Received: 30 November 2021 | Revised: 1 August 2022 | Accepted: 14 August 2022

DOI: 10.1002/bimj.202100383

Biometrical Journal

RESEARCH ARTICLE

Influence of cluster-period cells in stepped wedge cluster randomized trials

Philipp Mildenberger | **Jochem König**

Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center Johannes Gutenberg University Mainz, Mainz, Germany

Correspondence
Philipp Mildenberger, Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center Johannes Gutenberg University Mainz, D 55101 Mainz, Germany.
Email: pmildenb@uni-mainz.de

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

Stepped wedge cluster randomized trials (SWCRT) are increasingly used for the evaluation of complex interventions in health services research. They randomly allocate treatments to clusters that switch to intervention under investigation at variable time points without returning to control condition. The resulting unbalanced allocation over time periods and the uncertainty about the underlying correlation structures at cluster-level renders designing and analyzing SWCRTs a challenge. Adjusting for time trends is recommended, appropriate parameterizations depend on the particular context. For sample size calculation, the covariance structure and covariance parameters are usually assumed to be known. These assumptions greatly affect the influence single cluster-period cells have on the effect estimate. Thus, it is important to understand how cluster-period cells contribute to the treatment effect estimate. We therefore discuss two measures of cell influence. These are functions of the design characteristics and covariance structure only and can thus be calculated at the planning stage: the *coefficient matrix* as discussed by Matthews and Forbes and *information content (IC)* as introduced by Kasza and Forbes. The main result is a new formula for IC that is more general and computationally more efficient. The formula applies to any generalized least squares estimator, especially for any type of time trend adjustment or nonblock diagonal matrices. We further show a functional relationship between IC and the coefficient matrix. We give two examples that tie in with current literature. All discussed tools and methods are implemented in the R package `SteppedPower`.

KEYWORDS
cluster randomized trial, generalized least squares, information content, longitudinal data, stepped wedge

1 | INTRODUCTION

Stepped wedge cluster randomized trials (SWCRT) are a versatile alternative to parallel cluster randomized designs. They are increasingly popular in health services research for evaluation of complex interventions. In SWCRTs, the intervention to be investigated is introduced to clusters sequentially. SWCRTs are preferably used when logistical considerations make it feasible to implement the intervention sequentially or when the intervention under investigation is believed to have a benign effect because all clusters receive the intervention in the course of the study. However, this comes with a higher model complexity. The planning of SWCRT is challenging, due to many necessary design choices and assumptions on the correlation structure to be made. Hussey and Hughes (2007) introduced an approach for power calculation based on generalized least squares (GLS), which has since been refined (see, e.g., Li et al., 2021).

As treatment is allocated to clusters in an unbalanced manner and the proportion receiving the intervention under investigation is changing over time periods, the evidence contribution of cluster period cells can be highly variable. Hence, it may even be advisable and efficient to sample patients for outcome assessment in variable numbers for different cluster-period cells. Tools are needed for displaying the pattern of contributions of cluster-period cells to the treatment effect estimate.

Kasza and Forbes (2019) proposed the concept of *information content*, which is essentially the relative change in estimator variance when leaving out one cluster-period cell. Since explicit computation of *information content* requires recalculation of a modified effect estimator for every cluster for every time period and is hence computationally expensive, efficient calculation methods are of interest. Kasza and Forbes (2019) derived a formula that holds for all multiple-period cluster randomized trial designs with a factor adjustment for the time trend. This formula needs only one matrix inversion for each cluster-period cell instead of the two matrix inversions needed when computed explicitly.

Matthews and Forbes (2017) explored the *coefficient matrix* in the context of SWCRT. The intervention effect estimate is a linear combination of cluster-period cells; each entry of the coefficient matrix denotes the contribution of a particular cell, which we will call *cell contribution* throughout this paper.

We will explore these two methods to examine the influence of cluster-period cells on the treatment effect estimator. Section 2 will specify the setting and briefly explain the standard concept of power calculation for stepped wedge designs, following the GLS approach of Hussey and Hughes (2007). In Section 3, *cell contribution* and *information content* will be formally introduced. A new method to calculate the latter is derived, which generalizes to nonfactorial time trend specifications as we disclose the relationship between both measures. This is followed by illustrative examples in Section 4 and a critical reflection in Section 5.

1.1 | Real-world example

The PART-CHILD study evaluates a complex intervention aimed at improving participation centered care for children with special health care needs in interdisciplinary health care centers. The study is registered at www.drks.de with the ID DRKS00015054. The effectiveness evaluation is based on a stepped wedge cluster randomized trial implemented in 15 Social Pediatric Centers (Figure 1). All study sites receive the PART-CHILD intervention starting at one of five randomly allocated time points. After a prerollout period of 3 months, every 3 months three study sites start the intervention and thus cross from the control to the intervention condition. Together with a 3-month postrollout period after the last study sites received the intervention, this results in an overall duration of the effectiveness evaluation of 21 months comprising seven periods of 3 months. The first period under the intervention condition is considered as a training period. The primary outcome is a based on the collaborate score, which is based on a three-item questionnaire. As primary outcome, the degree of perceived shared decision-making with parents (CollaboRATEpediatric parent scale) is assessed on one randomly selected day per week during the entire study period, directly following care appointments. The scale comprises three items assessed on a 10-point Likert scale. The study was initially planned using the continuous CollaboRATEpediatric mean score. However, due to ceiling effects, the binary CollaboRATEpediatric top score was used for the primary analysis. Deviating from this, we consider the continuous score in this paper.

