RESEARCH ARTICLE

Theory of General Balance Applied to Step Wedge Designs

Simon Bond*[1,2]

[1]Cambridge Clinical Trials Unit, Cambridge University Hospitals NHS Foundation Trust , Cambridgeshire, United Kingdom
[2]Medical Research Council Biostatistics Unit, University of Cambridge, Cambridgeshire, United Kingdom

Correspondence
*Simon Bond, CCTU, Box 401
Addenbrooke's Hospital, Hills Road,
Cambridge CB2 0QQ, UK Email:
simon.bond@addenbrookes.nhs.uk

Summary

A standard idealised step-wedge design satisfies the requirements, in terms of the structure of the observation units, to be considered a *balanced design* and can be labelled as a criss-cross design (time crossed with cluster) with replication. As such Nelder's theory of general balance can be used to decompose the analysis of variance into independent strata (grand mean, cluster, time, cluster:time, residuals). If time is considered as a fixed effect then the treatment effect of interest is estimated solely within the cluster and time:cluster strata; the time effects are estimated solely within the time stratum. This separation leads directly to scalar, rather than matrix, algebraic manipulations to provide closed-form expressions for standard errors of the treatment effect estimate. We use the tools provided by the theory of general balance to obtain an expression for the standard error of the estimated treatment effect in a general case where the assumed covariance structure includes random-effects at the time and time:cluster levels. This provides insights that are helpful for experimental design regarding the assumed correlation within clusters over time, sample size in terms of numbers of clusters and replication within cluster, and components of the standard error for estimated treatment effect.

KEYWORDS:
cluster randomized clinical trials, linear mixed models, step-wedge design

1 | BACKGROUND AND MOTIVATION

The theory of general balance is presented within two papers by Nelder[1,2] and is expounded upon, in particular using Hasse diagrams to perform calculations, in Bailey.[3] The theory applies to experimental designs that meet the condition that the structure of observation units be represented by an arbitrary degree of nesting and crossing of *blocks*, but that the level of replication within any level is constant. As such an idealized step-wedge (SW) design is an example of time crossed with cluster.

The first of the two papers, focused on the structure of observational units, to the exclusion of treatment structure. It provides a calculus to decompose the implied covariance structure into orthogonal *strata*, with three different parametrisations that correspond to expected sums of squares, variance parameters of random effects, interpretable correlation coefficients; along with corresponding degrees of freedom. The use of Hasse diagrams adds some clarity to this calculus,[3] although stops short of providing all three versions of the parametrisation.

The second of the papers,[2] considers how to proceed with estimation of fixed treatment effects. Each fixed effect to be estimated may have information contained within any of the strata. Estimates from different strata have the useful property of mutual independence, conditional on the true values, and the covariance decomposition immediately provides expressions for the variance of each estimate within a stratum.

In the canonical SW setting, aside from the universal level, sometimes referred to as the grand mean, there are four possible strata levels: cluster;time; cluster:time; and residual error. The adjustment for time as a fixed, rather than random, effect is the default as assumed in Hussey & Hughes[4] and has the simplification that estimation of time effects is contained entirely within the time stratum. Hence the multi-dimensional information matrix that would need to be algebraically inverted in a modern analysis (see Pinheiro & Bates[5]) of a linear mixed model with fixed effects for intercept, time and treatment, is conveniently avoided. When the information matrix is decomposed by strata, in the cluster and cluster-time strata there are zeroes for all elements aside from diagonal element corresponding to the treatment effect. Hence the generalised inverse of the information matrix, which is needed to express the standard error of the strata-specific treatment effect estimate, is trivial to express in terms of a scalar inverse.

This paper is focused on applying the methods of general balance in the setting of a SW designs. The Hussey & Hughes[4] assumptions will be used predominantly as an exemplar and a subsequent generalisation provided to show how the methods can be used in an alternative setting. There are a numerous variations on these assumptions: fixed or random effects at any stratum; taking repeated observations within the same participant, rather than new participants within the same cluster, which adds in a new cluster-participant stratum. The methods demonstrated within this paper can be easily applied by the reader across different combinations of assumptions. At the study design stage a statistician typically has assumed numerical values for standard deviation, treatment effect size of interest, and within-cluster correlation values. The result of applying methods from this paper is to provide a clear exposition in terms of simple closed-form expressions for the standard error of treatment effect estimates at each stratum and overall and thus the statistical power. In contrast to the alternative of numerical calculations we gain an understanding of the implications of the assumptions, uncertainty around parameter values, and the interplay between the number of clusters, participants, and replications.

