Highly Efficient Stepped Wedge Designs for Clusters of Unequal Size

J.N.S. Matthews

School of Mathematics, Statistics & Physics and Institute of Health & Society,
Newcastle University,
Newcastle upon Tyne, NE1 7RU, U.K.

Abstract
The Stepped Wedge Designs (SWD) is a form of cluster randomized trial, usually comparing two treatments, which are divided into sequences and time periods. Clusters are allocated to sequences, with the treatment changing at different periods in the different sequences. Typically all sequences start with the standard treatment and end with the new treatment, which can make SWDs attractive to practitioners. The clusters allocated to the sequences will usually differ in size but the existing literature generally assumes that they have the same size. This paper considers the case when clusters have different sizes and determines optimal designs in some special cases and highly efficient designs in the general case, with bounds placed on the amount by which they fall short of optimal. The designs allocate the same proportion of subjects to each of the sequences of the SWD except for the extreme sequences, where treatments change after the first period or just before the final period, which receive a different proportion of subjects. The proportions depend on the cluster sizes, the duration of the study and the intra-class correlation. The paper concentrates on the cross-sectional design, where subjects are measured once but the results are extended to the closed-cohort design, in which each subject is measured in each period of the study.

Key words Closed-cohort design; Cluster randomized trial; Cross-sectional design; Optimal design; Stepped wedge design.

*john.matthews@ncl.ac.uk: corresponding author
Table 1: A classic stepped wedge design with $S$ sequences, comparing treatment A and B over $T$ periods, $S = T - 1$.

| Sequence | 1 | 2 | 3 | ... | $T - 1$ | $T$ |
|----------|---|---|---|-----|--------|-----|
| 1        | A | A | A | ... | A     | B   |
| 2        | A | A | A | ... | B     | B   |
| ...      |   |   |   |     |       |     |
| $S - 1$  | A | A | B | ... | B     | B   |
| $S$      | A | B | B | ... | B     | B   |

1 Introduction

A cluster randomized trial (see, e.g., Donner and Klar 2000; Hayes and Moulton 2017) is a form of randomized trial which has been used for many years in circumstances where individual randomization is either impractical or undesirable. In most applications clusters are allocated to a treatment and remain on that treatment for the duration of the trial. However, there has been increasing interest in designs where a cluster can experience more than one of the trial treatments, and one such variant which has received substantial attention is the Stepped Wedge Design (SWD): for an early review of its use see Mdege et al. (2011). The classic SWD compares two treatments, A and B say, and clusters are randomized to one of $S$ sequences, denoted by 1, 2, ..., $S$, as shown in Table 1. The trial is divided into $T$ treatment periods and treatment A is used in the first period, changing over to B in a period determined by the sequence, with clusters finishing on treatment B: no reversion from B to A is allowed.

For some areas of application the SWD has several advantages. Typically A is the standard treatment and B is a new treatment, so the staged changeover from A to B inherent in a SWD can be a useful way to assess the practical utility following the implementation into practice of a treatment previously found to be valuable in more highly controlled conditions. The proscription on changes from the new (B) treatment to the standard (A) can also be attractive to potential participants.

The SWD also has the advantage over standard cluster randomized trials that treatment estimation can be made within-cluster as each cluster will have experience of both treatments, albeit in a rather imbalanced way for those allocated to extreme sequences, such as sequence 1 or sequence $S$. Treatment A is largely applied at the start of the study, and B towards
the end, so that analysis of a SWD must allow for any trend in the outcome. The substantial imbalance between periods and treatments in the SWD leads to a noticeable loss of efficiency relative to more conventional designs (see Matthews and Forbes, 2017) and their use should be confined to cases where their practical advantages are decisive.

The statistical model most widely adopted for the analysis of SWDs is that proposed by Hussey and Hughes (2007), who use a mixed-effects model to accommodate the clustering and a period effect to allow for any trend. It assumes that the outcome on individual $k$ in cluster $i$ in period $j$, $Y_{ijk}$, follows the model

$$Y_{ijk} = \beta_j + X_{ij}\theta + \alpha_i + \epsilon_{ijk},$$  \hspace{1cm} (1)

where $\beta_j$ is the effect of period $j$, $j = 1, \ldots, T$ and $X_{ij} \in \{0, 1\}$ is a treatment indicator which is 1 only if treatment B is administered in cluster $i$ during period $j$. It is supposed that there are $C$ clusters and the effect of cluster $i$ is modelled by the $\alpha_i$, $(i = 1, \ldots, C)$, which are independent random variables with mean 0 and variance $\tau^2$. It is assumed that the residual terms $\epsilon_{ijk}$ are independent of each other, and of the $\alpha_i$, and have mean 0 and variance $\sigma^2_e$.

While a cluster always appears in all periods of the design, an important distinction should be made between cross-sectional and closed-cohort designs. In the latter the individuals in a cluster are observed repeatedly throughout the design, so $k$ in model (1) refers to the same individual, measured $T$ times, as would arise if, e.g., the residents in a care home were assessed throughout the study. In the former case each individual is observed just once, as would occur if the clusters were intensive care units and individuals are patients observed in just one period. The modification of (1) needed for the cohort design is presented in Section 4.

Since the Hussey & Hughes model was described many variations have been proposed, both to the method of analysis and to the basic design. These include allowing random period effects (Hooper et al., 2016), allowing more complex correlation structures (Hooper et al., 2016; Girling and Hemming, 2016; Kasza et al., 2018), adding sequences to the design where treatment allocation is unchanged so as to create hybrids between standard cluster-randomized and SWDs (Girling and Hemming, 2016), changing the proportions allocated to the first and last periods (Thompson et al., 2017), and allowing more than two treatments (Lyons et al., 2017; Grayling et al., 2018). Most contributions have assumed allocation of equal numbers of clusters of equal size to the sequences of the SWD. Varying the number of clusters per
sequence was considered in Girling and Hemming (2016) for hybrid designs and the optimal classic SWD design was determined by Lawrie et al. (2015) for the cross-sectional case and for a closed-cohort design by Li et al. (2018). These papers, along with almost all others have assumed that the number of individuals in a cluster (cohort design), or the number in a cluster-period cell (cross-sectional design) is the same for all clusters.