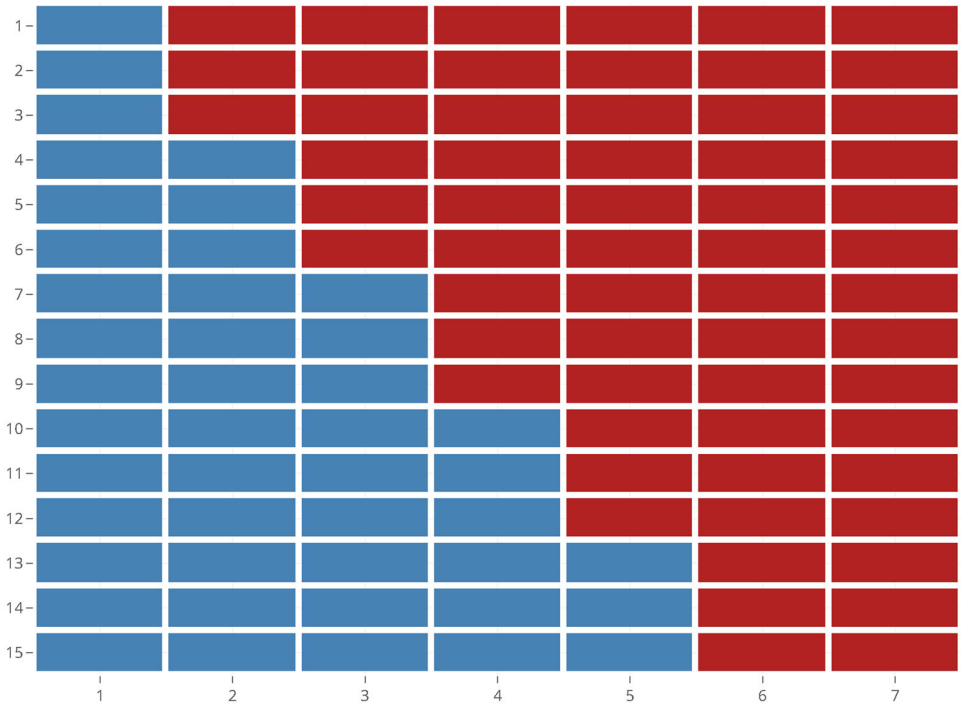

FIGURE 1 Design of the PART-CHILD study with 15 clusters in five sequences over seven quarters (i.e., 21 months). Control periods in blue, intervention periods in red

2 | NOTATION AND SETTING

A common approach to model the correlation in longitudinal studies is random effect. In general, such a model has the form

$$y_{ktn} = x_{kt}\theta + \mu_t + c_{kt} + e_{ktn}, \tag{1}$$

where y_{ktn} denotes the response in cluster $k \in \{1, \dots, K\}$ at time period $t \in \{1, \dots, T\}$ for individual $n \in \{1, \dots, N_{kt}\}$. x_{kt} indicates the treatment status (0 = control, 1 = interventional treatment) and θ the treatment effect. μ_t is the time trend at time t, $\mu = (\mu_1, \dots, \mu_T)$. It is generally recommended to incorporate a dedicated time effect into the model Barker et al. (2016). Hussey and Hughes (2007) suggest the use of a saturated (i.e., factorial) time trend, which—while performing best from a bias perspective—might not be feasible for every stepped wedge design. Thus, more parsimonious representations of the time trend are also considered in the literature Hemming et al. (2017); Nickless et al. (2018). Some authors even propose to use only a global mean μ_0 in trials with a short duration Zhou et al. (2020), although at the risk of a higher bias. A good overview can be found in Li et al. (2021).

Lastly, let c_{kt} be a random cluster effect for cluster k with $(c_{k1}, \dots, c_{kT}) \sim N(0, V_k)$ and e_{ktn} a residual error term with $e_{ktn} \sim N(0, \sigma^2)$. In practice, the structure of the covariance matrices V_k needs restraining assumptions for the model to be identifiable. The standard approach as introduced in Hussey and Hughes (2007) assumes the random cluster effect to be constant over time within each cluster, yielding an exchangeable within-cluster correlation structure. The random cluster effect can also be assumed to be first-order autoregressive, that is, $\text{corr}(c_{kt}, c_{kt'}) := r^{|t-t'|}$ with some autoregression parameter $r \in [0, 1]$. This is a special case of the *exponential decay model* of Kasza et al. (2019a). Often, all clusters are assumed to share the same correlation structure, yet there exist extensions that deviate from this assumption, notably when random intervention effects are included in the model Davey et al. (2015); Hughes et al. (2015). Further extensions exist, especially for closed cohort or open cohort designs in which individuals are observed over all or multiple study periods, respectively. The main result—the formula in Theorem 3.1—still holds for these extensions, with the only requirement being that the observations can be aggregated on cluster level. Such aggregation is possible if the cluster-period means form a sufficient statistic for θ Grantham et al. (2019).

Then, with $y_{kt} = 1/N_{kt} \sum_{n=1}^{N_{kt}} y_{ktn}$ and $e_{kt} = 1/N_{kt} \sum_{n=1}^{N_{kt}} e_{ktn}$, Equation (1) becomes

$$y_{kt} = x_{kt}\theta + \mu_t + c_{kt} + e_{kt}, \qquad e_{kt} \sim N\left(0, \frac{\sigma^2}{N_{kt}}\right).$$

Let us define $\beta := (\theta, \mu')' \in \mathbb{R}^p$ and $\omega_{kt} := c_{kt} + e_{kt}$. p is the total number of parameters to be estimated. This leads to a compact and more general notation of the above equation:

$$y_{kt} = X_{kt}\beta + \omega_{kt}.$$

In matrix notation:

$$y = X\beta + \omega, \tag{2}$$

where X is the corresponding design matrix and $\omega \sim N(0, \Omega)$, where Ω is a covariance matrix determined by V_k, σ, and N_{kt}. We are thus in a GLS setting, so the variance of $\hat{\beta}$ is

$$\text{Var}(\hat{\beta}) = (X'\Omega^{-1}X)^{-1}. \tag{3}$$

With $\text{Var}(\hat{\theta})$ the diagonal element of $\text{Var}(\hat{\beta})$ that corresponds to $\hat{\theta}(= \hat{\beta}_1)$, one can then calculate the power of a z-test as $\Phi(\frac{\theta - 0}{\sqrt{\text{Var}(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}})$.

3 | INFLUENCE OF CLUSTER-PERIOD CELLS

In this section, we discuss two influence diagnostics for GLS. *Cell contribution* yields estimates for the influence of observations on $\hat{\theta}$ (not for $\text{Var}(\hat{\theta})$), whereas the *information content* denotes the relative change in $\text{Var}(\hat{\theta})$ when one particular observation is omitted.

3.1 | Information content in generalized least squares estimation

Recall the general linear model

$$y = X\beta + \omega$$

with $\omega \sim N(0, \Omega)$ and Ω a positive definite, symmetric covariance matrix. Usually, Ω is a block diagonal matrix as defined in (2) but note that we do not require Ω to be block diagonal, that is, the observations do not necessarily need to stem from independent clusters for the following theorems to hold. Hence, for the remainder of this section, we establish a generalized and simplified notation: We denote single observations (i.e., cluster-period means in the context of stepped wedge designs aggregated on cluster level as explained in Section 2) with i instead of kt and thus $X \in \mathbb{R}^{I \times p}, I = KT$, with p the number of parameters to be fitted in the GLS model.

