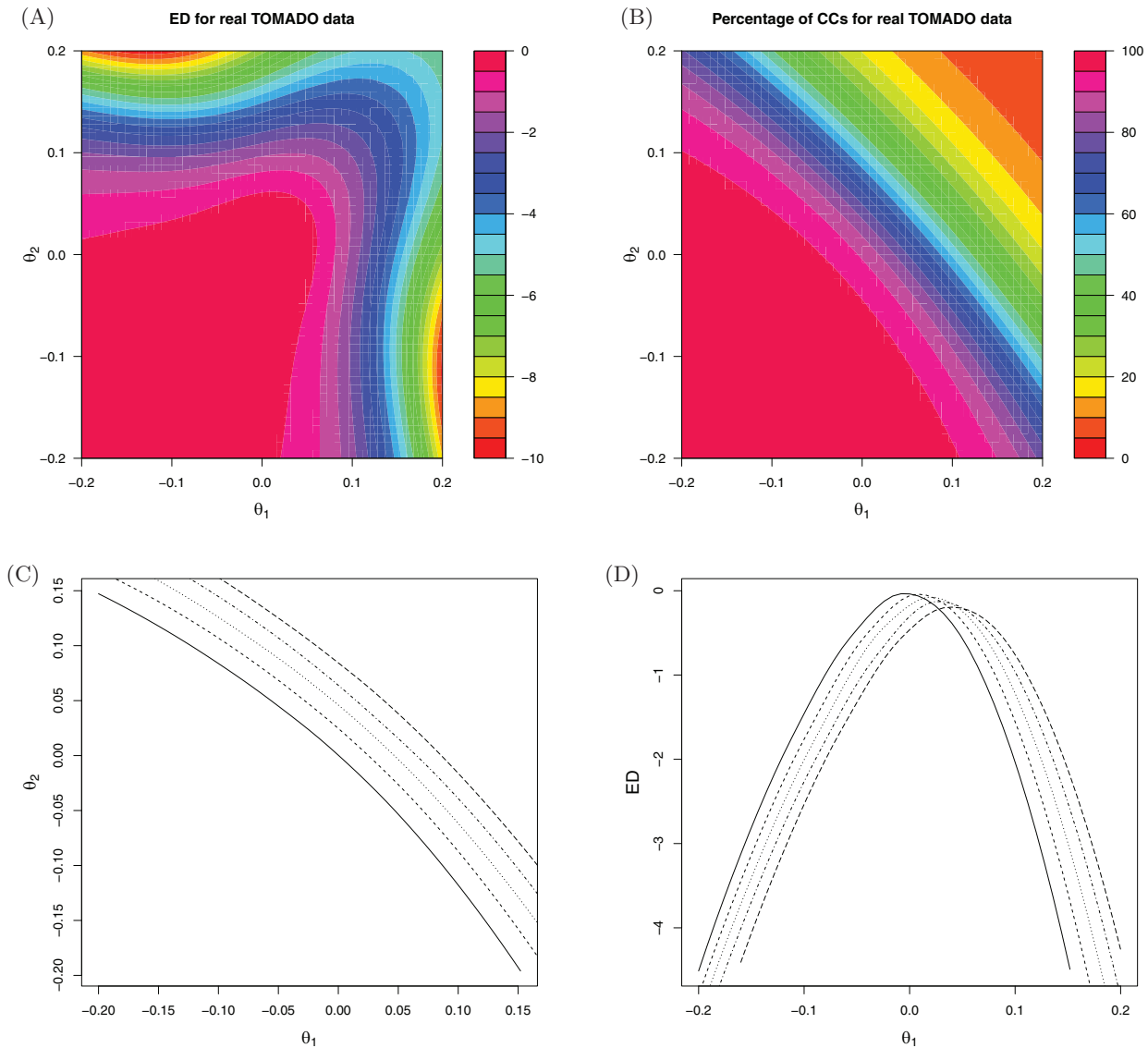


FIGURE 4 Contour plots show ED in contrast SPI-bMAD (A) and percentage expected CCs (B), based on parameter values from the TOMADO trial: see the text for details. Plot (C) shows the locus of points (θ_1, θ_2) , where the expected proportion of CC matches the observed value (74/83), for each of $\theta_0 = -1.78$ (solid line), -2.0 (dashed), -2.2 (dotted), -2.4 (dash-dotted), and -2.6 (long dashed). Plot (D) shows the ED in the contrast SPI-bMAD along the (θ_1, θ_2) loci from plot (C), plotted against θ_1 , for the θ_0 values and line types in (c). This figure appears in color in the electronic version of this paper.

