Published in final edited form as:

Stat Methods Med Res. 2018 April ; 27(4): 1202–1218. doi:10.1177/0962280216657852.

Non-randomized and randomized stepped-wedge designs using an orthogonalized least squares framework

Yirui Hu¹ and Donald R Hoover^{1,2}

¹Department of Statistics and Biostatistics, Rutgers University New Brunswick, New Jersey, USA

²Institute for Health, Health Care Policy and Aging Research at Rutgers the State University of New Jersey, New Jersey, USA

Abstract

Randomized stepped-wedge (R-SW) designs are increasingly used to evaluate interventions targeting continuous longitudinal outcomes measured at T -fixed time points. Typically, all units start out untreated, and randomly chosen units switch to intervention at sequential time points until all receive intervention. As randomization is not always feasible, non-randomized stepped-wedge (NR-SW) designs (units switching to intervention are not randomly chosen) have attracted researchers. We develop an orthogonalized generalized least squares framework for both R-SW and NRSW designs. The variance of the intervention effect estimate depends on the number of steps (S), length of step sizes (t_s), and number of units (n_s) switched at each step ($s=1, \dots, S$). If all other design parameters are equal, this variance is higher for the NR-SW than for the equivalent R-SW design (particularly if the intercepts of non-randomly stepped switching strata are analyzed as fixed effects). We focus on balanced stepped-wedge (BR-SW, BNR-SW) designs (where t_s and n_s remain constant across s) to obtain insights into optimality for variance of the estimated intervention effect. As previously observed for the BR-SW, the optimal choice for number of time points at each step is also $t_s \equiv 1$ for the BNR-SW. In our examples, when compared to BR-SW designs, equivalent BNR-SW designs even with intercepts of non-randomly stepped switching strata analyzed using fixed effects sacrifice little efficiency given an intra-unit repeated measure correlation $\rho \approx 0.50$. Compared to traditional difference-in-differences designs, optimal BNR-SW designs are more efficient with the ratio of variances of these designs converging to 0.75 when $T > 10$. We illustrate these findings using longitudinal outcomes in long-term care facilities.

Keywords

Non-randomized stepped-wedge; orthogonalized least squares; compound symmetry covariance approximation; power and sample size estimation; optimal design; fixed and random effects

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Yirui Hu, Rutgers University New Brunswick, 110 Frelinghuysen Road, 501 Hill Center, Piscataway, New Jersey, NJ 08901-8554, USA. yirui.hu@gmail.com.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

1 Introduction

Recently developed randomized stepped-wedge (SW) designs^{1,2} are applied to longitudinal outcomes repeatedly measured at T -fixed time points in N units being placed on the new intervention over time using a staggered schedule. For the examples used in this paper, a unit is a single medical facility undergoing a facility-wide intervention with facility-level measurements taken over time. Although we do not have the person-level data within these units, we extend to such designs in Appendix 1 where units could in fact be individual persons undergoing person-level interventions against chronic conditions being measured over time. Typically, at the first time point, all units are not on the new intervention (i.e. untreated). More units (who then remain on the intervention until end of study) are switched onto the new intervention (i.e. treated) at subsequent time points. A pooled comparison of the study outcome for “treated” versus “untreated” unit-measures that adjusts for secular time effect is made. The SW designs are increasingly implemented in diverse areas including: cardiovascular disease,^{3,4} cancer,^{5,6} HIV,⁷ respiratory disease,⁸ nutrition,⁹ maternal and child health,¹⁰ and health care financing.¹¹

Shifting of units onto intervention at different times complicates SW analysis in that we need to account for secular time effects and intra-unit correlation.^{1,12} Often the randomized stepped-wedge (R-SW) is applied to “cluster-randomized trials” where the “unit” is a cluster of a fixed $m > 1$ individuals with a shared intercept measured at fixed time points.^{1,2,12,13} The analysis is based on the means of all individuals in the cluster at each time. However, the R-SW is also used for cases where there is only 1 measure at each time point (i.e. $m = 1$) sometimes referred to as a wait list design.¹³ This is also the case for the examples in our analysis so our derivations are for $m = 1$ as is typically the case in facility-level analyses for health care settings.^{14,15} Appendix 1 gives a conversion between our setting and cluster-randomized trials with $m > 1$. Hussey and Hughes¹² provided approaches to sample size and power calculations for R-SW implementation of cluster-randomized trials. The R-SW often has greater statistical power than traditional designs including: randomized longitudinal parallel designs,¹⁶ where randomly chosen participants treated at all time points are compared to those untreated at all time points and randomized longitudinal difference-in-differences (R-DD) designs,¹⁷ where all units start out untreated and are measured for a fixed number of time points, then at the same fixed time point, a randomly chosen subset of units is switched to the intervention and remain switched until the end.