The design matrix X can be seen as a composite of a treatment vector x_1 and an adjustment matrix X_2, $X = (x_1, X_2)$. We can thus decompose $\text{Var}(\hat{\beta})$ and use the inversion rule for block matrices. Let further denote $W = \Omega^{-1}$ the inverse of the covariance matrix. Then

$$\text{Var}(\hat{\beta}) = (X'WX)^{-1} = \begin{pmatrix} x_1'Wx_1 & x_1'WX_2 \\ X_2'Wx_1 & X_2'WX_2 \end{pmatrix}^{-1} \tag{4}$$

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \left(x_1'Wx_1 - x_1'WX_2(X_2'WX_2)^{-1}X_2'Wx_1\right)^{-1} \\ &= \left(x_1'\left(W - WX_2(X_2'WX_2)^{-1}X_2'W\right)x_1\right)^{-1} \\ &= \frac{1}{x_1'(W - Q)x_1} \end{aligned} \tag{5}$$

with $Q := WX_2(X_2'WX_2)^{-1}X_2'W$.

Now to delete the ith observation, the design matrix X is essentially multiplied with a modified identity matrix where the ith diagonal element is replaced by 0. This is done such a way that further equations are more easily transformed. To this end, we define $e_i := (0, \dots, 1, 0, \dots, 0)' \in \mathbb{R}^I$ an indicator vector of the ith observation and observe that $X_{[i]} := (I - \frac{1}{w_{ii}} e_i e_i' W)X$ is the same as the matrix X with the ith row replaced by $(0, \dots, 0)'$. Similarly, we use $\hat{\beta}_{[i]}$ to denote the estimator for β that results when the ith observation is omitted.

$$\mathrm{Var}(\hat{\beta}_{[i]})^{-1} = X_{[i]}' W X_{[i]} = X' \left(I - \frac{1}{w_{ii}} W e_i e_i' \right)' W \left(I - \frac{1}{w_{ii}} e_i e_i' W \right) X.$$

After reordering observations in X such that the ith observation is at the first position—and Ω accordingly—we can express $\mathrm{Var}(\hat{\beta}_{[i]})$ as

$$\mathrm{Var}(\hat{\beta}_{[i]})^{-1} = \begin{pmatrix} x_1' W x_1 - \frac{1}{w_{ii}} x_1' W e_i e_i' W x_1 & x_1' W X_2 - \frac{1}{w_{ii}} x_1' W e_i e_i' W X_2 \\ X_2' W x_1 - \frac{1}{w_{ii}} X_2' W e_i e_i' W x_1 & X_2' W X_2 - \frac{1}{w_{ii}} X_2' W e_i e_i' W X_2 \end{pmatrix} =: \begin{pmatrix} a & B \\ B' & D \end{pmatrix}. \tag{6}$$

The variance of the treatment effect estimator when omitting the ith observation can then be calculated using the inversion rule for block matrices as

$$\mathrm{Var}(\hat{\beta}_{1[i]}) = \frac{1}{a - BD^{-1}B'}. \tag{7}$$

Inserting the expressions in (6) into Equation (7) yields the following theorem. A proof can be found in the Appendix.

Theorem 3.1. *In a GLS model with design matrix $X = (x_1, X_2)$, covariance matrix $\Omega = W^{-1}$ and $Q := WX_2(X_2'WX_2)^{-1}X_2'W$, the inverse of the variance of the effect estimator of x_1 after omitting the ith observation is*

$$\begin{aligned} Var(\hat{\beta}_{1[i]})^{-1} &= x_1'(W - Q)x_1 - \frac{(e_i'(W - Q)x_1)^2}{e_i'(W - Q)e_i} \\ &= Var(\hat{\beta}_1)^{-1} - \frac{(e_i'(W - Q)x_1)^2}{e_i'(W - Q)e_i}. \end{aligned} \tag{8}$$

For the information content follows

$$\frac{1}{\mathrm{IC}(i)} = 1 - \frac{(e_i'(W - Q)x_1)^2}{x_1'(W - Q)x_1 e_i'(W - Q)e_i}. \tag{9}$$

A key advantage is that this expression holds for every general linear model. It is thus applicable to any set of fixed effects in X_2—especially any time effect parameterization. The covariance structure does not need to be block diagonal—in mixed model parlance, it is also applicable for crossed random effects. Furthermore, the expression in (9) essentially yields the information content for all observations at once, because the expression can be efficiently vectorized; in particular, it does not need the computation of different matrix inverses for each observation.

3.1.1 | Generalizations and applications of Theorem 3.1

By varying the columns in X_2, various cases can be handled. In the context of stepped wedge designs, X_2 reflects the specification of the time period effect. When the time effect is modeled as a factor variable, X_2 consists of stacked identity matrices of dimension $p - 1$. If there are reasons to assume that the period effect is smooth, for example, constant, linear, or a spline, X_2 can be more parsimonious. But one could also add more columns to X_2, for example, to allow for a separate but common learning effect.

3.2 | Cell contributions *(coefficient matrix)*

In a GLS setting, the estimator $\hat{\beta}$ is a linear function of the data y

$$\hat{\beta} = \underbrace{(XWX)^{-1}(X'W)}_{=:\mathrm{H}} \cdot y$$

with X the design matrix and W the inverse of the covariance matrix Ω as above. The matrix H gives an impression of the contributions of cluster-period cells with regard to the estimated coefficients $\hat{\beta}$. In the stepped wedge context, the first row of H corresponds to the coefficient of the treatment status, that is, the treatment effect. Since the GLS estimator is the best linear unbiased estimator, it is necessarily centrosymetric if the trial design is skew-symetric Bowden et al. (2021). This characteristic pattern can be observed in all figures in Section 4, as all discussed examples are skew-symetric.

To demonstrate the similarities to *information content*, it can be expressed in the same terms.