5 | SOME COMMENTS ON DISCONNECTED DESIGNS

It was pointed out in Section 1 that missingness could mean that \tilde{D} is disconnected. However, disconnection with regard to CC analyses is rather different to that in the research cited in Section 1, which generally assumes that data are MCAR and all available data are analyzed using a model with a carryover treatment effect. In this paper, where \tilde{D} comprises only the CCs, and using model (1), \tilde{D} cannot be disconnected unless there are some sequences with no CCs. To see this write Equation (3) as $C = \sum N_s A_s$,

with all $N_s > 0$ and note that disconnection would imply the existence of a non-zero vector x such that $x^T C x = 0$. It would follow that $x^T A_s x = 0$ for all s , as the A_s are non-negative definite, so $x^T (\sum M_s A_s) x = 0$, that is, D is disconnected, contrary to assumption.

As a consequence, an upper bound on the probability that \tilde{D} is disconnected is $1 - \prod_{s=1}^S [1 - (1 - \pi_s)^{M_s}]$, with π_s given in Equation (14). When the M_s are reasonably large, as in the TOMADO trial, this bound is likely to demonstrate that there is little chance that \tilde{D} is disconnected. Using parameter values from the TOMADO trial, and assuming the missingness process has $\theta_{kj} = 0$ for

$k > 2$, and $\theta_0 = -1.78$, then the medians of the π_s over the region $-0.2 < \theta_1, \theta_2 < 0.2$ are around 0.85 and occur at similar points in the plane. For example, at $(\theta_1, \theta_2) = (0.144, 0.168)$, the π_s are 0.89, 0.85, 0.88, 0.87, 0.88, 0.85, 0.88, 0.88: with the M_s specified in Section 4.2 the above upper bound on the probability of disconnection is less than 10^{-7} . However, in smaller trials the upper bound given above will be much less useful: some comments on a more careful analysis are in Web Appendix F.

6 | DISCUSSION

In this paper, we have argued that a CC analysis of a complete crossover design provides a way to estimate a PSS estimand: although couched in different terms, this is essentially the point made in Matthews et al. (2014) for the AB/BA design. This estimand will give some indication of the treatment efficacies among those able to tolerate all the treatments, so is likely to be relevant to those involved with administering the treatments in practice. A complete crossover design is able to do this in a way that a parallel group design cannot because recruits are, at some stage of the trial, offered all the treatments under investigation. It is, perhaps, surprising, that this potential advantage of crossover trials was not mentioned in the ICH E9(R1) guideline (International Council for Harmonisation, 2019).

The ICH guideline discusses other estimands and these may be relevant to other users. The treatment policy strategy estimand aligns most closely with an ITT approach and it is notable that writing 35 years ago Lewis (1987) cast doubt on the relevance of the ITT principle for crossover designs. Knowledge of the HS estimand may be useful for those involved with development of the treatments, as it is the quantity that would be estimated from Equation (1) if all planned observations were made. The populations on which HS and PSS estimands are based differ by the patients who drop out of the study before they complete their allocated treatments. With our definition, an estimate of τ_{PSS} can always be obtained from a CC analysis. If all available data are analyzed using REML this is true for τ_{HS} , provided that any missing data are MAR.