While randomization of units to intervention arms is preferred as a gold standard to reduce bias and improve efficiency,¹⁸ it is not always feasible, particularly in health service settings, because of resource and logistical constraints.¹⁹ Thus, for example, non-randomized difference-in-differences (NR-DD) studies are applied to estimate impact of new interventions or policies.²⁰ In NR-DD designs, all units start out untreated and are followed for a fixed number of visits; then, non-randomly chosen units are switched to the new policy or intervention and compared to controls continuing to be untreated. However, non-randomized stepped-wedge (NR-SW) designs have recently attracted researchers.^{21,22} While it is known that generalized linear mixed models^{12,23} can evaluate intervention effects on continuous outcomes under normal approximation, these have not been formally applied to NR-SW designs including for power and sample size estimation.

This paper develops a unifying orthogonalized framework for stepped-wedge designs and obtains simple formulas for variance of intervention effects on continuous outcomes. Section 2 develops a general linear model for both R-SW and NR-SW designs and proposes an orthogonalized design matrix to simplify derivation of variance estimates. Section 3 presents general least squares (GLS) estimates for variance of intervention effect using a within unit compound symmetry repeated measure correlation and discusses the general framework for power estimation. Section 4 focuses on a special but common case of stepped-wedge designs where equal numbers of units are switched at equally spaced times that we denote as balanced designs (BR-SW for randomized and BNR-SW for non-randomized) and derives simple closed form solutions for variance of the intervention effect estimate for these designs. Variances of optimal designs for BR-SW and BNR-SW studies are compared. Section 5 compares the optimal BNR-SW design to the optimal NR-DD studies. Section 6 presents illustrative examples from long-term care facilities and Section 7 summarizes and discusses possible future work.

2 Stepped-wedge models

2.1 Notation

Let T be the number of measured time points, S be the number of these time points at which one or more units are transitioned onto intervention (i.e. “steps”) and N be the total number of units. Typically, at the first step, all of the units start in the control (untreated) condition and baseline measurements are taken, although we expand this to allow some units treated at baseline. Once switched onto the intervention, units remain treated until the end. At the last step, often all units have switched to the intervention, but we expand to allow some units to remain untreated. Let $s = 1, \dots, S$ enumerate the ordered stratum (of units) that is switched to intervention per step. Here a stratum is a group of units that share a common characteristic, i.e. the time when the intervention is first delivered. Let A_s denote the s th “shifting strata” (the subset of units that switched to the intervention at step s) and n_s be the number of units in A_s .

Let $i = 1, \dots, N$ enumerate the units with the enumeration ordered by the stratum. For instance, A_1 contains $\{1, 2, \dots, n_1\}$, i.e. stratum $s = 1$ with units $i \in A_1$ switched to the intervention at the first step; A_2 contains $\{n_1 + 1, n_1 + 2, \dots, n_1 + n_2\}$, i.e. stratum $s = 2$ with units $i \in A_2$ switched to the intervention at the second step and so on. The number of

consecutive time periods or step size per step is denoted $(t_0, t_1, \dots, t_S)$ with $\sum_{s=0}^S t_s = T$, but note that $t_0 = 0$ if some (i.e. n_1) units are already treated when the study starts. If all units have not been shifted to treatment by the end of the study, $t_S = 0$ and n_S denotes the number

of units never shifted onto treatment. Let $j_s = \sum_{l=0}^{s-1} t_l$ denote the first time point that units in the s th stratum are treated (with $j_s = \infty$ if some units are never treated); and Z_{ij} denote if unit i is treated at time j (0=no, 1=yes) with $Z_{ij} = 0$ if $j < j_s$ and $Z_{ij} = 1$ if $j \geq j_s$ where s is the ordered stratum that unit i belongs to. Figure 1 illustrates the general SW study where all units start out untreated ($t_0 > 0$), and all units are treated after the last step ($t_S > 0$).