2 | NOTATION

We assume the same notation as Hussey & Hughes[4] and will resolve this with Nelder,[1,2] providing explanation as needed.

Let Y_{ijk} be the response for individual k at time j from cluster i (i in $1, \ldots, I$; j in $1, \ldots, J$; k in $1, \ldots, N$). We assume that

$$Y_{ijk} = \mu_{ij} + e_{ijk},$$
$$\mu_{ij} = a_i + \beta_j + x_{ij}\theta,$$
$$a_i \sim N(0, \tau^2),$$
$$e_{ijk} \sim N(0, \sigma_e^2).$$

The random elements a_i and e_{ijk} are assumed to be mutually independent. We also define $\sigma^2 = \sigma_e^2 + \tau^2$, the overall variation (Hussey & Hughes[4] attach a different definition to σ^2 with no indices).

The correlation within clusters is captured by the random effects, a_i; the fixed time effects are β_j; the treatment indicator x_{ij} is constant within cluster-time. The SW design imposes a condition on this covariate x_{ij} to ensure that only uni-directional switching happens, so $x_{ij} = 1 \implies x_{ij'} = 1$ for all $j' > j$, however the results presented here apply to any choice of treatment mapping; the treatment effect size is θ. As is common practice in study design, we assume that the values of the covariance parameters τ^2, σ_e^2, are known.

3 | EXPERIMENTAL UNITS

We can classify the experimental units as criss-cross (of time and cluster) with split plot (replications).[1] Under the assumptions of balance this is condensed into: $(I \times J) \to N$. The equivalent Hasse diagram is given in figure 1 .

Expressing the degrees of freedom identity under the calculus is the first step, which in this case gives terms for intercept, cluster, time, time-cluster, residual respectively:

$$IJN = 1 + (I - 1) + (J - 1) + (I - 1)(J - 1) + IJ(N - 1). \tag{1}$$

The individual terms can be obtained by working down the Hasse diagram where $d_{universal} = 1$ and $d_F = n_F - \sum_{G > F} d_G$, where n_F is the number of levels within stratum F. The notation $G > F$ is defined precisely in Bailey[3] but simply means strata G that are above stratum F.

Averaging matrices are defined to transform the vector Y_{ijk} to a vector of averages repeated over the index that has been dropped, for example $Y_{ijk} \mapsto \overline{Y}_{i,\cdot,\cdot} = \sum_{j,k} Y_{ijk}/JN$, the ijkth component of $\mathbf{B}_{cluster}\mathbf{Y}$, where

$$\mathbf{B}_{cluster} = (1/JN)\,\mathbf{I}_{cluster} \otimes \mathbf{J}_{time} \otimes \mathbf{J}_{individual}.$$

Here bold font is used to indicate a matrix where $\mathbf{J}$ is a square matrix of 1s and $\mathbf{I}$ the identity matrix, both with dimensions equal to the number of levels within the stratum given in the suffix. The Kronecker product is indicated by $\otimes$. Generalisation of this example to other combinations of indices is trivial.

Next, we map each term in the degrees of freedom equation (1) to a set of orthogonal matrices that are used to calculate the sums of squares within each stratum: $\mathbf{C}_{universal}, \mathbf{C}_{cluster}, \mathbf{C}_{time}, \mathbf{C}_{cluster:time}, \mathbf{C}_{individual}$. For example,

$$\mathbf{C}_{cluster:time} = \mathbf{B}_{cluster:time} - \mathbf{B}_{cluster} - \mathbf{B}_{time} + \mathbf{B}_{universal},$$

is obtained by

- starting with the term $(I-1)(J-1)$, the fourth term on the right hand side of equation (1)

- replacing I with *cluster*, J with *time*, and 1 with *universal*.

- symbolically expanding to obtain the expression *cluster:time - cluster - time + universal*

- mapping each term to the index of an averaging matrix and preserving the coefficients.