In practice, the assumption that all clusters have the same size is unlikely to hold and it is possible that the efficiency of the analysis might be improved by taking account of the clusters sizes when allocating them to sequences. Given the inherent inefficiency of the SWD, it is important to investigate any avenues which might improve the efficiency of the design. This article addresses the problem of finding the optimal allocation to sequences, with the focus initially being on the cross-sectional SWD, with outcomes which obey \([\Pi]\). The number of individuals seen in a cluster-period cell for cluster \(i\) is denoted by \(N_i\): some period to period variation in the number recruited in cluster \(i\) is inevitable but this is ignored in our formulation in the interests of parsimony. Our approach assumes that the number of periods, \(T\) (and hence \(S = T - 1\)) has been chosen and that a value of the intra-class correlation (ICC) \(\rho = \tau^2 / \left(\tau^2 + \sigma^2_e\right)\) is assumed. For most cluster trials it is reasonable to assume that very good estimates of the \(N_1, \ldots, N_C\) are known at the outset. Uncertainties in these values can be addressed by appropriate sensitivity analyses. Our approach emphasises analytical over numerical methods, even if this requires some approximation, as the general form of an optimal, or near-optimal, design can provide greater insight than an isolated numerical solution. The aim is to find an allocation which provides a highly efficient value for \(\text{var}(\hat{\theta})\): Section 2 calculates a suitable form for \(\text{var}(\hat{\theta})\); Section 3 finds optimal designs for some special cases, introduces symmetric designs and finds highly efficient designs for the general case; the methods are extended to the closed-cohort design in Section 4 and some concluding remarks are in Section 5.

## 2 Computing the variance of \(\hat{\theta}\)

### 2.1 Variance of \(\hat{\theta}\) in terms of the cluster sizes

The model in \([\Pi]\) induces a model for the cluster by period means \(Y_{ij+} = N_i^{-1} \sum_{k=1}^{N_i} Y_{ijk}\), namely \(Y = Z\phi + e\), where \(Y^T = (Y_1^T, \ldots, Y_C^T)\) and \(Y_i\) is the \(T\)-dimensional vector \((Y_{i1+}, \ldots, Y_{iT+})^T\), which has variance \(V_i\). The \((CT)\times(T+1)\) design matrix for the induced model is \(Z, \phi = (\beta_1, \ldots, \beta_T, \theta)^T\).
and \( \text{var}(e) \) is the \( CT \times CT \) block diagonal matrix \( V = \text{diag}(V_1, V_2, \ldots, V_C) \).

For the model in (1) \( V_i = \tau^2 J_T + N_i^{-1} \sigma^2 e I_T \), where \( I_T \) is the \( T \times T \) identity matrix and \( J_T = 1^T 1_T \), with \( 1_T \) being the \( T \times 1 \) vector of ones. Consequently \( \text{var}(\hat{\theta}) \) is the bottom right hand element of \( (Z^TV^{-1}Z)^{-1} \). Now

\[
Z = \begin{pmatrix} I_T & D_1 \\ I_T & D_2 \\ \vdots & \vdots \\ I_T & D_C \end{pmatrix},
\]

where \( D_i \) is the \( T \times 1 \) vector of 0s and 1s indicating the treatment allocation for cluster \( i \). It follows that

\[
(Z^TV^{-1}Z)^{-1} = \left( \sum V_i^{-1} D_i \right)^{-1},
\]

and therefore

\[
\text{var}(\hat{\theta}) = \sum D_i^T V_i^{-1} D_i - \left( \sum D_i^T V_i^{-1} \right) \left( \sum V_i^{-1} \right)^{-1} \left( \sum V_i^{-1} \right).
\]

Further progress requires each term of the above to be evaluated and details are given in Appendix A: the result is that

\[
\sigma_e^2 \text{var}(\hat{\theta})^{-1} = E - H \frac{G}{N} - \frac{TF^2 + WE^2 - 2EF}{N(1 - WT)},
\]

where

\[
E = \sum_{i=1}^C N_i r_i \quad F = \sum_{i=1}^C N_i w_i r_i
\]

\[
G = \sum_{i=1}^C \sum_{i'=1}^C N_i N_{i'} r_{ii'} \quad H = \sum_{i=1}^C N_i w_i r_i^2
\]

and also

\[
N = \sum_{i=1}^C N_i \quad W = \sum_{i=1}^C N_i w_i / N.
\]

In the above equations \( r_i \) is the number of times treatment B is allocated to the \( i \)th cluster, \( r_{ii'} \) denotes the number of periods in which cluster \( i \) and cluster \( i' \) both receive treatment B, and

\[
w_i = N_i \tau^2 / (\sigma_e^2 + N_i \tau^2 T) = N_i / (\lambda + N_i T),
\]

where \( \lambda = \sigma_e^2 / \tau^2 = (1 - \rho) / \rho \).
2.2 Variance of $\hat{\theta}$ in terms of sequence allocations

The clusters are to be allocated to the $S$ sequences in Table 1. If cluster $i$ is allocated to sequence $\ell$ then for all such clusters $r_i = r_\ell$, where $r_\ell$ denotes the number of occurrences of B in sequence $\ell$, which means that the expressions in (3), (4) can be re-written in terms of sums over sequences instead of sums over clusters. Note that with the sequences numbered as in Table 1 $r_\ell = \ell$ and $r_\ell\ell' = \min\{\ell, \ell'\}$, $\ell, \ell' = 1, \ldots, S$.

An optimal design can be sought by ascertaining the allocation of the $N_i$ to sequences which maximizes (2). For an analytic, as opposed to a numerical solution, there is merit in providing a succinct and general description of an optimal design. With this in mind we define $p_i = N_i / N$, the proportion of patients in cluster $i$. We also define $p_\ell = \sum_{i \in \ell} p_i$, where $\ell = 1, \ldots, S$ indexes the sequences, and $i \in \ell$ is taken to be the set of clusters allocated to sequence $\ell$, so $p_\ell$ is the proportion of the available individuals allocated to sequence $\ell$. Note that $\sum_\ell p_\ell = \sum_i p_i = 1$. In addition, define $q_i$ by

$$q_i = N_i w_i / N = N_i^2 / \{N(\lambda + N_i T)\},$$

and analogously to $p_\ell$, $q_\ell = \sum_{i \in \ell} q_i$, so $\sum_\ell q_\ell = \sum_i q_i = W$. Note that $W$ is determined once the $N_i$ and $\lambda$ are known - it does not depend on the allocation of clusters to sequences. These definitions allow us to rewrite the quantities in (3) and (4) as

$$E = N \sum_\ell \ell p_\ell$$

$$F = N \sum_\ell \ell q_\ell$$

(6)

$$G = N^2 \sum_\ell \sum_{\ell'} \min\{\ell, \ell'\} p_\ell p_{\ell'}$$

$$H = N \sum_\ell \ell^2 q_\ell.$$  

(7)