We expand the general SW design to three special subcases: (1) Not all units are shifted onto treatment (or $t_S = 0$), even though the last n_S units are not shifted onto treatment, they do constitute a “shifting strata” and step “ S ” can be thought of as “never shifted”. Note when $S = 2$, this subcase reduces to a DD design. (2) The study does not begin until after the first n_1 units have been put on treatment. The fact that these units are never untreated is captured by $t_0 = 0$. (3) Both previous conditions $t_0 = t_S = 0$ happen with some units started on treatment and some units never shifted to treatment. Note this subcase reduces to a parallel design when $S = 2$.

2.2 Statistical model and orthogonal coding for design matrix

Let Y_{ij} be the measurement at the j th time point from unit i . Again for this paper, we only have single facility-level measures over time. If there are m patient-level measures nested within each facility then $\bar{Y}_{ij}$ can be analyzed using the conversion in Appendix 1.

For any given unit i at time j , we can model the outcome of interest as $Y_{ij} = \mu_i + \beta_j + \theta Z_{ij} + \varepsilon_{ij}$. Here μ_i is the main effect for unit i , β_j is the main effect for time j , θ is the effect of the intervention and ε_{ij} is random error. The intervention effect (θ) is modeled as an “immediate jump” effect and remains constant in the post-intervention measurements. Now $\mu_i \sim N(\alpha_0, \tau^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ with all previous terms being independent. Randomization, also known as random allocation of the units into shifting strata, results in each unit having an equal probability of being assigned to each of the S shifting strata. The purpose of randomization is to eliminate allocation bias and achieve shifting strata similar in baseline characteristics.²⁴ If we subsume the random unit deviation ($\mu_i - \alpha_0$) into the error term ε_{ij}^* (i.e. $\varepsilon_{ij}^* = (\mu_i - \alpha_0) + \varepsilon_{ij}$), the R-SW model is

$$Y_{ij} = \alpha_0 + \beta_j + \theta Z_{ij} + \varepsilon_{ij}^* \quad (1)$$

Now $\varepsilon_{ij}^* \sim N(0, \sigma^2)$ where $\sigma^2 = \tau^2 + \sigma_e^2$ and is independent between different units as shown in equation (6) but has correlation $\rho = \frac{\tau^2}{\tau^2 + \sigma_e^2}$ between two time points j and j' within the same unit i .

As discussed in Appendix 1, we should caution that our notation for ρ uses the wait list design or an already averaged unit as a single observation. This differs from that used in most cluster-randomized stepped-wedge design papers as our response is a single measure (Y_{ij}) as opposed to the average of m independent observations for a cluster i at time j ($\bar{Y}_{ij}$).

To convert between the two notations, our $\rho = \frac{\tilde{\rho}}{\tilde{\rho} + \frac{1-\tilde{\rho}}{m}}$ where $\tilde{\rho}$ denotes the “ ρ ” used in “cluster randomization notation” papers.^{12,25–27}

For a non-randomized design with S steps (and thus S shifting strata), we assume that the non-randomization is associated with the central tendency (mean) of the observations within the strata, but not otherwise with the trajectories. The mean effect is no longer a common α_0 , but differs by α_s ($s = 1, \dots, S$) that captures the non-randomization displacement for

being in shifting stratum s . The shared random error of the unit effect (i.e. $\mu_j - \alpha_s$) is again subsumed into the variance and within unit covariance of ε_{ij}^* (i.e. $\varepsilon_{ij}^* = (\mu_j - \alpha_s) + \varepsilon_{ij}$). The model being fit for NR-SW is thus

$$Y_{ij} = \alpha_s I_{\{i \in A_s\}} + \beta_j + \theta Z_{ij} + \varepsilon_{ij}^* \quad (2)$$

We consider in equation (2) the strata effects (i.e. $\alpha_s - \alpha_0$) to be “fixed” rather than “random” effects.²⁸ In many NR-SW settings, it may be hard to argue that strata effects are random (for example with respect to strata order and hence number of time points treated), normally distributed, and/or that a random effects state model is numerically stable.²⁹ The Hausman Test²⁸ for this admissibility of random effects models could be used in such settings. Or correlation of α_s with s could be examined, and random effects models not be used for non-zero correlation.

However, if the strata effects are considered random in the NR-SW model, then the assumption on the covariance of ε_{ij}^* will be different from that in equation (6). For the NR-SW with the strata effects considered random, measurements from units in different strata are independent, and the repeated measure correlation within the same unit is ρ , but now the correlation of the error ε_{ij} and ε_i from two different units in the same stratum is ρ_s where $0 \leq \rho_s \leq \rho$ as shown in (7).