Alternatively we work down the Hasse diagram apply the equation $\mathbf{C}_F = \mathbf{B}_F - \sum_{G>F} \mathbf{C}_G$. It is shown that these form a set of orthogonal matrices [1] and that the sums of squares are thus independent and defined as $\mathbf{Y}'\mathbf{C}_F\mathbf{Y}$.

4 | COVARIANCE DECOMPOSITION

The covariance matrix of the $\mathbf{Y}$ can be written in three different parametrisations

$$\mathbf{V} = \sigma^2 \sum_F \rho_F \mathbf{W}_F = \sigma^2 \sum_F f_F \mathbf{U}_F = \sigma^2 \sum_F \zeta_F \mathbf{C}_F, \qquad (2)$$

where the F is used to index the strata.

The first of these parametrisations is the easiest to conceptualise, and propose suitable values for in study design stages; ρ_F represents the correlation between two observations that are within the same level of F, but not common to any strata below F.

Following Hussey & Hughes,[4] we are assuming that

- $\rho_{universal} = 0$ the universal or grand mean,

- $\rho_{cluster} = \tau^2/\sigma^2$, cluster level,

- $\rho_{time} = 0$, as per the assumption of fixed time effects,

- $\rho_{cluster:time} = \rho_{cluster}$, patients within the same cluster-period are no more correlated than those in the same cluster, but different periods,

- $\rho_{residual} = 1$, a mathematical formality that a random variable is perfectly correlated with itself.

The third assumption may be less plausible in some contexts.

The second parametrisation in (2) corresponds to expressing the model as a mixed effect model, as in section 2 where the f_F is the assumed variance parameter (or zero in its absence) of an independent random intercept scaled by the overall variance, at the corresponding stratum level.

The third parametrisation in (2) is the expected mean sum-of-squares under the null hypothesis and provides the covariance matrix for estimates obtained in each stratum. The core of this paper lies in showing how to derive expressions for ζ_F. These values are of importance as they readily lead to expressions for the stratum-specific and overall information, and standard error of the treatment effects (4), which in turn can be used in standard formulae for power and significance level.

Nelder[1] derives these relationship, gives precise definitions to $\mathbf{W}_F, \mathbf{U}_F$ and a calculus to obtain f_F and ζ_F from ρ_F. Note that we have scaled the ζ by σ^2. The detailed steps of the calculations and results are given in table 1 . The equivalent obtained by working up and down the Hasse diagram using the rules descried in Bailey[3] and stated below is in figure 1 .

- $d_F = n_F - \sum_{G>F} d_G$

- $f_F = \rho_F - \sum_{G>F} f_G$

- $\zeta_F = \sum_{G \leq F} n_{residual}/n_G f_G$

If observations at distinct time points are taken on different participants then the assumption of a constant correlation within a cluster appears plausible as a consequence of time-invariant characteristics unique to each cluster. A variant of the SW design repeatedly observes participants longitudinally over time. Including an additional stratum for each patient, nested below the cluster level and above the residual level, may account for this. The methods described herein allow the reader to easily derive equivalent formulae. However an extra stratum makes no provision to potentially allow the correlation to decrease for within-participant observations that are further apart in time. Varying correlation falls outside the scope of this paper.
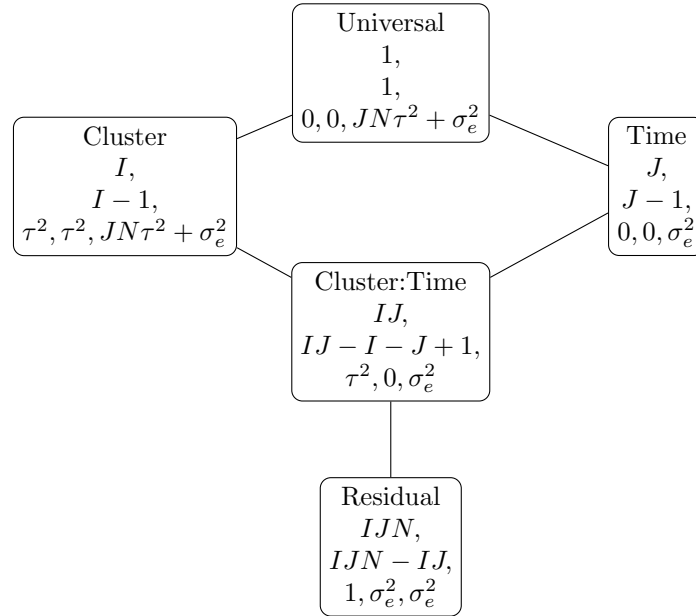

FIGURE 1 Hasse Diagram: Number of levels, degrees of freedom, $\rho\sigma^2$, $f\sigma^2$, $\zeta\sigma^2$

5 | ESTIMATION

The second of Nelder's papers[2] covers how to estimate treatment effects and provide standard errors, moving away from the null hypothesis.