Theorem 3.2. *In a GLS model with design matrix* $X = (x_1, X_2)$*, covariance matrix* $\Omega = W^{-1}$ *and* $Q := WX_2(X_2'WX_2)^{-1}X_2'W$*, the* cell contribution *of the ith cell (syn.: observation) is*

$$h_i = \frac{x_1'(W-Q)e_i}{x_1'(W-Q)x_1}. \tag{10}$$

To show this, we use the decomposition of $(X'WX)^{-1}$ as in (4) and apply the inversion rule for block matrices. We are only interested in the first row of $(X'WX)^{-1}$, that is, we only need the two blocks in the upper row of the inverse. The first block of the inverse is $\mathrm{Var}(\hat{\beta}_1)$ that was already calculated in Equation (5), the second block in the first row of $\mathrm{Var}(\hat{\beta})$ is

$$\begin{aligned} B_2 &= -\left(x_1'Wx_1 - x_1'WX_2(X_2'WX_2)^{-1}(X_2'Wx_1)\right)^{-1} x_1'WX_2(X_2'WX_2)^{-1} \\ &= -\left(x_1'(W-Q)x_1\right)^{-1} x_1'WX_2(X_2'WX_2)^{-1}. \end{aligned}$$

Then, using the decomposition $X = (x_1, X_2)$, we get

$$h_i = \left(\mathrm{Var}(\hat{\beta}_1), B_2\right)X'We_i = \frac{x_1'We_i}{x_1'(W-Q)x_1} - \frac{x_1'Qe_i}{x_1'(W-Q)x_1}$$

and formula (10) follows.

3.3 | Relation between information content and cell contribution

With the results from Sections 3.1 and 3.2:

Lemma 3.3. *In a GLS model with design matrix* $X = (x_1, X_2)$*, covariance matrix* $\Omega = W^{-1}$ *and* $Q := WX_2(X_2'WX_2)^{-1}X_2'W$*, the inverse* information content *can be expressed using the* cell contribution h_i *of the ith observation as*

$$\frac{1}{IC(i)} = 1 - h_i^2 \frac{x_1'(W-Q)x_1}{e_i'(W-Q)e_i}.$$

This follows directly from (9) and (10). Notably, the inverse information content is a linear function of the squared cell contributions h_i^2 if and only if the diagonal of $W - Q$ is constant. The diagonal is constant for stepped wedge designs with equal cluster sizes $N_{kt} = N$ and a compound symmetry structure.

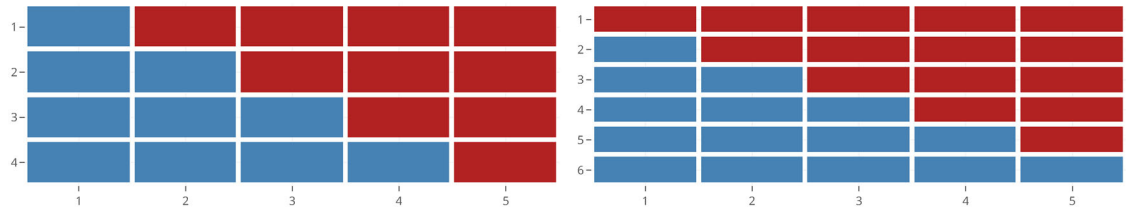

FIGURE 2 Allocation of cluster-period cells to control (blue) and interventional treatment (red) for two stepped wedge designs. The plot on the left visualizes a design with four clusters and five time periods, on the right is a hybrid design with six clusters, two of which do not switch

3.4 | Computation

The methods are implemented in the R-package `SteppedPower` that is available on CRAN. The package can compute power of GLS settings with user-friendly specifications for most of the common extensions to stepped wedge designs. It further implements the influence diagnostics covered in this paper. Besides, it offers a plot method that produced the figures shown in the next section.

4 | EXAMPLES

The influence of single cluster-period cells depends on the assumed correlation structure and the specification of temporal effects. We present two scenarios: the first compares correlation structures and illustrates the different strengths of the two influence diagnostics; the second compares different specifications of time effects. Plots produced by the `SteppedPower` package are shown for visualization. A demonstration can be found in the supporting material.

4.1 | Covariance specifications

We first consider a stepped wedge design of four sequences with one cluster each. In the second step, we will add two more clusters, which stay under control and interventional treatment, respectively. One can also think of the latter design as a combination of a stepped wedge design with four clusters and a parallel design consisting of two clusters. It is therefore referred to as hybrid design Girling and Hemming (2016). The allocation of cells to control and interventional treatment is illustrated in Figure 2. We assume a cross-sectional setting with 100 individuals in each cell. Both designs adjust for temporal effects with a factor variable.

Figure 3 shows the cell contributions for the design on the left in Figure 2 for different covariance structure assumptions. The upper row has a rather small intracluster correlation (ICC) of 0.01, the bottom row has a higher ICC of 0.1. The plots in the left column assume an exchangeable correlation structure, the ones in the right column an autoregressive correlation. A common pattern as observed in Matthews and Forbes (2017)—which might be counterintuitive from a user's perspective—becomes evident. Some cells under control—mainly those that are far from the main diagonal—contribute positively to the treatment effect estimator, whereas some cells under intervention contribute negatively. That is, a higher outcome in the first period of the fourth cluster leads to a higher estimated mean under intervention althought the cluster to receive the control treatment. This is an observation that is more clearly seen in the coefficient matrix than with the information content, because the coefficient matrix contains information about the sign of the contribution. The same pattern can be observed when we look at a hybrid design where we add two clusters that stay under control and interventional treatment, respectively, as shown in Figure 4. Perhaps, even more intriguing, the two clusters that do not switch have a very similar contribution pattern under high ICC and an exchangeable correlation structure (Figure 4, bottom left), even though the two clusters are in completely different treatment arms. This pattern can be seen only in the cell contribution image and is a consequence of the implicit adjustment for period effects.

Notably, the pattern is in accordance with the centrosymmetric structure of the information content matrix. Again, considering the coefficient matrix brings additional insight, because the information content conceals the fact that the first and the sixth clusters share a positive and negative sign in the first and fifth periods, respectively (see also Figure 5, bottom row).