For both R-SW and NR-SW in equation (1) and (2), the coding for intervention effect (Z_{ij}) is effectively (0, 1) with 0 for control and 1 for intervention. For t_0 (possibly zero) baseline measures, all units stay in control and the coding is 0. In the “build-up” steps, the coding switches from 0 to 1 sequentially as $j \rightarrow j_s$. Eventually for the last step (unless $t_S = 0$), every unit receives intervention and the coding is 1. However, to obtain an orthogonalized decomposition of the intervention parameters from the time parameters, we re-parameterize equations (1) and (2) as below.

For R-SW

$$Y_{ij} = \alpha_0 + \beta_j^* + \theta Z_{ij}^* + \varepsilon_{ij}^* \quad (3)$$

For NR-SW

$$Y_{ij} = \alpha_s I_{\{i \in A_s\}} + \beta_j^* + \theta Z_{ij}^* + \varepsilon_{ij}^* \quad (4)$$

where $Z_{ij}^* = 0$ for if $j < j_1$ or if $j = j_S$ as all units are in the same treatment condition, otherwise for $j_s \leq j < j_{s+1}$

$$Z_{ij}^* = \begin{cases} -\frac{\sum_{l=1}^s n_l}{N}, & \text{if } Z_{ij}=0 \\ \frac{N - \sum_{l=1}^s n_l}{N}, & \text{if } Z_{ij}=1 \end{cases} \quad (5)$$

Equivalently, $Z_{ij}^* = Z_{ij} - \frac{\sum_{l=1}^s n_l}{N}$ and thus $\beta_j^* = \beta_j + \frac{\sum_{l=1}^s n_l}{N} \theta$.

The orthogonal coding of intervention effect (5) for unit-time can be found in Appendix 2. The advantage of the proposed orthogonal coding is it simplifies the solution for the GLS estimates obtained in Section 3.

3 GLS variance formula and power estimation

3.1 General formula for GLS estimate

The matrix forms of equations (3) and (4) can be written as: $Y = X\beta + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 V)$. Here X represents the design matrix and Y is a vector of outcomes. For equation (4) with the general parameter vector $\beta = (\alpha_1, \dots, \alpha_s, \beta_1^*, \dots, \beta_{T-1}^*, \theta)$, the corresponding X has columns $(I_{\{j \in A_1\}}, \dots, I_{\{j \in A_S\}}, J_1, \dots, J_{T-1}, Z^*)$ with N rows per column. Z^* is a column with orthogonal coding in equation (5) (as shown in Appendix 2) and $J_1, \dots, J_{T-1}$ are columns corresponding to dummy $T-1$ independent time, coded (1, 0, -1) where -1 at reference time

$-b$, 1 at time j , and 0 at all other times, so here and elsewhere with for $j=T$, $\beta_T = -\sum_{j=1}^{T-1} \beta_j$

under the fixed effects constraint $\sum_{j=1}^T \beta_j = 0$. Similarly for equation (3) except that

$\beta = (\alpha_0, \beta_1^*, \dots, \beta_{T-1}^*, \theta)$.

As shown below for the R-SW in equation (3) and the fixed effects NR-SW in equation (4), the covariance matrix V is the overall correlation matrix of ε_{ij}^* , which is made up of N block diagonals of V_0 with all off-block diagonal matrix elements being 0. Each V_0 is the correlation matrix of repeated measures within each single unit, with dimension T .

$$V = \begin{pmatrix} V_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_0 \end{pmatrix}_{NT}, \text{ where } V_0 = \begin{pmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}_T \quad (6)$$

The NR-SW with the strata effects treated as random uses equation (3) but with V now being a block diagonal of shifting stratum variances V_s ($s = 1, \dots, S$) where the random strata effects are subsumed in ρ_s .

$$V = \begin{pmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_S \end{pmatrix}_{NT} \text{ with } V_s = \begin{pmatrix} V_0 & \cdots & \rho_s \\ \vdots & \ddots & \vdots \\ \rho_s & \cdots & V_0 \end{pmatrix}_{n_s T} \quad (7)$$