It assumes that there is a general design matrix $\mathbf{X}$, with IJN rows and columns for all fixed effects including intercept, treatment effect, and time effects. We define the vector of coefficients $\boldsymbol{\beta}$ to satisfy $E\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$. The response vector is projected within each stratum's subspace, and a stratum-specific estimate is provided

$$\widehat{\boldsymbol{\beta}}_F = \left(\mathbf{X}'\mathbf{C}_F\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{C}_F\mathbf{Y} \tag{3}$$

$$\operatorname{Var}\widehat{\boldsymbol{\beta}}_F = \sigma^2\zeta_F\left(\mathbf{X}'\mathbf{C}_F\mathbf{X}\right)^{-1} \tag{4}$$

Define the basis $X_{time} = \{\mathbf{e}_1, \ldots, \mathbf{e}_J\}$, where $\mathbf{e}_j$ is a vector length IJN, where the rows corresponding to time j are 1, and 0 otherwise. Thus X_{time} can be used as the columns of the design matrix for time. Consider the mapping $\mathbf{B}_F\mathbf{e}_j$. Each cluster level average of $\mathbf{e}_j$ is $1/J$, which also equates to the universal average of $\mathbf{e}_j$. The cluster-time, and time level the average is either 1 or 0, indicating if the time point at which to average equals the j suffix in $\mathbf{e}_j$. The projection matrices $\mathbf{C}_F$ are differences of averages (section 3), which cancel out for cluster and cluster-time and thus the columns for times, in $\mathbf{X}$ are orthogonal to $\mathbf{C}_{cluster}$

and $\mathbf{C}_{cluster:time}$. Hence the time fixed effects are entirely contained within the remaining time stratum. Similarly the intercept is entirely contained in the Universal stratum level.

The matrix inversion denoted in (3) and (4) will thus involve singular matrices with rows and columns of zeroes corresponding to the columns of $\mathbf{X}$ that are orthogonal to the $\mathbf{C}_F$. So the stratum-specific equations (3) and (4) are modified to drop such columns from $\mathbf{X}$. The equations now reduce to scalar equations in terms of θ, the treatment effect, for the cluster and cluster-time strata, which greatly simplifies the algebraic inversion. We can replace the general design matrix $\mathbf{X}$, with the column vector $\mathbf{x} = \{x_{ijk}\}$, as defined in section 2 where the right-most index varies the fastest. Note the assumption that $x_{ijk} = x_{ij}$ is constant for all k, which is true for a SW design.

We must obtain expressions for the variances of the treatment effect estimates in each stratum using the specific case of the $\mathbf{x}$ associated with the specific SW design and the specific values of $\mathbf{C}_{cluster}, \mathbf{C}_{cluster:time}$. The explicit calculation of individual terms $\mathbf{x}'\mathbf{B}_F\mathbf{x}$ is trivial to evaluate by summing x_{ijk} over the indexes that correspond to a $\mathbf{J}$ in the Kronecker matrix products that define $\mathbf{B}_F$ (section 3). We also note that $x_{ij} = x_{ij}^2$ as $x_{ij} \in \{0, 1\}$.