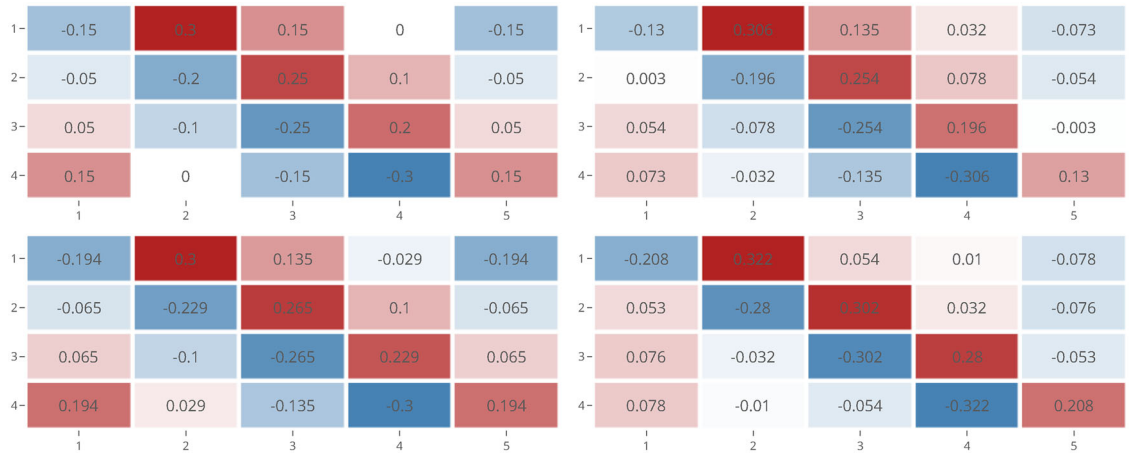

FIGURE 3 Cell contributions to the treatment effect estimator in a stepped wedge design of four clusters and five periods with 100 individuals in each cluster. Intracluster correlation 0.01 (top row) and 0.1 (bottom row), with an assumed exchangeable correlation structure (left) and a first-order autoregressive correlation structure with correlation parameter $r = 2/3$ (right). Positive cell contributions in red, negative contributions in blue

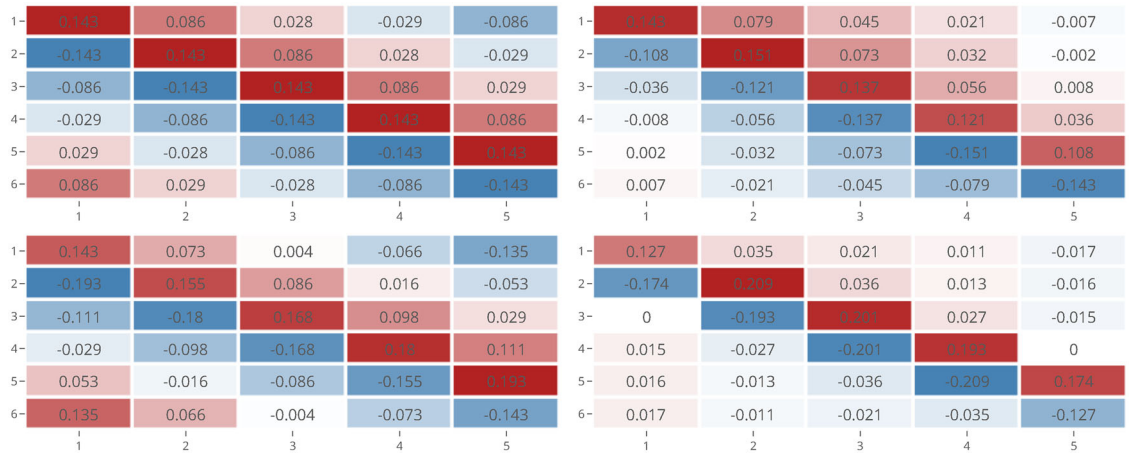

FIGURE 4 Cell contributions to the treatment effect estimator in a hybrid stepped wedge design of six clusters and five periods with 100 individuals in each cluster. Bottom cluster does not switch to interventional treatment, top cluster starts with interventional treatment. Intracluster correlation 0.01 (top row) and 0.1 (bottom row), with an assumed exchangeable correlation structure (left) and a first order autoregressive correlation structure with correlation parameter $r = 2/3$ (right). Positive cell contributions in red, negative contributions in blue

The combined influence of entire periods or clusters can also be shown with `SteppedPower`. For cell contribution, marginal plots in Figure 5 show the sum of absolute weights per row and per column, reflecting the contribution of entire clusters and periods, respectively. Note that each column can be seen as an unbiased treatment estimator that compares weighted means, that is, for each period, the sum of cell contributions under interventional treatment is the same as the sum of cell contributions under control treatment, as observed by Matthews and Forbes (2017). Note that this holds here because we adjusted for the time trend with a factor, so the model estimates a period-specific mean μ_t for each period t.

For the information content, the marginal plots show the relative change in estimator variance if entire periods or clusters are omitted. A notable observation is that the relative information content of entire clusters depends on the ICC: For high ICC, clusters that switch either early, late, or not at all have a lower information content than clusters that switch in the middle of the trial; for low ICC, all clusters have a very similar information content. Interestingly, the fact that all clusters are equally important for the effect estimator variance is not easily spotted with the coefficient matrix.

We also observe that for low ICC, the relative information content is approximately constant across time periods, whereas for high ICC, the relative information content is lowest in the middle period. This pattern was also observed

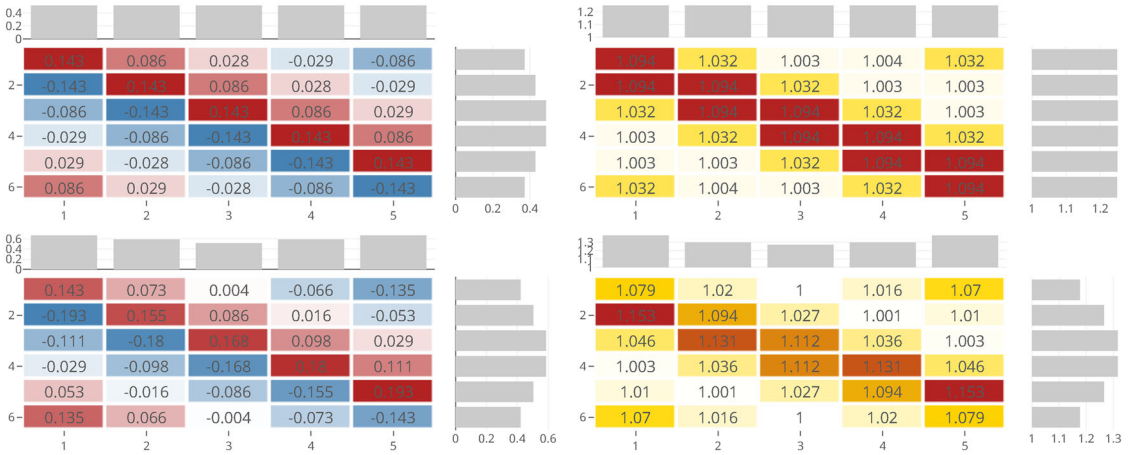

FIGURE 5 Cell contributions (left) and information content (right) in a hybrid stepped wedge design of six clusters and five periods with 100 individuals in each cluster. Bottom cluster does not switch to interventional treatment, top cluster starts with interventional treatment. Correlation structure is assumed to be exchangeable. Intracluster correlation is 0.01 (top row) and 0.1 (bottom row). Marginal plots (in gray) indicate the influence of entire clusters or periods, respectively.

in Kasza and Forbes (2019). When implementation periods are incorporated, this imbalance across clusters becomes even more pronounced (Kasza et al., 2019b, figure 4).