With V_0 as defined above and all V_s correlations not in the V_0 being ρ_s , the intra-stratum correlation is mediated by the non-randomization selection effect. The GLS estimate for β is $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$ and has variance $\Lambda = (X'V^{-1}X)^{-1}\sigma^2$ where Λ is a square matrix of order $(S+T)$ for NR-SW and $(T+1)$ for R-SW. The variance of $\hat{\theta}$ is the last diagonal element of Λ . This $\hat{\beta}$ is the best linear unbiased estimator (BLUE) for β and uniform minimum variance (UMVU) if Y_{ij} is normally distributed.³⁰

Although the inverse of Λ is complicated, the orthogonalized coded $X = (I_{\{i \in A_1\}}, \dots, I_{\{i \in A_S\}}, J_1, \dots, J_{T-1}, Z^*)$, for the fixed effects NR-SW based on equation (4) with V as defined by equation (6) simplifies $X'V^{-1}X$ with most cross-products being zero which simplifies derivation of $(X'V^{-1}X)^{-1}$.

As Appendix 3 proves, $\text{Var}(\hat{\theta}) = \frac{1}{\text{OTD} - \text{IDP}}\sigma^2$, where $\text{OTD} = \sum_{s=1}^S n_s Z_s^{*'} V_0^{-1} Z_s^*$ stands for the “orthogonalized treatment dispersion” (Z_s^* is the orthogonal coding for the s th

stratum in Table 3 as defined Appendix 2) $\text{IDP} = \sum_{s=1}^S \frac{[1+(T-1)\rho]}{n_s T} (I'_{\{i \in A_s\}} V^{-1} Z_s^*)^2$ denotes “intercept dispersion penalty” and reflects reduction from OTD by the treatment term being dispersed about different intercepts α_s (rather than always about a common α_0) due to the fixed effects non-randomization. Full expansions of OTD and IDP are presented in Appendix 3.

For the fixed effects NR-SW

$$\text{Var}(\hat{\theta}_{NR-SW}) = \frac{1}{\text{OTD} - \text{IDP}}\sigma^2 = \frac{(1-\rho)\sigma^2}{\sum_{l=1}^{S-1} t_l \left(\frac{\sum_{h=1}^l n_h}{N} \right) \left(\frac{N - \sum_{h=1}^l n_h}{N} \right) - \frac{1}{T} \sum_{s=1}^S n_s \left[\sum_{l=1}^{S-1} t_l \left(\frac{N - \sum_{h=1}^l n_h}{N} \right) - \sum_{h=1}^{s-1} t_h \right]^2}$$

(8)

Note that here and elsewhere, $\text{Var}(\hat{\theta}_{NR-SW})$ denotes the variance for “fixed effects” modeling.

For the R-SW, $\text{IDP} = 0$, therefore the variance reduces to

$$\begin{aligned} \text{Var}(\hat{\theta}_{R-SW}) &= \frac{1}{\text{OTD}}\sigma^2 \\ &= \frac{(1-\rho)\sigma^2}{\frac{1}{(1-\rho)} \sum_{l=1}^{S-1} t_l \left(\frac{\sum_{h=1}^l n_h}{N} \right) \left(\frac{N - \sum_{h=1}^l n_h}{N} \right) - \frac{\rho}{[1+(T-1)\rho]} \sum_{s=1}^S n_s \left[\sum_{l=1}^{S-1} t_l \left(\frac{N - \sum_{h=1}^l n_h}{N} \right) - \sum_{h=1}^{s-1} t_h \right]} \end{aligned}$$

(9)

For the same given SW design, $Var(\hat{\theta}_{R-SW})$ is lower with randomization of units into the shifting strata (compared to fixed effects NR-SW) by the ratio of equation (9)/equation (8)

being $\frac{OTD - IDP}{OTD} < 1$.

Finally, the solution to $Var(\hat{\theta})$ for the random effects NR-SW from equation (7) is difficult as the ρ_s elements in the V_s lead to numerically complicated inverses and thus is not presented here.

Both $Var(\hat{\theta}_{NR-SW})$ and $Var(\hat{\theta}_{R-SW})$ are invariant to t_0 and t_S conditional on the sum $t_0 + t_S = E$ (where E denotes the number of time points where all units are homogeneous with respect to intervention assignment which occurs on the front or back “edges” of the SW) for an otherwise identical design. Because of the orthogonal coding, $Z_{ij}^* = 0$ for $j < j_1$ and $j > j_S$, meaning that observations falling in these periods contribute equally to IDP and OTD and doing so only by dampening V_0^{-1} . For example, in a stepped-wedge design with $S = 3$ and $(t_0, t_1, t_2, t_3) = (1, 1, 1, 2)$, $Var(\hat{\theta})$ is invariant to t_0 (or t_3) given the sum of the two is some fixed value E where $E = t_0 + t_3 = 3$. Thus, $(t_0, t_3) = (1, 2)$ and $(t_0, t_3) = (2, 1)$ (with $t_1 = t_2 = 1$ remaining the same) achieve the same $Var(\hat{\theta})$ since $1+2 = 2+1 = 3$.