2.3 Matrix expression for the variance of $\hat{\theta}$

The calculations in Sections 2.1 and 2.2 indicate that

$$N^{-1} \sigma^2_e \text{var(}\hat{\theta})^{-1} = \sum_\ell \ell p_\ell - \sum_\ell \ell^2 q_\ell - \sum_\ell \sum_{\ell'} \min\{\ell, \ell'\} p_\ell p_{\ell'}$$

$$- T(\sum_\ell \ell q_\ell)^2 + W(\sum_\ell \ell p_\ell)^2 - 2(\sum_\ell \ell p_\ell)(\sum_\ell \ell q_\ell)$$

$$1 - WT.$$  

(8)
It is convenient to write this in terms of quadratic forms in $P$ and $Q$, the $S$-dimensional vectors with $\ell$th element $p_\ell$ and $q_\ell$, respectively. Using the fact that $\sum p_\ell = 1$ and $\sum q_\ell = W$, (8) can be written as

\[
N^{-1}\sigma^2_\epsilon \text{var}(\hat{\theta})^{-1} = P^T \Xi P - Q^T \tilde{\Lambda} P - \left( \frac{TQ^T \Delta Q + WP^T \Delta P - 2Q^T \Delta P}{1 - WT} \right),
\]

(9)

where

\[
\Xi_{\ell,\ell'} = \frac{1}{2} | \ell - \ell' |
\]

\[
\tilde{\Lambda}_{\ell,\ell'} = \ell^2
\]

\[
\Delta_{\ell,\ell'} = \ell \ell',
\]

so $\tilde{\Lambda} = y1_S^T$ and $\Delta = zz^T$, where $z^T = (1, 2, \ldots, S)$, $y^T = (1^2, 2^2, \ldots, S^2)$.

The aim is to find an allocation of the $N_i$ to sequences which yields $P, Q$ that maximize (9). It is helpful to define

\[
u_i = p_i - \frac{q_i}{W};
\]

as $\sum p_i = 1$ and $\sum q_i = W$, these quantities will sum to zero. If $U$ denotes the $S$-dimensional vector of the $u_i$ (defined from $u_i$ as $p_i$ is from $p_\ell$), then $Q$ can be eliminated from (9) using $Q = WP - WU$. This gives

\[
N^{-1}\sigma^2_\epsilon \text{var}(\hat{\theta})^{-1} = P^T \Xi P + WU^T \tilde{\Lambda} P
- 2WU^T \Delta P - \frac{TW^2}{1 - WT} U^T \Delta U
\]

(10)

where $\Lambda = \frac{1}{2}(\tilde{\Lambda} + \tilde{\Lambda}^T)$.

3 Finding optimal and approximately optimal designs

Our approach to finding highly efficient designs starts by identifying optimal designs in three special cases

3.1 Some special cases

The three special cases are: i) when the clustering effect is dominant, $\lambda \to 0$; ii) there is no cluster effect, $\lambda \to \infty$; and iii) when all the clusters have the
same size, \( N_i = n \). It is useful to note that
\[
0 < W = \frac{\sum N_i w_i}{\sum N_i} \leq \frac{1}{T},
\]
because \( 0 < w_i \leq T^{-1} \), with \( W = T^{-1} \) when \( \lambda = 0 \) and \( W \to 0 \) as \( \lambda \to \infty \). When \( N_i = n \) for all \( i \), \( W = n/(\lambda + nT) \).

For these special cases, all the terms involving \( U \) in (10) vanish. This follows from the following observations.

For \( \lambda \to 0 \): \( 1 - WT \to 0 \) linearly in \( \lambda \): \( u_i = p_i - q_i/W \to 0 \), linearly in \( \lambda \), and hence all terms in \( U \) tend to 0.

For \( \lambda \to \infty \): \( \frac{q_i}{W} \to N_i^2/\sum_j N_j^2 \) (i.e. \( u_i \) is finite in the limit) and \( W \to 0 \).

For \( N_i = n \): \( W = n/(\lambda + nT) \) and \( q_i = p_i n/(\lambda + nT) = p_i W \) so \( U = 0 \).

The optimal designs for these three cases will follow if we identify the value of \( P \) which maximises \( P^T \Gamma P \), subject to \( 1_T S P = 1 \), where
\[
\Gamma = \Gamma(W) = \Xi - W\Lambda + W\Delta.
\]
Introducing the Lagrangian \( \mathcal{L} = P^T \Gamma P - \mu 1_S^T P \), differentiating and setting equal to zero gives the equation for the optimal \( P \), \( P_{opt} \), as
\[
P_{opt} = \frac{\Gamma^{-1}1_S}{1_S^T \Gamma^{-1}1_S}.
\]
If \( e_S \) is defined to be the \( S \)-dimensional vector with first and last element equal to 1 and zero elsewhere, then scalar \( a \) and \( b \) can be found so that \( \Gamma(a1_S + be_S) \propto 1_S \). Consequently the optimal symmetric design gives one weight to sequences 1 and \( S \) and another weight to all the other sequences, namely
\[
P_{opt}^T = \left( \frac{1}{2} - \frac{1}{2}W(S - 2), W, W, \ldots, W, W; \frac{1}{2} - \frac{1}{2}W(S - 2) \right). \quad (11)
\]

The optimal designs for the cases with \( \lambda = 0 \), \( N_i = n \) and \( \lambda \to \infty \) are shown in Table 2. Note that writing the weights for the case \( N_i = n \) in terms of \( \rho = 1/(1 + \lambda) \) gives the expressions presented in Theorem 1 of Lawrie et al. (2015).
Table 2: Proportions of patients on the sequences of optimal SWDs for extreme values of the intra-class correlation and for the case of equal cluster-period size.

| λ | \(\ell = 2, \ldots, S - 1\) | \(\ell = 1\) and \(S\) |
|---|---|---|
| 0 | \(\frac{1}{T}\) | \(\frac{3}{2T}\) |
| \(N_i = n\) | \(\frac{n}{\lambda + nT}\) | \(\frac{\lambda + 3n}{2(\lambda + nT)}\) |
| \(\infty\) | 0 | \(\frac{1}{2}\) |

### 3.2 The general case: symmetric designs and approximate optimality

There is an inherent symmetry in the design in Table 1 which can usefully be exploited in the search for highly efficient designs. As the stochastic element of (1) is temporally reversible, the design with the ‘reflected’ allocations, i.e. with the allocations to sequence 1 & \(S\), 2 & \(S - 1\), and in general sequences \(\ell & S - \ell + 1\), interchanged will have the same value of \(\text{var}(\hat{\theta})\). A design where \(p_\ell = p_{S-\ell+1}\) and \(q_\ell = q_{S-\ell+1}\), \(\ell = 1, \ldots, S\) is referred to as symmetric and for these designs \(P = RP\) and \(Q = RQ\), where \(R\) is a matrix with 1s on the trailing diagonal and 0s elsewhere. In Appendix B it is explained that among optimal designs there will always be a symmetric design, so in seeking an optimal design only symmetric designs need to be considered.