While the method developed in Sections 2 and 3 can be applied to any complete crossover design, the numerical illustrations are largely based on a Markovian missingness process and the widely used Williams' crossover designs. Perhaps the principal observation from Figures 1 and 2 is that the relative difference between the PSS and HS estimands, ED, decreases as the correlation between observations in successive periods increases. Data are MCAR at the origin in the (θ_1, θ_2) region, where the ED must vanish, and it is important to note that the ED remains small in an extended region around the origin for both designs used

for the illustration. The ED is less than 3% across most of the plotted region when $\rho = 0.8$ and only exceeds 5% in relatively small regions close to $(\theta_1, \theta_2) = (-1, 1)$ and $(1, -1)$. The ED is rather larger when $\rho = 0.3$ but this is probably very much at the lower end of the range of ρ likely to be encountered in crossover trials. The justification for using a crossover trial would not be strong for such low ρ . The value for the real TOMADO data is larger than 0.6, and the 10 values presented in Elbourne et al. (2002) for the correlations between outcomes are between 0.49 and 0.91 with a median of 0.75.

CCs are analyzed because these patients have been able to tolerate all the treatments. In practice, some patients who would have tolerated all treatments may fail to complete the trial for reasons not related to the treatment. If the investigator can be assured that the reasons for a patient dropping out of the trial are unrelated to their outcomes, that is, they are MCAR, then it may be that an analysis estimating τ_{PSS} could incorporate partial information from these patients. The results in Figures 1 and 2 show that τ_{PSS} is closer to zero than τ_{HS} and that the two coincide if all missing data are MCAR, so it is reasonable to conjecture that this adjustment to the population on which τ_{PSS} is based would reduce the discrepancy between τ_{PSS} and τ_{HS} .

The missingness model that we have used, Equation (4), is of a form widely used in the missing data literature. The use of the probit, rather than the more usual logistic, link is a minor adjustment made for mathematical convenience. The model assumes that the chance of dropout is related to the values of the outcome variables. This would be realistic in many settings, such as pain relief trials, where patients may not tolerate ineffective treatments. However, there may be other settings where the probability of dropout might need to be modeled otherwise, perhaps by including terms for allocated treatment. This possibility is something which might be addressed in further work.

If data are MAR, that is, along the $\theta_1 = 0$ axis in Figures 1, 2, and 4, then τ_{HS} can be estimated using REML on all available data. Away from this axis, such estimates will be biased. The estimate of τ_{PSS} obtained from a CC analysis, together with plots such as Figure 4C,D, evaluated at the estimated model parameters, may provide the analyst with some information on the likely range of values for τ_{HS} . Our results suggest that this difference is unlikely to be large unless so many observations are missing that this would, in itself, undermine the credibility of the study.

ACKNOWLEDGMENTS

This work was supported by the NIHR Research Methods Opportunity Funding Scheme (Grant number: RMOFS 2012/05]. The authors are grateful to Dr Tim Quinnell for permission to use the data from the TOMADO trial

(NIHR HTA Programme, Project number 08/110/03), to Dr Michael Grayling for his helpful comments and especially to the reviewers and Associate Editor, whose comments led to a radical change in the presentation.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this paper.