3.2 Variance and power estimation

Hussey and Hughes¹² determined power for R-SW designs by using a Wald test for intervention effect. Similarly, we consider two hypotheses for intervention effect with S steps and T total visits: $H_0 : \theta = 0$; $H_1 : \theta = \pm\theta_1$. Here θ_1 is the minimum detectable difference (or effect size δ_1 expressed as a multiple of σ , i.e., $\theta_1 = \delta_1\sigma$) for a stepped-wedge design given α , β , N . For practical repeated measure designs, the sample sizes are often large enough to permit normal approximation of the non-central t distribution when $df > 30$.³¹ With α and β being Type I and Type II errors, these are met for a given θ_1 if $Var(\hat{\theta})$ is such that

$$\theta_1 = (z_{1-\frac{\alpha}{2}} + z_{1-\beta}) \sqrt{Var(\hat{\theta})} \quad (10)$$

For the fixed effects NR-SW and the R-SW, respectively, $Var(\hat{\theta})$ is obtained from the GLS variance estimates in equations (8) and (9). Thus, if Φ is the cumulative distribution function for standard normal $N(0, 1)$, rearranging equation (10) gives the power for conducting a two-sided test of size α as

$$Power = \Phi\left(\frac{\theta_1}{\sqrt{Var(\hat{\theta})}} - z_{1-\frac{\alpha}{2}}\right) \quad (11)$$

4 Balanced SW designs and optimality properties

This section focuses on a specific design we define as “balanced” stepped-wedge, which numerically simplifies formulas and thus enables derivation of optimality properties. The SW design is balanced if the same number of units ($n_1 = \dots = n_S = n$) are switched per step with equal step sizes ($t_0 = t_1 = \dots = t_S = t$). For the general balanced design given t and n , starting at time $t+1$, n units switch to intervention and then t measures are taken before the next switch. This continues until all units have switched and after t more measures the study ends; thus, $T = (1 + S) * t$ and $N = n * S$. Such balanced designs could occur in practice if say it took t time points to ramp up application to n new units for each new step. Balanced designs can be either non-randomized (BNR-SW) or randomized (BR-SW). This section first presents and compares simplified formulas for variance of the fixed effects BNR-SW and BR-SW, then investigates the optimal design to find the optimal values of S (or equivalently t) to achieve greatest power for a balanced design with fixed T and N . Finally, efficiency of the optimal fixed effects BNR-SW design is compared to that of the NR-DD for any given T and ρ .

4.1 Power and sample size estimation for balanced SW designs

For a balanced fixed effects BNR-SW design, variance in equation (8) simplifies to

$$\text{Var}(\hat{\theta}_{\text{BNR-SW}}) = \frac{12S(S+1)}{NT(S-1)(S+2)}(1-\rho)\sigma^2 \quad (12)$$

Again here and elsewhere $\hat{\theta}_{\text{BNR-SW}}$ denotes the estimated intervention effect for the fixed effects balanced nonrandomized model as we did not derive closed form variance estimates for random effects non-randomized models.

For a balanced BR-SW design, variance in equation (9) simplifies to

$$\text{Var}(\hat{\theta}_{\text{BR-SW}}) = \frac{6S[1+(T-1)\rho](1-\rho)}{NT(S-1)[1+(T-1-\frac{St}{2})\rho]}\sigma^2 \quad (13)$$

Note that equation (13) is the same as formulas in Woertman et. al.²⁵ and Hussey and

Hughes¹² after the previously noted conversion $\rho = \frac{\tilde{\rho}}{\tilde{\rho} + \frac{1-\tilde{\rho}}{m}}$ (also illustrated in Appendix 1), where $\tilde{\rho}$ denotes the intra-class correlation in their cluster-randomized SW papers.