$$\mathbf{x}'\mathbf{B}_{universal}\mathbf{x} = 1/(IJN)\mathbf{x}'(\mathbf{J}_{cluster} \otimes \mathbf{J}_{time} \otimes \mathbf{J}_{individual})\mathbf{x} = 1/(IJN) \sum_{ijk} x_{ijk} \sum_{i'j'k'} x_{i'j'k'}$$

$$= N/(IJ) \left\{ \sum_{ij} x_{ik} \right\}^2$$

$$\mathbf{x}'\mathbf{B}_{cluster}\mathbf{x} = 1/(JN)\mathbf{x}'(\mathbf{I}_{cluster} \otimes \mathbf{J}_{time} \otimes \mathbf{J}_{individual})\mathbf{x} = 1/(JN) \sum_{ijk} x_{ijk} \sum_{j'k'} x_{ij'k'}$$

$$= N/J \sum_i \left\{ \sum_j x_{ij} \right\}^2$$

$$\mathbf{x}'\mathbf{B}_{time}\mathbf{x} = 1/(IN)\mathbf{x}'(\mathbf{J}_{cluster} \otimes \mathbf{I}_{time} \otimes \mathbf{J}_{individual})\mathbf{x} = 1/(IN) \sum_{ijk} x_{ijk} \sum_{i'k'} x_{i'jk'}$$

$$= N/I \sum_j \left\{ \sum_i x_{ij} \right\}^2$$

$$\mathbf{x}'\mathbf{B}_{cluster:time}\mathbf{x} = 1/N\mathbf{x}'(\mathbf{I}_{cluster} \otimes \mathbf{I}_{time} \otimes \mathbf{J}_{individual})\mathbf{x} = 1/N \sum_{ijk} x_{ijk} \sum_{k'} x_{ijk'}$$

$$= N \sum_{ij} x_{ij}^2 = N \sum_{ij} x_{ij}$$

Using the notation of Hussey & Hughes,[4] $U = \sum_{ij} x_{ij}, V = \sum_i (\sum_j x_{ij})^2, W = \sum_j (\sum_i x_{ij})^2$, together with the expression for ζ in table 1 we obtain the information values from the cluster, and cluster:time strata shown in table 2

Having obtained multiple estimates of the treatment effect at the different strata the analysis is not yet concluded; the statistician must use these to report a *single* treatment effect estimate and associated standard error that is optimal in some way. Given that the estimates from the two strata are known to be independently distributed, the minimum variance linear combination is shown in Scheffe[6] to be the overall least squares estimate, assuming that the covariance parameters are known, and thus is also the maximum likelihood estimate under normality assumptions. The optimal linear combination takes weights proportional to the information values, formally the inverse variance, for each individual estimate and the resulting estimate has information equal to the sum of the individual information values. Simple algebraic manipulation verifies the formula given, but not derived, in Hussey & Hughes,[4] as per table 2 .

At the analysis stage we *could* continue in the same manner of projecting the response vector $\mathbf{Y}$ and the design matrix $\mathbf{X}$ by strata, using the $\mathbf{C}$ projections, and then fitting a standard linear regression model within each stratum as per Chambers et al.[7] If working by-hand, one would calculate the projections by iteratively working down the Hasse diagram taking the average for each level of the stratum and subtracting off the projection from the matching level of any parent strata. The residual mean squares at the strata provide estimates of $\sigma^2 \zeta$ that, using the formula in table 3 for the general case, could be inverted to provide estimates of the random effects variance parameters. In the case of the Hussey & Hughes assumptions, there are multiple estimates of σ_e^2 at

the residual, cluster-time, and time strata, which need to be combined as a sum with weights proportional to the residual degrees of freedom at each stratum. A further complication is that these residual mean squares are calculated conditional on the fixed effect parameters estimated *within* each stratum. Ultimately, we want to replace the individual-stratum fixed effect estimates with pooled values, which are obtained using weights derived from covariance parameters. Hence an iterative calculation to achieve convergence would be required. In *practice* at the analysis stage we would apply modern software[5] using restricted maximum likelihood estimation (REML) to estimate *both* the covariance parameters and fixed effects, using optimal weighting of observations to account for any deviations from the assumption of exact balance for any missing clusters/times or unequal replication. Supplementary material in the form of R code is provided to illustrate the calculations.