In the upper row of Figure 5, one can see that patterns of cluster contribution and cluster information content differ. This can occur because the former expresses influence on the point estimate, whereas the latter expresses influence on the variance of the treatment effect estimator. Also recall that with Lemma 3.3, the inverse information content is a linear function of the squared cell contribution in this setting. If we consider the sum of squared cell contributions in each cluster, we would get a similar pattern.

4.2 | Time effect specifications

We now consider the impact of time effect specifications on cell influences. Based loosely on the introductory example, we consider 10 clusters that are observed over 2 years where each quarter represents an observational period. This leads to a stepped wedge design with seven sequences, with the three sequences in the middle containing two clusters. The ICC is set to 0.075 that was also assumed in the planning of the Part-Child study.

We compare four models with different time specifications for this setting. The first model only contains a global mean, that is, we require $\mu_1 = \mu_t \forall t$ in Equation (1). In the second model, the time trend is linear, that is, $\mu_t = \mu_0 + C \cdot t$. Then, a seasonal time trend is assumed, that is, $\mu_t = \mu_{t+4}, t \in \{1, \dots, 4\}$. The fourth model contains the usual factor parameterization as in (1). Figure 6 shows the implications on cell influence. It becomes evident that influence patterns strongly depend on the type of time effect specification. Notably, columns add up to zero only if time is specified as a factor.

Grantham et al. (2020) showed that variances of the treatment effect estimator coincide for linear and factor time specifications if sequences are equally sized and all clusters are observed in every period, that is, for balanced complete stepped wedge designs. It can thus be expected that those two specifications produce similar influence patterns in this example—although not equal, because we have a slight imbalance across sequences.

5 | DISCUSSION

In this article, we have derived a new method for an a priori calculation of *information content* in GLS models. We have shown a connection of *information content* and *cell contributions*, the latter had previously been studied by Matthews and Forbes (2017) under the term *coefficient matrix*. We believe that the general formula provided in Theorem 3.1 may further the understanding of information content and its dependence on covariance structures.

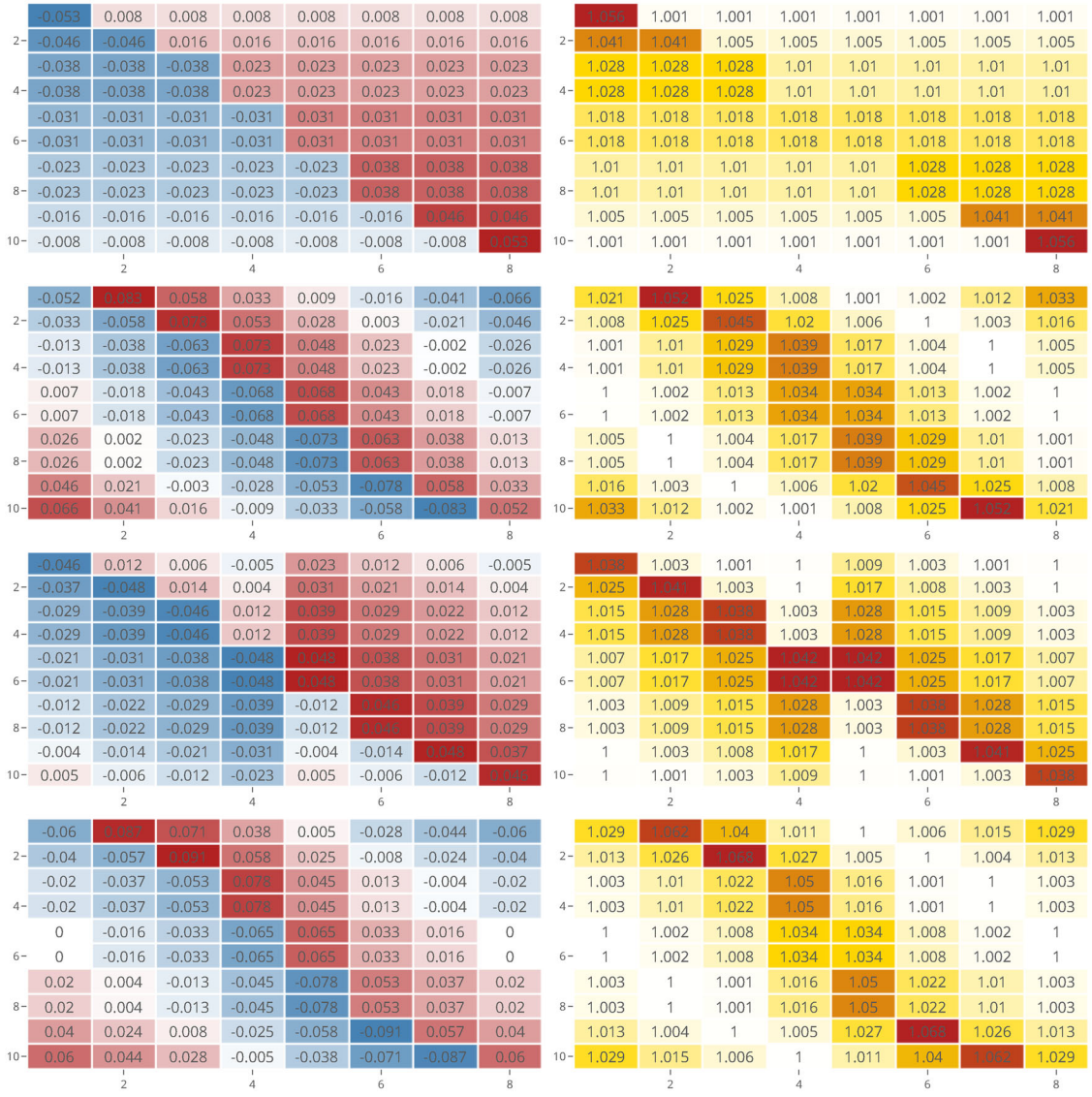

FIGURE 6 Cell contributions (left) and information content (right) in a stepped wedge design of 10 clusters and 8 periods with different time specifications. From top to bottom: none, linear, seasonal, factor

Cells without influence on the treatment effect estimator or on its variance can occur. Kasza et al. (2021) discuss the omission of such noninformative cells when time is modeled as a factor variable. It might still be beneficial to keep these cells in the study design, as they might contribute to the estimation of the covariance parameters. The underlying GLS model is assumed to know the correlation parameters. So, the estimator investigated here $\hat{\theta}_{GLS}$ will have different properties than an estimator from a mixed model $\hat{\theta}_{MM}$ or from generalized estimating equations $\hat{\theta}_{GEE}$. It is sensible to assume that the cells that do not influence $\hat{\theta}_{GLS}$ or $\text{Var}(\hat{\theta}_{GLS})$ have an influence on the estimated covariance $\hat{\Omega}$ and hence on $\hat{\theta}_{MM}$ and $\hat{\theta}_{GEE}$ or their variances. This aspect deserves further research.