As Appendix 4 shows (with σ_{FE-NR}^2 , σ_R^2 , ρ_{FE-NR} , ρ_R defined in that Appendix), compared to fitting any R-SW, the NR-SW with fixed strata effects (FE) under the same setting will result in a lower model within population measurement variance on Y_{ij} ($\sigma_{FE-NR}^2 < \sigma_R^2$) together with a smaller within unit correlation of Y_{ij} and Y_{ij}' ($\rho_{FE-NR} < \rho_R$) due to elimination of variance about the α_s from the total variance about a common α_0 . However, from equations (8) and (12), the $\text{Var}(\hat{\theta}_{\text{NR-SW}})$ only depends on σ^2 and ρ through the product

$(1 - \rho)\sigma^2$. To that end, this product is unchanged by application of the fixed effects NR-SW design in that $(1 - \rho_{FE-NR})\sigma_{FE-NR}^2 = (1 - \rho_R)\sigma_R^2 = \sigma_e^2$. This invariance property means that the “randomized study design” effect parameters σ^2 and ρ can be directly used in equations (8) and (12) for the estimation of the variance of the intervention effect estimate no matter the impact of the fixed effects NR-SW on the final σ^2 and ρ .

4.2 Optimal t for balanced SW design

A balanced SW design may have a fixed total number of longitudinal times T because of budget and/or time constraints. For example, a study may be funded for $T = 6$ monthly measures on each unit. Finding the optimal balanced SW design with regards to the step size t from all possible integer step sizes ($t = 1, 2$ and 3) that can maximize power (or minimize the sample size needed to obtain a given power) would be important. We start with fixed effects balanced non-randomized designs (BNR-SW), which corresponds to finding the optimal t^* (or equivalently S^*) that maximizes the power (by minimizing the variance in equation (8)), given T , N , and ρ .

$$S^* = \arg \max_S \text{Power} = \arg \min_S \text{Var}(\hat{\theta}_{BNR-SW})$$

$$t^* = \frac{T}{1 + S^*}$$

The derivative of $\log(\text{Var}(\hat{\theta}_{BNR-SW}))$ in equation (8) with respect to S is

$(\frac{1}{S} - \frac{1}{S-1}) + (\frac{1}{1+S} - \frac{1}{2+S})$ and is negative for $S \geq 1$ meaning $\text{Var}(\hat{\theta}_{BNR-SW})$ monotonically decreases as S increases. The optimal S^* should be as large as possible, and the corresponding optimal t^* should be as small as possible. Accordingly, $t^* = 1$ maximizes power. Likewise for randomized balanced BR-SW designs, the derivative of

$\log(\text{Var}(\hat{\theta}_{BR-SW}))$ in equation (13) with respect to S , is $(\frac{1}{S} - \frac{1}{S-1}) - \frac{T\rho}{2(1+S)^2}$, which is also negative for $S \geq 1$ and thus is optimized by $t^* = 1$ as has been previously observed or surmised.^{12,25–27}

In the Supplementary Appendix (available at: smm.sagepub.com), we investigate designs that are equivalent in terms of N , T , and S with $(t_1 = \dots = t_{S-1} = t)$ but t_0 and t_S unconstrained, which we denote as “internally balanced”. For these designs, variance is often minimized with t_0 and t_S being less than t , which for $t = 1$ is at $t_0 = t_S = 0$. However, if many pre-existing baseline ($t_0 \gg 1$) or/and post full implementation ($t_S \gg 1$) measures will be available, the Supplementary Appendix shows that reducing the number of steps (i.e. S) may increase power.

4.3 Variance ratio of balanced fixed effects designs BNR-SW to BR-SW designs

For the balanced setting with the same N , T , S and t , the ratio of $\frac{\text{Var}(\hat{\theta}_{BR-SW})}{\text{Var}(\hat{\theta}_{BNR-SW})}$ in equation (9)/equation (8) reduces to equation (13)/equation (12) which is a function of T and ρ that falls between $\frac{1}{2}$ and 1 as shown below.

$$\frac{1}{2} < \frac{\text{Var}(\hat{\theta}_{BR-SW})}{\text{Var}(\hat{\theta}_{BNR-SW})} = \frac{(T+t)[1+(T-1)\rho]}{2T[1+(\frac{T+t}{2}-1)\rho]} < 1 \quad (14)$$

In particular, with optimal choice $t^* = 1$, the ratio in equation (14) reduces to

$$\frac{1}{2} < \frac{\text{Var}(\hat{\theta}_{BR-SW})}{\text{Var}(\hat{\theta}_{BNR-SW})} \big|_{\{t^*=1\}} = \frac{(T+1)[1+(T-1)\rho]}{2T[1+(\frac{T-1}{2})\rho]} < 1 \quad (15)$$

The ratio plots in Figure 2 summarize the variance comparisons for optimal balanced BR-SW versus fixed effect BNR-SW designs ($t = 1$) as functions of T and ρ . The ratio of variances in equation (14) converges to 1 as ρ increases for any T , but at a slower rate as T becomes larger.