TABLE 1 Hussey & Hughes Covariance Parametrisations

Stratum	ρ	f	$\sigma^2 \zeta$
Overall Mean	0	$\rho = 0$	$\sigma^2(IJNf + INf_{time} + JNf_{cluster} + Nf_{cluster:time} + f_{individual}) = JN\tau^2 + \sigma_e^2$
Cluster	τ^2/σ^2	$\rho_{cluster} - \rho = \tau^2/\sigma^2$	$\sigma^2(JNf_{cluster} + Nf_{cluster:time} + f_{individual}) = JN\tau^2 + \sigma_e^2$
Time	0	$\rho_{time} - \rho = 0$	$\sigma^2(INf_{time} + Nf_{cluster:time} + f_{individual}) = \sigma_e^2$
Cluster:Time	τ^2/σ^2	$\rho_{cluster:time} - \rho_{cluster} - \rho_{time} + \rho = 0$	$\sigma^2(Nf_{cluster:time} + f_{individual}) = \sigma_e^2$
residual	1	$\rho_{individual} - \rho_{cluster:time} = \sigma_e^2/\sigma^2$	$\sigma^2 f_{individual} = \sigma_e^2$

TABLE 2 Information Levels by Stratum

Stratum	Hussey & Hughes	General Case
Cluster	$(NV/J - NU^2/IJ)/(JN\tau^2 + \sigma_e^2)$	$(NV/J - NU^2/IJ)/(JN\tau_{cluster}^2 + N\Delta^2 + \sigma_e^2)$
Time	No Information	$(NW/I - NU^2/IJ)/(IN\tau_{time}^2 + N\Delta^2 + \sigma_e^2)$
Cluster:Time	$(NU - NV/J - NW/I + NU^2/IJ)/\sigma_e^2$	$(NU - NV/J - NW/I + NU^2/IJ)/(N\Delta^2 + \sigma_e^2)$

6 | GENERALISATION

We can relax the assumption of Hussey & Hughes [4] in terms of two additional parameters. We can allow random, rather than fixed, time effects, which requires an extra correlation parameter, ρ_{time}, to quantify the degree of correlation between two participants observed in the same time period but from distinct clusters. There may be further, super-additive, correlation between participants in the same cluster-period, where $\rho_{cluster:time} = \Delta/\sigma^2 + \rho_{cluster} + \rho_{time}$, and we used Δ to represent the increment in correlation. We modify table 1 to achieve table 3. The equivalent Hasse diagram is figure 2. The corresponding information values from all three strata are given in table 2.
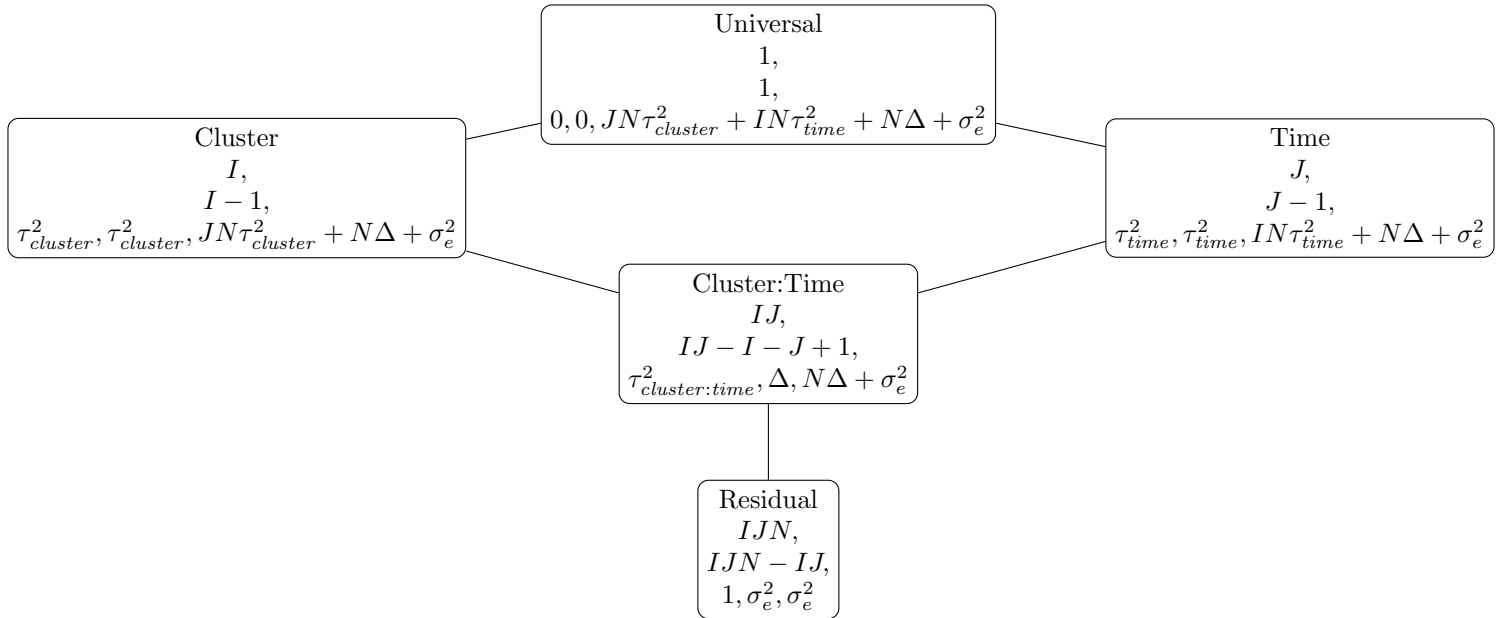