It should be noted: i) that if \(P\) and \(Q\) are symmetric, then so too is \(U\), i.e. \(RU = U\); and ii) that \(\Delta + \Delta R = (S + 1)z_1^T\), and therefore \(\Delta U = \frac{1}{T}(\Delta U + \Delta RU) = \frac{1}{T}(S + 1)z_1^T U = 0\). This means that for symmetric designs the expression in (10) simplifies to

\[
N^{-1}\sigma^2\text{var}(\hat{\theta})^{-1} = P^T(\Xi - W\Lambda + W\Delta)P + WU^T\tilde{\Lambda}P
\]

(12)

The variance of \(\hat{\theta}\) depends on \(\lambda\), the \(N_i\) and the design through the \(p_\ell\) and \(q_\ell\), \(\ell = 1, \ldots, S\). Optimising over these quantities is complicated by the fact that while \(p_i\) determines \(q_i\) (given \(N, T, \lambda\)), it does so non-linearly \((q_i = Np_i^2/\{\lambda + NTp_i\})\) and hence \(p_i\) does not determine \(q_\ell\). However, for values of \(N_i\), \(T\) and \(\lambda\) which are likely to arise in practice, the relation between \(q_i\) and \(p_i\) will often be very close to linear (see Figures 1 and 2) and we will exploit this fact to make some analytic progress.

Given this observation it is possible that for many proposed studies the elements of the vector \(WU, WP_\ell - q_\ell\), will be small in magnitude relative to
Figure 1: The relationship between the $p_i$ and $q_i$ for $\lambda = 10$. Circles denote $T = 4$ and crosses $T = 10$. The second column shows the values of $W p_i - q_i$, with a horizontal line at zero: note the vertical scale in second column. First row is for 40 clusters, with $N_i$ drawn from a Poisson distribution of mean 30, the second row is for 20 clusters with $N_i$ from a Poisson distribution of mean 30 and 20 with mean 10.
Figure 2: The relationship between the $p_i$ and $q_i$ for $\lambda = 100$. Circles denote $T = 4$ and crosses $T = 10$. The second column shows the values of $W_{p_i} - q_i$, with a horizontal line at zero: note the vertical scale in second column. First row is for 40 clusters, with $N_i$ drawn from a Poisson distribution of mean 30, the second row is for 20 clusters with $N_i$ from a Poisson distribution of mean 30 and 20 with mean 10.
In this case it is at least plausible that the term $WU^T \tilde{\Lambda} P$ in (12) could be neglected, and that the design found by maximising just the first term in (12) would be highly efficient. The calculations in Section 3.1 would then apply, with $W$ taking the value determined by the proposed $N_i, T$ and $\lambda$, and would not be restricted to the special cases in Table 2. The resulting design $P_{opt}$ would be as given in (11).

That $P_{opt}$ is the unique maximum of $P^T \Gamma(W)P$, subject to $P$ symmetric and $1_S^T P = 1$, can be established by the following argument. Any alternative symmetric design can be written as $P_{opt} + V$, where $V$ is symmetric, $1_S^T V = 0$ and $V$ is such that the elements of $P_{opt} + V$ remain in $[0,1]$. It should be noted that

$$(P_{opt} + V)^T \Gamma(P_{opt} + V) = P_{opt}^T \Gamma P_{opt} + V^T \Gamma V,$$

as the terms linear in $V$ vanish because $\Gamma P_{opt} \propto 1_S$. Also $1_S^T V = 0$ implies that $V^T \Lambda V = 0$ and as $V$ is symmetric $V^T \Delta V = 0$, so $V^T \Gamma V = V^T \Xi V$. It is shown in [Bünger (2014)] that

$$\Xi^* = \Xi - \frac{1}{4} (S - 1) 1_S 1_S^T$$

is negative semi-definite, with a one-dimensional kernel spanned by $e_S$. As $V$ cannot lie wholly in the kernel it follows that $0 > V^T \Xi^* V = V^T \Xi V$.

Consequently the design $P_{opt}$ might be referred to as an approximately optimal design. However, further investigation of the effect of neglecting the final term in (12) is required.

### 3.3 The effect of the term $WU^T \tilde{\Lambda} P$

An initial view of the relative importance of the terms in (12) can be seen in Figures 3 and 4, where in each plot the values of $P^T \Gamma P$ and $WU^T \tilde{\Lambda} P$ are shown for 1000 random symmetric allocations (these figures appear in colour in the electronic version of this article). The upper cloud of points are the values of the first term and the lower cloud shows the second term. Two patterns of $N_i$ are considered and $\lambda = 10$ and $\lambda = 100$ are illustrated (corresponding to ICCs of approximately 0.1 and 0.01). In one pattern 40 $N_i$ are from a Poisson distribution with mean 30, representing reasonably homogenous clusters. In the other pattern, 20 of the $N_i$ are from a Poisson distribution with mean 10 and the remainder are from a Poisson distribution of mean 30, chosen to represent the case sometimes encountered in practice, where some clusters are ‘large’ and some are ‘small’, as might arise, e.g., with
Figure 3: $\lambda = 10$. The first row is for 40 clusters, Poisson mean 30, the second row has 20 clusters from a Poisson mean 30 and 20 with mean 10. First column $T = 4$, second $T = 10$. The values of $P^T\Gamma P$ (black circles) and $WU^T\tilde{A}P$ (red crosses) are shown for 1000 symmetric random allocations. This figure appears in colour in the electronic version of this article, and the colours of the symbols refer to that version.

In all cases the value of $P^T\Gamma P$ is substantially larger than $WU^T\tilde{A}P$. The latter varies about 0, and it is possible that relatively large positive values of both terms arise for the same allocation, such the maximum of $P^T\Gamma P + WU^T\tilde{A}P$ exceeds that of $P^T\Gamma P$. However, given that the value of $WU^T\tilde{A}P$ is substantially smaller than that of $P^T\Gamma P$, maximising \cite{12} is unlikely to produce a design with a markedly smaller $\text{var}(\hat{\theta})$ than that available for the approximately optimal design.