ORCID

John N.S. Matthews  <https://orcid.org/0000-0003-2559-037X>

REFERENCES

- Basu, S. & Santra, S. (2010) A joint model for incomplete data in crossover trials. *Journal of Statistical Planning and Inference*, 140, 2839–2845.
- Bate, S.T., Godolphin, E.J. & Godolphin, J.D. (2008) Choosing crossover designs when few subjects are available. *Computational Statistics & Data Analysis*, 52, 1572–1586.
- Elbourne, D., Altman, D., Higgins, J., Curtin, F., Worthington, H., & Vail, A. (2002) Meta-analyses involving cross-over trials: methodological issues. *International Journal of Epidemiology*, 31, 140–149.
- Godolphin, J.D. (2006) The specification of rank reducing observation sets in experimental design. *Computational Statistics & Data Analysis*, 51, 1862–1874.
- Godolphin, P.J. & Godolphin, E.J. (2019) Robustness of crossover trials against subject drop-out—examples of perpetually connected designs. *Statistical Methods in Medical Research*, 28, 788–800.
- González-Farías, G., Domínguez-Molina, J.A. & Gupta, A.K. (2004) The closed skew-normal distribution. In: Genton, M.G. (Ed.) *Skew-elliptical distributions and their applications: a journey beyond normality*, Boca Raton, FL: Chapman and Hall, pp. 25–42.
- Ho, W.-K., Matthews, J.N.S., Henderson, R., Farewell, D. & Rodgers, L. R. (2012) Dropouts in the AB/BA crossover design. *Statistics in Medicine*, 31, 1675–1687.
- International Council for Harmonisation. (2019) Addendum on estimands and sensitivity analysis in clinical trials to the guidelines on statistical principles in clinical trials E9(R1). Available at: <http://www.ich.org/page/efficacy-guidelines>. Accessed April 25 2022.
- Jones, B. & Kenward, M.G. (2014) *Design and analysis of cross-over trials*. 3rd ed. Boca Raton, FL: Chapman and Hall.
- Kackar, R. & Harville, D. (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear-models. *Journal of the American Statistical Association*, 79, 853–862.
- Leuchs, A.-K., Zinserling, J., Brandt, A., Wirtz, D. & Benda, N. (2015) Choosing appropriate estimands in clinical trials. *Therapeutic Innovation & Regulatory Science*, 49, 584–592.
- Lewis, J. A. (1987) Migraine trials—crossover or parallel group. *Neuroepidemiology*, 6, 198–208.
- Liu, F. (2011) Estimation bias in complete-case analysis in crossover studies with missing data. *Communications in Statistics: Theory and Methods*, 40, 812–827.
- Low, J.L., Lewis, S.M. & Prescott, P. (1999) Assessing robustness of crossover designs to subjects dropping out. *Statistics and Computing*, 9, 219–227.
- Majumdar, D., Dean, A.M. & Lewis, S.M. (2008) Uniformly balanced repeated measurements designs in the presence of subject dropout. *Statistica Sinica*, 18, 235–253.
- Mallinckrodt, C., Molenberghs, G. & Rathmann, S. (2017) Choosing estimands in clinical trials with missing data. *Pharmaceutical Statistics*, 16, 29–36.
- Matthews, J.N.S., Henderson, R., Farewell, D., Ho, W.-K. & Rodgers, L.R. (2014) Dropout in crossover and longitudinal studies: is complete case so bad? *Statistical Methods in Medical Research*, 23, 60–73.
- Molenberghs, G. & Kenward, M.G. (2007) *Missing Data in Clinical Studies*. Wiley, Chichester.
- National Research Council. (2010) *The prevention and treatment of missing data in clinical trials*. Washington, DC: The National Academies Press.
- Prescott, P. & Mansson, R. (2001) Robustness of balanced incomplete block designs to randomly missing observations. *Journal of Statistical Planning and Inference*, 92, 283–296.
- Quinnell, T.G., Bennett, M., Jordan, J., Clutterbuck-James, A.L., Davies, M.G., Smith, I.E. et al. (2014) A crossover randomised controlled trial of oral mandibular advancement devices for obstructive sleep apnoea-hypopnoea (TOMADO). *Thorax*, 69, 938–945.
- Rubin, D.B. (1976) Inference and missing data. *Biometrika*, 63, 581–592.
- Senn, S.J. (2002) *Cross-over trials in clinical research*. 2nd ed. Chichester: Wiley.
- Varghese, C., Rao, A.R. & Sharma, V.K. (2002) Robustness of williams square change-over designs. *Metrika*, 55, 199–208.
- Wang, X. & Chinchilli, V.M. (2021) Analysis of crossover designs with nonignorable dropout. *Statistics in Medicine*, 40, 64–84.
- Williams, E.J. (1949) Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research A: Physical Sciences*, 2, 149–168.

SUPPORTING INFORMATION

Web Appendices and Figures referenced in Sections 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library. R code to compute key quantities, such as the formulae in Equations (14) and (15), is also available on this website.

How to cite this article: Matthews, J.N.S., Bazakou, S., Henderson, R. & Sharples, L.D. (2023) Contrasting principal stratum and hypothetical strategy estimands in multi-period crossover trials with incomplete data. *Biometrics*, 79, 1896–1907. <https://doi.org/10.1111/biom.13777>