FIGURE 2 Hasse Diagram Generalised Structure

7 | EFFICIENCY FACTOR

The evaluation of U, V, and W provide expressions for the variance of the treatment effect estimate for all values of the design matrix **X**. The most efficient design, with smallest estimated treatment effect variance, is when individual-level randomisation is used, blocked at the cluster-time level. Assuming equal and balanced randomisation within each cluster-time block, it is easily shown that $U = IJ/2, V = IJ^2/2, W = JI^2/2$. Hence from table 2, under either set of assumptions, the cluster and time strata provide exactly zero information, and the numerator in the cluster-time stratum is $NIJ/4$, thus leading to a well-known expression for the variance under individual-level randomisation.

For any other design the *efficiency factor* is the ratio of the estimated treatment effect variance to the variance from the individual-level randomisation design, assuming the same number of observations overall. Sample sizes may be calculated by initially assuming individual-level randomisation and then scaling up the resulting sample size by the efficiency factor for the actual design choice. The efficiency factor thus gives a simple summary to enable a fair comparison between design choices.

Beyond making detailed calculations for U, V, W in specific cases, it is helpful to provide a rule-of-thumb for the efficiency factor of the SW design. We consider an abstraction that approximates finite sums with continuous integrals (for example $\sum_{x=0}^{n} f(x) \approx \int_0^n f(x)dx$), and for simplification we assume $I = J$ and that the summand/integrand is the indicator function that $i \leq j$. This leads to $U = J^2/2, V = W = J^3/3$ and the numerators in table 2 are all equal to $NJ^2/12$. Hence the efficiency

factor for the Hussey & Hughes design is approximated by

$$\frac{2 + (JN - 2)\rho_{cluster}}{3\left\{1 + (JN - 1)\rho_{cluster}\right\}},$$

which goes from $2/3$ to $1/3$ as $\rho_{cluster}$ goes from 0 to 1.

8 | DISCUSSION

In the set of assumptions used initially in Hussey & Hughes[4] the information within the time-cluster stratum increased linearly with N. Increasing the sample size overall will increase the power up to a maximum of 100%. The cluster term has a the $JN\tau^2$ in the denominator and so increasing N will only increase the information at the cluster level to an asymptote. Mathematically it can be seen that in the SW design $J = O(I), U = O(I^2), V = O(I^3), W = O(I^3)$, and so the information at the cluster stratum is $O(I)$.

Moving to the general parametrisation, unless $\Delta = 0$ all the information expression across the strata are $O(I)$. So with a fixed number of clusters, the power can only increase to an asymptote strictly less than 100% as the overall sample size is increased. This limitation is common to standard cluster-randomised designs.

The assumption $\Delta = 0$ may be plausible if the setting is a treatment implemented by training a ward in a hospital thus any correlation is due to a common set of staff within a ward, rather than any temporal preservation of correlation. In an educational setting, where the clusters are defined by a teacher using a new teaching method and the time periods are year cohorts of pupils in the teacher's class, $\Delta = 0$ is a more questionable assumption: pupils within the same year-class will interact and influence each other more than pupils from different years who share the same teacher. The parametrisation considered above adds on a point mass of extra correlation for observations in the same cluster and period. An alternative would be a parametric correlation structure that allows the increment $\Delta(j_1, j_2)$ to be a function of the proximity of the periods within the same cluster $|j_1 - j_2|$, an auto-regressive structure for example. However, that is beyond the scope of this article.