A more thorough analysis starts by noting that

\[ WU^T\tilde{A}P = W \sum_{\ell=1}^{s} u_{\ell}\ell^2. \]
Figure 4: $\lambda = 100$. The first row is for 40 clusters, Poisson mean 30, the second row has 20 clusters from a Poisson mean 30 and 20 with mean 10. First column $T = 4$, second $T = 10$. The values of $P^T \Gamma P$ (black circles) and $WU^T \tilde{\Lambda} P$ (red crosses) are shown for 1000 symmetric random allocations. This figure appears in colour in the electronic version of this article, and the colours of the symbols refer to that version.
As the $u_\ell$ are symmetric and sum to zero, this can be rewritten as

$$WU^T\tilde{\Lambda}P = \frac{1}{2}W \sum_{\ell=1}^{S} u_\ell \{ \ell^2 + (S + 1 - \ell)^2 \} = W \sum_{\ell=1}^{S} u_\ell \{ \ell - \frac{1}{2}(S + 1) \}^2.$$  

In consequence, $WU^T\tilde{\Lambda}P$ is maximised for symmetric designs by a design in which all the clusters where $u_i > 0$ are allocated equally to sequences 1 and $S$ and those with $u_i < 0$ are allocated to the sequences minimising $(\ell - \frac{1}{2}(S + 1))^2$, which will be sequence $\ell = \frac{1}{2}(S + 1)$, ($S$ odd - i.e the middle sequence) or to sequences $\frac{1}{2}S$ and $\frac{1}{2}S + 1$, ($S$ even, i.e. the middle two sequences). So the maximum value of $WU^T\tilde{\Lambda}P$ is

$$A_{\text{max}} = \frac{1}{4}WU_+(S - 1)^2 \quad S \text{ odd}$$  

or

$$A_{\text{max}} = \frac{1}{4}WU_+S(S - 2) \quad S \text{ even},$$  

where $U_+ = \sum_{u_i>0} u_i$. Noting that $u_i > 0$ if and only if $N_i/(\lambda + N_iT) < W$, this design will allocate the smaller clusters to sequences 1 and $S$ and the larger sequences to the middle sequences - a design which is very different from the approximately optimal design.

The maximised value $P_{\text{opt}}^T\Gamma P_{\text{opt}}$ is

$$P_{\text{opt}}^T\Gamma P_{\text{opt}} = M_{\text{opt}} = \frac{1}{12}(S - 1)(3 - 3(S - 1)W + S(S - 2)W^2),$$  

so the maximum of (12) cannot exceed $M_{\text{opt}} + A_{\text{max}}$, and hence $\text{var}(\hat{\theta}_{\text{approx}})/\text{var}(\hat{\theta}_{\text{exact}})$ cannot exceed $1 + (A_{\text{max}}/M_{\text{opt}})$. Indeed, this is likely to be a very conservative upper bound, because the $M_{\text{opt}}$ and $A_{\text{max}}$ arise from quite different designs.

A numerical study was conducted to elucidate further the role of the term $WU^T\tilde{\Lambda}P$. All designs had 40 clusters and three patterns of cluster-period size were considered. In each 20 clusters are generated independently, and the remaining 20 clusters are duplicates of the first 20, to ensure that a symmetrical design is possible. The patterns are those in Figures 3 and 4, with the addition of the case with smaller $N_i$, where all clusters are chosen from a Poisson distribution with mean 10. The same combinations of $\lambda$ and $T$ are also considered.

Once the $N_i$ are chosen, the value of $W$ and $U_+$ can be found for each combination of $\lambda$ and $T$, and (15) and (13) can be evaluated. For each
Table 3: Results of $4 \times 10^6$ symmetric allocations for twelve combinations of $\lambda$, $T$ and cluster type. Cluster type 1, 40 clusters, mean $N_1$ 10; type 2, 40 clusters, 20 mean 10 and 20 mean 30; type 3, 40 clusters mean 30. $M_{\text{opt}}$ is [15], $A_{\text{max}}$ is [13]. $E_1$ is the upper bound on the efficiency $1 + A_{\text{max}}/M_{\text{opt}}$. $A_{\text{Emp}}^{\text{max}}$ is the maximum of $WU^T \bar{\Lambda}P$ and $M + A$ is the maximum of (12) over the $4 \times 10^6$ allocations; $E_2$ is the efficiency based on $M_{\text{opt}}$ and $M + A$.

| Cluster | $\lambda$ | $T$ | $W$   | $M_{\text{opt}}$ | $A_{\text{max}}$ | $E_1$ | $A_{\text{Emp}}^{\text{max}}$ | $M + A$ | $E_2$ |
|---------|-----------|-----|-------|------------------|------------------|------|------------------------|--------|------|
| 1       | 10        | 4   | 0.2034| 0.3173           | 0.0051           | 1.0161| 0.0051                 | 0.3201 | 1.0086 |
| 1       | 10        | 10  | 0.0915| 0.8877           | 0.0172           | 1.0194| 0.0148                 | 0.8985 | 1.0121 |
| 1       | 100       | 4   | 0.0788| 0.4243           | 0.0068           | 1.0160| 0.0068                 | 0.4261 | 1.0042 |
| 1       | 100       | 10  | 0.0530| 1.2704           | 0.0514           | 1.0404| 0.0437                 | 1.2939 | 1.0185 |
| 2       | 10        | 4   | 0.2236| 0.3014           | 0.0058           | 1.0192| 0.0058                 | 0.3032 | 1.0061 |
| 2       | 10        | 10  | 0.0953| 0.8564           | 0.0185           | 1.0216| 0.0149                 | 0.8656 | 1.0108 |
| 2       | 100       | 4   | 0.1235| 0.3841           | 0.0116           | 1.0303| 0.0116                 | 0.3867 | 1.0067 |
| 2       | 100       | 10  | 0.0697| 1.0885           | 0.0711           | 1.0653| 0.0562                 | 1.1194 | 1.0285 |
| 3       | 10        | 4   | 0.2308| 0.2958           | 0.0010           | 1.0034| 0.0010                 | 0.2965 | 1.0024 |
| 3       | 10        | 10  | 0.0968| 0.8449           | 0.0028           | 1.0033| 0.0025                 | 0.8467 | 1.0020 |
| 3       | 100       | 4   | 0.1369| 0.3724           | 0.0035           | 1.0094| 0.0035                 | 0.3743 | 1.0049 |
| 3       | 100       | 10  | 0.0751| 1.0353           | 0.0170           | 1.0164| 0.0150                 | 1.0452 | 1.0095 |

Combination $4 \times 10^6$ random symmetric allocations of clusters to sequences (not necessarily distinct) were generated and the maximum of each of $P^T \Gamma P$, $WU^T \bar{\Lambda}P$ and $P^T \Gamma P + WU^T \bar{\Lambda}P$ were evaluated.