The cell contributions patterns depend on the assumed correlation structure. Different specifications of the correlation can lead to differing point estimators. This underlines the importance to prespecify the correlation structure used for primary analysis for valid inference.

Generalizing this approach to nonnormal outcomes is possible, but with a few caveats. One can apply influence diagnostics on a linear predictor with the delta method. But for common nonnormal distributions such as binomial or Poisson, the variance is a function of the mean. Thus, the variances of cluster cells depend not only on the design assumptions, but also on the particular realizations. A priori influence diagnostics as described in this paper would thus only yield *estimated* influence. To further complicate the issue, the particular realization of a potential time trend also has an effect.

More recently, permutation-based methods for the analysis of SWCRT have been introduced; see Hughes et al. (2020) and Kennedy-Shaffer et al. (2020). Comparable cell influence metrics might be beneficial for a deeper understanding of the differences between (semi-)parametric and nonparametric analysis approaches.

The impact of unequal cluster sizes in the context of SWCRT has gained more attention in recent years, see Harrison et al. (2020), Martin et al. (2019), and Matthews (2020). Kasza et al. (2021) explore the implications of unequal cell sizes on the information content. During a trial, one can update the information content using ad hoc estimates for the covariance structure and parameters. It might be interesting to develop methods for dynamic over- or undersampling of upcoming cluster-period cells based on updated information content.

ACKNOWLEDGMENT

We thank Irene Schmidtmann (head of Biostatistics Division, IMBEI) for fruitful discussions. We thank Federico Marini for his support with the realization of the `R` package. We thank Freia DeBock, Michael Eichinger, and Michael Urschitz for sharing the PART-CHILD example. This research was conducted as part of the PhD thesis of Philipp Mildenberger at the University Medical Center Mainz.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Philipp Mildenberger https://orcid.org/0000-0002-7367-1708
Jochem König https://orcid.org/0000-0003-4683-0360

REFERENCES

Barker, D., McElduff, P., D'Este, C., & Campbell, M. (2016). Stepped wedge cluster randomised trials: A review of the statistical methodology used and available. *BMC Medical Research Methodology*, *16*(1), 1–19.

Bowden, R., Forbes, A. B., & Kasza, J. (2021). On the centrosymmetry of treatment effect estimators for stepped wedge and related cluster randomized trial designs. *Statistics & Probability Letters*, *172*, 109022.

Davey, C., Hargreaves, J., Thompson, J. A., Copas, A. J., Beard, E., Lewis, J. J., & Fielding, K. L. (2015). Analysis and reporting of stepped wedge randomised controlled trials: Synthesis and critical appraisal of published studies, 2010 to 2014. *Trials*, *16*(1), 1–13.

Girling, A. J., & Hemming, K. (2016). Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statistics in Medicine*, *35*(13), 2149–2166.

Grantham, K. L., Forbes, A. B., Heritier, S., & Kasza, J. (2020). Time parameterizations in cluster randomized trial planning. *The American Statistician*, *74*(2), 184–189.

Grantham, K. L., Kasza, J., Heritier, S., Hemming, K., & Forbes, A. B. (2019). Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment. *Statistics in Medicine*, *38*(11), 1918–1934.

Harrison, L. J., Chen, T., & Wang, R. (2020). Power calculation for cross-sectional stepped wedge cluster randomized trials with variable cluster sizes. *Biometrics*, *76*(3), 951–962.

Hemming, K., Taljaard, M., & Forbes, A. (2017). Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials*, *18*(1), 1–11.

Herrler, A., Görig, T., Georg, S., De Bock, F., Ullrich, C., & Eichinger, M. (2021). Assessment of shared decision making in pediatrics: Developing German scales for patients aged 7–18 years, parents and parent-proxy reports (collaboratepediatric). *Patient Education and Counseling*, *104*(3), 634–641.

Hughes, J. P., Granston, T. S., & Heagerty, P. J. (2015). Current issues in the design and analysis of stepped wedge trials. *Contemporary Clinical Trials*, *45*, 55–60.

Hughes, J. P., Heagerty, P. J., Xia, F., & Ren, Y. (2020). Robust inference for the stepped wedge design. *Biometrics*, *76*(1), 119–130.
Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*, *28*(2), 182–191.
Kasza, J., Bowden, R., & Forbes, A. B. (2021). Information content of stepped wedge designs with unequal cluster-period sizes in linear mixed models: Informing incomplete designs. *Statistics in Medicine*, *40*(7), 1736–1751.
Kasza, J., & Forbes, A. B. (2019). Information content of cluster-period cells in stepped wedge trials. *Biometrics*, *75*(1), 144–152.
Kasza, J., Hemming, K., Hooper, R., Matthews, J. N. S., & Forbes, A. B. (2019a). Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Statistical Methods in Medical Research*, *28*(3), 703–716.
Kasza, J., Taljaard, M., & Forbes, A. B. (2019b). Information content of stepped-wedge designs when treatment effect heterogeneity and/or implementation periods are present. *Statistics in Medicine*, *38*(23), 4686–4701.
Kennedy-Shaffer, L., De Gruttola, V., & Lipsitch, M. (2020). Novel methods for the analysis of stepped wedge cluster randomized trials. *Statistics in Medicine*, *39*(7), 815–844.
Li, F., Hughes, J. P., Hemming, K., Taljaard, M., Melnick, E. R., & Heagerty, P. J. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Statistical Methods in Medical Research*, *30*(2), 612–639.
Martin, J. T., Hemming, K., & Girling, A. (2019). The impact of varying cluster size in cross-sectional stepped-wedge cluster randomised trials. *BMC Medical Research Methodology*, *19*(1), 1–11.
Matthews, J. N. S. (2020). Highly efficient stepped wedge designs for clusters of unequal size. *Biometrics*, *76*(4), 1167–1176.
Matthews, J. N. S., & Forbes, A. B. (2017). Stepped wedge designs: insights from a design of experiments perspective. *Statistics in Medicine*, *36*(24), 3772–3790.
Nickless, A., Voysey, M., Geddes, J., Yu, L.-M., & Fanshawe, T. R. (2018). Mixed effects approach to the analysis of the stepped wedge cluster randomised trial—Investigating the confounding effect of time through simulation. *PLoS One*, *13*(12), e0208876.
Zhou, X., Liao, X., Kunz, L. M., Normand, S.-L. T., Wang, M., & Spiegelman, D. (2020). A maximum likelihood approach to power calculations for stepped wedge designs of binary outcomes. *Biostatistics*, *21*(1), 102–121.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mildenberger, P., & König, J. (2022). Influence of cluster-period cells in stepped wedge cluster randomized trials. *Biometrical Journal*, 1–13. https://doi.org/10.1002/bimj.202100383