If it is assumed that $\tau^2_{time} = 0$ then we revert to fixed time effects, whilst still assuming $\Delta \neq 0$. In this case the time stratum provides no information on the treatment effect. The random effects are a stronger assumption, which does rely on the parametric assumption of normality, in contrast to the non-parametric assumption of the fixed effects; the extra strength of the assumption does provide extra information on the treatment effects if it holds. In either case the smaller the value of $\tau^2_{cluster}$ or τ^2_{time}, the more similar the observations between clusters or periods, compared to within, the stronger the information provided at the respective strata.

In recent research[8] the focus is on the coefficients of the linear combination of Y used in estimation of the treatment effect. They demonstrate that the weighted sum of the individual stratum estimates that form the maximum likelihood estimate can also be represented as a weighted sum from two simpler models. The first component, termed as *vertical* model is a fixed effect model that ignores the cluster effect, only adjusting for time; in which case there is an equivalent Hasse diagram, and the resulting orthogonal matrix is $\mathbf{B}_{cluster:time} - \mathbf{B}_{time}$. The second component comes from a two-way fixed effect model adjusting for time and cluster; the resulting orthogonal matrix coincides with $\mathbf{C}_{cluster:time}$ derived in section 3. It is straightforward to show that any linear combination of these two matrices can be expressed as an linear combination of the $\mathbf{C}_{cluster:time}$ and $\mathbf{C}_{cluster}$ as used in this paper. The two routes to deriving and interpreting the components of information that constitute the maximum likelihood estimate are thus complimentary.

References

1. Nelder J. The Analysis of Randomized Experiments with Orthogonal Block Structure. I. Block Structure and the Null Analysis of Variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.* 1965;283(1393):147–162.

2. Nelder J. The Analysis of Randomized Experiments with Orthogonal Block Structure. II. Treatment Structure and the General Analysis of Variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.* 1965;283(1393):163–178.

3. Bailey RA. *Design of Comparative Experiments.* Cambridge University Press; 2008.

4. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemporary clinical trials.* 2007;28:182–191.

5. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. Springer; 2000. chapter 2.

6. Scheffé H.. *The Analysis of Variance*. Wiley; 1959. pages 176-178.

7. Chambers JM, Freeny A, Heiberger RM. Analysis of Variance; Designed Experiments. In: Chambers JM, TJ Hastie, eds. *Statistical Models in S*, Chapman & Hall / CRC 1992. chapter 5.

8. Matthews JNS, Forbes AB. Stepped wedge designs: insights from a design of experiments perspective. *Statistics in Medicine.* 2017;36(24):3772–3790. sim.7403.

TABLE 3 General Covariance Parametrisations

Stratum	ρ	f	$\sigma^2\zeta$
Overall Mean	0	$\rho = 0$	$\sigma^2(IJNf + INf_{time} + JNf_{cluster} + JNf_{cluster} + Nf_{cluster:time} + f_{individual})$
			$= JN\tau^2_{cluster} + IN\tau^2_{time} + N\Delta + \sigma^2_e$
Time	τ^2_{time}/σ^2	$\rho_{time} - \rho = \tau^2_{time}/\sigma^2$	$\sigma^2(INf_{time} + Nf_{cluster:time} + f_{individual}) = IN\tau^2_{time} + N\Delta + \sigma^2_e$
Cluster	$\tau^2_{cluster}/\sigma^2$	$\rho_{cluster} - \rho = \tau^2/\sigma^2$	$\sigma^2(JNf_{cluster} + Nf_{cluster:time} + f_{individual}) = JN\tau^2_{cluster} + N\Delta + \sigma^2_e$
Cluster:Time	$\tau^2_{cluster:time}/\sigma^2$	$\rho_{cluster:time} - \rho_{cluster} - \rho_{time} + \rho = \Delta/\sigma^2$	$\sigma^2(Nf_{cluster:time} + f_{individual}) = N\Delta + \sigma^2_e$
residual	1	$\rho_{individual} - \rho_{cluster:time} = \sigma^2_e/\sigma^2$	$\sigma^2 f_{individual} = \sigma^2_e$