The empirical maximum of $P^T \Gamma P$ is omitted from Table 3 as it coincides with the value [13]. The maximum of (12) is in column $M + A$ and these exceed $M_{\text{opt}}$, but never by more than 3% and usually much less than that. The estimated efficiency $E_1 = 1 + A_{\text{max}}/M_{\text{opt}}$ is always less than 1.065, but again is usually much less than this. A similar pattern is revealed by column $E_2$, where the maximum of $M + A$ is compared with $M_{\text{opt}}$. The values of $E_2$ are substantially lower than those of $E_1$, confirming that the bound $1 + A_{\text{max}}/M_{\text{opt}}$ is conservative. These figures indicate that the approximately optimal design is indeed highly efficient. The efficiency is slightly worse for larger $\lambda$ and $T$, but is entirely adequate even in these cases.

One of the advantages of the approximately optimal design is that the adequacy of the approximation can be assessed by computing $W$ and $U_+$, both of which are available a priori.
3.4 Designs with equal allocation

The design which allocates equally to sequences, i.e. with $P_i = S - 1/S$, is perhaps the most commonly used version of the SWD. For this design $P^T \Gamma P = \{(S^2 - 1)(2 - W S)\}/(12S)$ and consequently the efficiency of the equal allocation design can be found as

$$\frac{\text{var}(\hat{\theta}_{\text{equal}})}{\text{var}(\hat{\theta}_{\text{opt}})} = \frac{(3 - 3(S - 1)W + S(S - 2)W^2)S}{(S + 1)(2 - WS)}.$$  \hfill (16)

This is a decreasing function of $W$, as would be anticipated given that the optimal design makes the allocations more equal as $W$ increases. When $W = 1/(S + 1)$, its maximum value, the above ratio is $1 + (S - 2)/(S^3 + 4S^2 + 5S + 2)$, which never exceeds 76/75. At $W = 0$, the ratio is $3/2(1 - (S + 1)^{-1})$, which increases to $3\frac{1}{2}$ as $S$ increases. So, for small $W$, the equal allocation design can be noticeably inefficient.

4 Closed-cohort designs

In the closed-cohort SWD, the values $Y_{i1k}, Y_{i2k}, \ldots, Y_{iTk}$ are observed on the same individual and therefore the model \[ \text{Section 2.1} \] needs to be amended to allow for dependence between these quantities. A random variable $\zeta_{ik}$, specific to individual $k$ in cluster $i$ is included as

$$Y_{ijk} = \beta_j + X_{ij}\theta + \alpha_i + \zeta_{ik} + \epsilon_{ijk},$$  \hfill (17)

where $\zeta_{ik}$ has zero mean, variance $\omega^2$ and is independent of all other random variables in the model. The vector $Y_i$ defined in Section 2.1 now has variance $V_i = (\tau^2 + N_i^{-1}\omega^2)J_T + N_i^{-1}\sigma_e^2I_T$, and the inverse of this matrix can be written as

$$V_i^{-1} = \sigma_e^{-2}\left(N_iI_T - \frac{N_i(N_i + \lambda_2)}{N_iT + \lambda_2T + \lambda_1}J_T\right),$$

where $\lambda = \sigma_e^2/\tau^2$ has been rewritten as $\lambda_1$ and $\lambda_2 = \omega^2/\tau^2$.

The $\text{var}(\hat{\theta})$ for a closed-cohort SWD can be found by repeating the calculations in Section 2.1 but using the above expression for $V_i^{-1}$. These are briefly rehearsed in Appendix C and give very similar results to the cross-sectional case. If $\tilde{w}_i$ and $\hat{q}_i$ are defined as

$$\tilde{w}_i = \frac{N_i + \lambda_2}{N_iT + \lambda_2T + \lambda_1}; \quad \hat{q}_i = \frac{N_i\tilde{w}_i}{N} = \frac{N_i(N_i + \lambda_2)}{N(N_iT + \lambda_2T + \lambda_1)}.$$
then (2) remains valid, provided that \( w_i \) is replaced by \( \tilde{w}_i \) in the definitions of \( F, H \) and \( W \). Also, (8) still obtains if \( \tilde{q}_\ell \), defined from \( \tilde{q}_i \) as \( q_\ell \) was defined from \( q_i \), is used in place of \( q_\ell \). Also \( W \) must be replaced by \( \tilde{W} \), with \( \tilde{W} = \sum \tilde{q}_i = \sum \tilde{q}_\ell \): note that \( 0 < \tilde{W} \leq T^{-1} \). Also, if \( \tilde{u}_i = p_i - \tilde{W}^{-1}\tilde{q}_i \), then \( \tilde{U} \) can be defined analogously to \( U \).

Exact optimal designs for the special cases considered Section 3.1 can now be found by maximising \( P^T \Gamma(\tilde{W})P \), because the terms in \( \tilde{W} \tilde{U} \) vanish in the same way that those in \( WU \) did for the cross-sectional case. The cases \( \lambda_1 = 0 \) and \( \lambda_1 \to \infty \), now for fixed \( \lambda_2 \), remain unchanged with \( \tilde{W} = T^{-1}, 0 \) respectively. For fixed \( \lambda_1 \), \( \lambda_2 = 0 \) is simply the cross-sectional case, whereas \( \lambda_2 \to \infty \) gives \( \tilde{W} = T^{-1} \). It is to be expected that the cases \( \lambda_2 \to \infty \) and \( \lambda_1 = 0 \) should coincide because the former corresponds to \( \omega^2 \) becoming the dominant variance component, which would be expected to have a similar effect to \( \tau^2 \) being dominant, which corresponds to \( \lambda_1 \to 0 \).

The case of equal cluster sizes \( N_i = n \) puts weight
\[
\tilde{W} = \frac{n + \lambda_2}{NT + \lambda_2 T + \lambda_1} = \frac{n\tau^2 + \omega^2}{nT\tau^2 + T\omega^2 + \sigma_e^2}
\]
on sequences 2 to \( S - 1 \), with complementary and equal weights on the other sequences. The second form above can readily be found to reproduce the result in [Li et al. (2018)](2018): note that their parameter \( \sigma^2_\pi \) needs to be set to 0 for their model to correspond with (17).

For the general case the relation between \( p_i \) and \( \tilde{q}_i \) is very similar to that between \( p_i \) and \( q_i \), so the arguments rehearsed in Section 3.2 apply to the closed-cohort case, hence the approximately optimal design is found by maximising \( P^T \Gamma(\tilde{W})P \). The efficiency of the design can be assessed in the same way as before, using (13) or (14), provided \( W, U_+ \) are replaced with \( \tilde{W}, \tilde{U}_+ \), the latter defined in the obvious way.