APPENDIX A: PROOF OF 3.1

Given the block decomposition of $\mathrm{Var}(\hat{\beta}_{[i]})$ as in (6)

$$\mathrm{Var}(\hat{\beta}_{[i]})^{-1} = \begin{pmatrix} x_1'Wx_1 - \frac{1}{w_{ii}}x_1'We_ie_i'Wx_1 & x_1'WX_2 - \frac{1}{w_{ii}}x_1'We_ie_i'WX_2 \\ X_2'Wx_1 - \frac{1}{w_{ii}}X_2'We_ie_i'Wx_1 & X_2'WX_2 - \frac{1}{w_{ii}}X_2'We_ie_i'WX_2 \end{pmatrix} =: \begin{pmatrix} a & B \\ B' & D \end{pmatrix}$$

and the resulting equation (7)

$$\mathrm{Var}(\hat{\beta}_{1[i]})^{-1} = a - BD^{-1}B'$$

one can show Theorem 3.1. Since $(X_2'We_i)' = e_i'WX_2 \in \mathbb{R}^I$ are vectors, we calculate D^{-1} using the Sherman–Morrison formula:

$$D^{-1} = (X_2'WX_2)^{-1} + \frac{\frac{1}{w_{ii}}(X_2'WX_2)^{-1}X_2'We_ie_i'WX_2(X_2'WX_2)^{-1}}{1 - \frac{1}{w_{ii}}e_i'WX_2(X_2'WX_2)^{-1}X_2'We_i}.$$

We then combine this to BDB^{-1} and get

$$BD^{-1}B' = \underbrace{B(X_2'WX_2)^{-1}B'}_{\mathcal{P}_1} + \underbrace{B\frac{(X_2'WX_2)^{-1}X_2'e_ie_i'WX_2(X_2'WX_2)^{-1}}{w_{ii} - e_i'WX_2(X_2'WX_2)^{-1}X_2'We_i}B'}_{\mathcal{P}_2}.$$

Let $Q := WX_2(X_2'WX_2)^{-1}X_2'W$ as above.

$$\begin{aligned}\mathcal{P}_1 &= \left(x_1'WX_2 - \frac{1}{w_{ii}}x_1'We_ie_i'WX_2\right)(X_2'WX_2)^{-1}\left(X_2'Wx_1 - \frac{1}{w_{ii}}X_2'We_ie_i'Wx_1\right)\\ &= x_1'Qx_1 - \frac{2}{w_{ii}}x_1'We_ie_i'Qx_1 + \frac{1}{w_{ii}^2}(x_1'We_i)^2e_i'Qe_i,\\ \mathcal{P}_2 &= \frac{(e_i'Qx_1)^2}{w_{ii}-e_i'Qe_i} - \frac{2}{w_{ii}}\frac{x_1'We_ie_i'Qe_ie_i'Qx_1}{w_{ii}-e_i'Qe_i} + \frac{(e_i'Wx_1)^2}{w_{ii}^2}\frac{(e_i'Qe_i)^2}{w_{ii}-e_i'Qe_i}.\end{aligned}$$

For better readability, let us replace $e_i'We_i = w_{ii}$ and $e_i'Qe_i := q_{ii}$. Let further $\kappa := x_1'We_i = e_i'Wx_1$ and $\lambda := x_1'Qe_i = e_i'Qx_1$ because W, Q are symmetric matrices. Note that $w_{ii}, q_{ii}, \kappa, \lambda$ are all scalars.

$$\begin{aligned}\mathcal{P}_1 &= x_i'Qx_1 - \frac{2\kappa\lambda}{w_{ii}} + \frac{q_{ii}\kappa^2}{w_{ii}^2}\\ &= x_i'Qx_1 + \frac{1}{w_{ii}-q_{ii}}\left(-2\kappa\lambda + \frac{2q_{ii}\kappa\lambda}{w_{ii}} + \frac{q_{ii}\kappa^2}{w_{ii}} - \frac{q_{ii}^2\kappa^2}{w_{ii}^2}\right),\\ \mathcal{P}_2 &= \frac{1}{w_{ii}-q_{ii}}\left(\lambda^2 - \frac{q_{ii}\kappa\lambda}{w_{ii}} - \frac{q_{ii}\kappa\lambda}{w_{ii}} + \frac{q_{ii}\kappa^2}{w_{ii}^2}\right),\\ \mathcal{P}_1 + \mathcal{P}_2 &= x_i'Qx_1 + \frac{1}{w_{ii}-q_{ii}}\left(-2\kappa\lambda + \frac{q_{ii}\kappa^2}{w_{ii}} + \lambda^2\right).\end{aligned}$$

Inserting this expression into (7) yields

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_{1[i]})^{-1} &= a - (\mathcal{P}_1 + \mathcal{P}_2) = x_1'Wx_1 - \frac{\kappa^2}{w_{ii}} - (\mathcal{P}_1 + \mathcal{P}_2)\\ &= x_1'(W-Q)x_1 - \frac{1}{w_{ii}-q_{ii}}\left(\kappa^2 - \frac{q_{ii}\kappa^2}{w_{ii}} - 2\kappa\lambda + \frac{q_{ii}\kappa^2}{w_{ii}} + \lambda^2\right)\\ &= \mathrm{Var}(\hat{\beta}_1)^{-1} - \frac{(e_i'(W-Q)x_1)^2}{e_i'(W-Q)e_i},\end{aligned}$$

which proves the theorem.