5 Discussion

The approach followed in this paper has produced exact optimal designs for some special cases and, more usefully, has shown that essentially the same form of design is highly efficient in all cases. The design gives the same weight \( W \) to sequences 2 to \( S - 1 \) and a second weight to the extreme sequences. When there is no cluster-level variance component, so all the observations (in the cross-sectional design) are independent, \( W = 0 \) and
the optimal design uses only the first and last sequences. This is sensible because independent observations mean that within-sequence contrasts have no advantage and within-period differences will eliminate the period parameters most effectively. By using only the first and last sequences, all the comparisons between sequences but for the first and last periods provide information on the treatment effect. The form of the design in the other extreme case, $W = T^{-1}$ is more intriguing, with the first and last sequences still receiving 50% more weight than the other sequences.

Investigators who find the simplicity of the equally weighted design compelling can take some comfort from the ratio in (16), which indicates that the equally weighted design has good efficiency except for smaller values of $W$. An advantage of the method in this paper is that much depends on the value of $W$ and in many studies the investigator will have sufficient information at the outset to compute a value, or perhaps a plausible range of values, for $W$. Not only will this allow an assessment of the efficiency of the equally weighted design, it will also allow an assessment of the efficiency of the approximately optimal design.

The results presented in this paper are in terms of the proportions of individuals allocated to the different sequences. However, the proportions of clusters allocated to sequences do not feature explicitly in the specification of the designs. This may seem surprising because, in rather loose terms, the amount of information carried by a given number of individuals in a cluster will vary with the intra-cluster correlation. So the number of clusters, as well as the number of individuals allocated to a sequence, might be expected to play a role. In fact, how the individuals are distributed among clusters is reflected in the quantity $W$ (or $\tilde{W}$). For example, if 30 individuals are allocated to a sequence in a single cluster for a cross-sectional study with $T = 4$ and $\lambda = 100$, then the $q_i$ for that cluster is $4.09/N$, whereas if the 30 individuals had been allocated in three clusters of size 10, the sum of the corresponding $q_i$ would be $2.14/N$.

Some issues remain open and will form the focus of future work. The best way to randomize clusters to achieve a design as close to optimal in real applications deserves consideration. In practice some sets of $N_i$ may not permit a perfectly symmetric design, so it would be useful to investigate how best to organise allocation so that a design as close to symmetric is obtained. It may be that for smaller numbers of clusters it is awkward to arrange allocation so that the specifications given in this paper are met. In these cases it is possible that exhaustive numerical searches will provide
a practical alternative. This paper has worked with the the classic SWD shown in Table 1: extensions to other forms of SWD merit attention - such as designs with sequences where treatment allocation is constant (see Girling and Hemming [2016], and designs in which either the first or last period, or both, are omitted or perhaps shortened (see Thompson et al. [2017]).

References

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). Optimum Experimental Designs, with SAS. Oxford University Press, Oxford.

Bünger, F. (2014). Inverses, determinants, eigenvalues, and eigenvectors of real symmetric Toeplitz matrices with linearly increasing entries. Linear Algebra and its Applications 459, 595–619.

Donner, A. and Klar, N. (2000). Design and Analysis of Cluster Randomization Trials in Health Research. Arnold, London.

Girling, A. J. and Hemming, K. (2016). Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. Statistics in Medicine 35, 2149–2166.

Grayling, M. J., Mander, A. P., and Wason, J. M. S. (2018). Admissible multi-arm stepped-wedge cluster randomized trial designs. arXiv:1710.03496v2.

Hayes, R. J. and Moulton, L. H. (2017). Cluster Randomized Trials. Chapman & Hall/CRC, Boca Raton, 2nd edition.

Hooper, R., Teerenstra, S., de Hoop, E., and Eldridge, S. (2016). Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Statistics in Medicine 35, 4718–4728.

Hussey, M. A. and Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials 28, 182–191.

Kasza, J. and Forbes, A. B. (2017). Information content of cluster-period cells in stepped wedge trials. Biometrics.

Kasza, J., Hemming, K., Hooper, R., Matthews, J. N. S., and Forbes, A. B. (2018). Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Statistical Methods in Medical Research, ?–?.
Lawrie, J., Carlin, J. B., and Forbes, A. B. (2015). Optimal stepped wedge
designs. *Statistics and Probability Letters* **99**, 210–214.

Li, F., Turner, E. L., and Preisser, J. S. (2018). Optimal allocation of clusters
in cohort stepped wedge designs. *Statistics and Probability Letters* **137**, 257–263.

Lyons, V. H., Li, L., Hughes, J. P., and Rowhani-Rahbar, A. (2017). Proposed variations of the stepped wedge design can be employed to accommodate multiple interventions. *Journal of Clinical Epidemiology* **86**, 160–167.

Matthews, J. N. S. and Forbes, A. B. (2017). Stepped wedge designs: insights from a design of experiments perspective. *Statistics in Medicine* **36**, 3772–3790.

Mdege, N. D., Man, M., Taylor, C. A., and Torgerson, D. J. (2011). Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *Journal of Clinical Epidemiology* **64**, 936–948.

Thompson, J. A., Fielding, K., Hargreaves, J., and Copas, A. (2017). The optimal allocation of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. *Clinical Trials* **14**, 639–647.

**Appendix A**

The matrix \( V_i = \tau^2 J_T + \sigma_e^2 N_i^{-1} I_T \), so

\[
V_i^{-1} = N_i \frac{\tau^2}{\sigma_e^2} (I_T - w_i J_T),
\]

where \( w_i = N_i \tau^2 / (\sigma_e^2 + N_i \tau^2 T) = N_i / (\lambda + N_i T) \). Therefore \( \sum V_i^{-1} = \sigma_e^{-2} N (I_T - W J_T) \) and from this it follows that

\[
(\sum V_i^{-1})^{-1} = \frac{\sigma_e^2}{N} (I_T + \frac{W}{1 - W T} J_T).
\]

**Evaluating** \( (\sum D_i^T V_i^{-1} (\sum V_i^{-1})^{-1} (\sum V_i^{-1} D_i)) \)

If we denote by \( r_i \) the number of times treatment B appears in the sequence to which cluster \( i \) is allocated, then \( r_i = D_i^T D_i = 1_i^T D_i \). Consequently
\[ V_i^{-1}D_i = \sigma_e^{-2}N_i(D_i - r_iw_i1_T) \quad \text{and therefore} \]
\[
\sum_{i=1}^{C} V_i^{-1}D_i = \sigma_e^{-2} \left\{ \sum_{i=1}^{C} N_i D_i - \left( \sum_{i=1}^{C} N_i r_i w_i \right) 1_T \right\}.
\]

It follows that \((\sum V_i^{-1})^{-1}(\sum V_i^{-1}D_i)\) is
\[
\frac{1}{N} \left( \sum_{i=1}^{C} N_i D_i - M1_T \right)
\]
where
\[
M = \frac{1}{1 - WT} \sum_{i=1}^{C} N_i r_i w_i - \frac{W}{1 - WT} \sum_{i=1}^{C} N_i r_i = \frac{F - WE}{1 - WT},
\]
and we have written \(E = \sum_{i=1}^{C} N_i r_i\) and \(F = \sum_{i=1}^{C} N_i w_i r_i\). It follows that
\[
(\sum D_i^T V_i^{-1})(\sum V_i^{-1})^{-1}(\sum V_i^{-1}D_i) = N^{-1}\sigma_e^{-2} \left( G + \frac{TF^2 + WE^2 - 2EF}{1 - WT} \right)
\]
where \(G = (\sum N_i D_i^T)(\sum N_i D_i)\).

**Evaluating \(\sum D_i^T V_i^{-1}D_i\)**

We have
\[
\sum_{i=1}^{C} D_i^T V_i^{-1}D_i = \sigma_e^{-2} \sum_{i=1}^{C} N_i D_i^T(I_T - w_i J_T)D_i
\]
\[
= \sigma_e^{-2} \left( \sum_{i=1}^{C} N_i r_i - \sum_{i=1}^{C} N_i w_i r_i^2 \right)
\]
\[
= \sigma_e^{-2} (E - H),
\]
with \(H = \sum_{i=1}^{C} N_i w_i r_i^2\).

Putting the above together we obtain
\[
\sigma_e^2 \text{var}(\hat{\theta})^{-1} = E - H - \frac{G}{N} - \frac{TF^2 + WE^2 - 2EF}{N(1 - WT)}.
\]
Appendix B

Suppose that $D_0$ denotes a design with total sample size $N$ and allocation vectors $P$ and $Q$. The estimator of $\theta$ based on this design has variance $v_0$. Suppose also that $D_1$ denotes the design with the same pattern of cluster allocation but with half the number of clusters, then $P$ and $Q$ are unchanged but the total sample size is now $\frac{1}{2}N$. The data from $D_1$ is denoted by $Y_1$, and the estimator of $\theta$ is $a_1^T \hat{Y}_1$, for some suitable vector $a_1$: the expectation of $a_1^T Y_1$ is $\theta$ and its variance is $v_1$. As $P, Q$ are unchanged and $N$ has halved in value, $v_1 = 2v_0$. Suppose that $\tilde{D}_1$ is the reverse of $D_1$, i.e. where $P, Q$ are replaced with $RP, RQ$ respectively, with $R$ being the matrix which reverses the order of a vector. As the stochastic part of the model in equation (1) of the main article is temporally reversible, the estimator of $\theta$ from $\tilde{D}_1$, namely $a_1^T \tilde{Y}_1$, is also $v_1$: this would also be the case under any centro-symmetric dispersion structure - see Kasza and Forbes (201).

Consider now the design $D_0^R = D_1 + \tilde{D}_1$, comprising a sample of size $\frac{1}{2}N$ allocated to each of $D_1$ and $\tilde{D}_1$: $D_0^R$ can be thought of as a symmetrised version of $D_0$. A linear unbiased estimator of $\theta$ is $\frac{1}{2}a_1^T Y_1 + \frac{1}{2}a_1^T \tilde{Y}_1$ and this has variance $\frac{1}{2}(v_1 + v_1) = v_0$. As $\frac{1}{2}a_1^T Y_1 + \frac{1}{2}a_1^T \tilde{Y}_1$ may not be the best linear unbiased estimator for $D_0^R$, this shows that the best estimator for this design has variance less than or equal to that of $D_0$. So, the symmetrised version of any design must be at least as good as the original design, so when searching for optimal designs attention can be restricted to symmetric designs where $P = RP, Q = RQ$.

The manoeuvre to go from $D_0$ to $D_1$ would be accomplished if, e.g., the sizes of the clusters allocated to sequence $\ell$ remain the same, but the number of each size is halved. In practice this might not be possible but the approach is in line with the use of so-called continuous designs in optimal design theory, where the weights allocated to design points can vary continuously, even if they do not correspond to whole numbers of units at each design point (see Atkinson et al., 2007, p.119).

Appendix C

The model for the closed-cohort design, namely

$$Y_{ijk} = \beta_j + X_{ij}\theta + \alpha_i + \zeta_{ik} + \epsilon_{ijk},$$
gives rise to a vector of responses on an individual of \((Y_{i1k}, Y_{i2k}, \ldots, Y_{iTk})^T\), 
\(k = 1, \ldots, N_i\) and the mean of these for cluster is \(Y_i\), which has variance 
\(V_i = (\tau^2 + \omega^2)J_T + N_i^{-1}\sigma^2_e I_T\), with inverse 
\[V_i^{-1} = \frac{N_i}{\sigma^2_e} (I_T - \tilde{w}_i J_T); \quad \tilde{w}_i = \frac{N_i + \lambda_2}{N_i T + \lambda_2 T + \lambda_1}\]

where \(\lambda_1 = \frac{\sigma^2_e}{\tau^2}\) (i.e. what was called \(\lambda\) in the cross-sectional case) and 
\(\lambda_2 = \frac{\omega^2}{\tau^2}\). It follows that \(\sum V_i^{-1} = N \sigma^2_e (I_T - \tilde{W} J_T)\), where 
\[\tilde{W} = \sum_{i=1}^C \frac{N_i}{N} \tilde{w}_i.\]

Also note that 
\[(\sum V_i^{-1})^{-1} = \frac{\sigma^2_e}{N} (I_T + \frac{\tilde{W}}{1 - \tilde{W} T} J_T).\]

It is also the case that \(V_i^{-1} D_i = \sigma^2_e N_i (D_i - \tilde{w}_i \tilde{r}_i 1_T)\).

These expressions are all identical to those found for the cross-sectional case, provided \(w_i\) is replaced with \(\tilde{w}_i\). The calculations which led to the different forms of \(\text{var}(\tilde{\theta})\) for the cross-sectional case are the same for the cohort case if \(W\) is replaced with \(\tilde{W}\) and \(q_i(q_\ell)\) are replaced with \(\tilde{q}_i (\tilde{q}_\ell)\), where 
\[\tilde{q}_i = \frac{N_i}{N} \tilde{w}_i = \frac{N_i (N_i + \lambda_2)}{N (N_i T + \lambda_2 T + \lambda_1)}.\]