5 Comparing balanced fixed effects BNR-SW to NR-DD

It is also of interest to assess the relative efficiency of NR-SW compared to the more traditional NR-DD study with the same N and T as there may be settings where an investigator is not able to randomize and has to choose between SW and DD designs. The optimal NR-DD design (Fixed Effects) which minimizes $\text{Var}(\hat{\theta}_{NR-DD})$ switches n^* units to

intervention after b^* time points where for N even, $n^* = \frac{N}{2}$,

$$b^* = \begin{cases} \frac{T}{2}, & \text{if } T \text{ is even} \\ \frac{T-1}{2} \text{ or } \frac{T+1}{2} & \text{if } T \text{ is odd} \end{cases}$$

The $\text{Var}(\hat{\theta}_{NR-DD})$ in this optimal NR-DD design is $\frac{4T}{Nb^*(T-b^*)}(1-\rho)\sigma^2$.

Taking the ratio of the variance of fixed effect balanced NR-SW versus optimal NR-DD gives equation (16).

$$\frac{\text{Var}(\hat{\theta}_{\text{BNR-SW}})}{\text{Var}(\hat{\theta}_{\text{NR-DD}})|\{\text{optimal } b^*\}} = \frac{3S(S+1)b^*(T-b^*)}{T^2(S-1)(S+2)} = \begin{cases} \frac{3(S+1)S}{4(S-1)(S+2)}, & \text{if } T \text{ is even} \\ \frac{3(S+1)S}{4(S-1)(S+2)} \frac{(T-1)(T+1)}{T^2}, & \text{if } T \text{ is odd} \end{cases}$$

(16)

In particular, for optimal balanced NR-SW with $t^* = 1$, the variance in equation (12)

simplifies to $\text{Var}(\hat{\theta}) = \frac{12(T-1)(1-\rho)}{N(T-2)(T+1)}\sigma^2$ and the ratio in equation (16) reduces to equation (17).

$$\frac{\text{Var}(\hat{\theta}_{\text{BNR-SW}})|\{\text{optimal } t^*=1\}}{\text{Var}(\hat{\theta}_{\text{NR-DD}})|\{\text{optimal } b^*\}} = \frac{3(T-1)b^*(T-b^*)}{T(T-2)(T+1)} = \begin{cases} \frac{3(T-1)T}{4(T-2)(T+1)}, & \text{if } T \text{ is even} \\ \frac{3(T-1)^2}{4T(T-2)}, & \text{if } T \text{ is odd} \end{cases}$$

(17)

The ratio in equations (16) and (17) depends only on T and S (or equivalently T and t based

on $t = \frac{T}{1+S}$). Figure 3 uses equation (16) to illustrate the ratio of variance for balanced fixed effects NR-SW versus optimal NR-DD for different values of t . Each symbol stands for one specific value of t . Note that to fit a balanced SW, T must be divisible by t . Under fixed effects for the same T ($T > 3$) and N , the ratio in equation (17) for optimal balanced NR-SW

and NR-DD indicates $\frac{\text{Var}(\hat{\theta}_{\text{BNR-SW}})}{\text{Var}(\hat{\theta}_{\text{NR-DD}})} < 1$. Now the two designs produce the same variance when $T = 3$. But the ratio in equation (17) increases roughly as T increases, and converges to 0.75 when $T > 15$. However, for a given T as t increases, the variance reduction of $\text{Var}(\hat{\theta})$ from the fixed effects NR-SW versus the optimal NR-DD, reduces and can reverse, but again for any fixed t , as T increases, the ratio decreases and converges to 0.75.

6 Examples from NEW JERSEY long-term care facilities

Both urinary incontinence and ulcers (bedsores) are common chronic conditions for residents that are improved by better treatment at long-term care facilities (LTCF). Five Star Quality Data³² over seven quarters from Spring 2012 through Fall 2013 reported the average percentages of all long-stay residents that had incontinence and ulcers in each of 270 New Jersey long-term facilities. The overall quarter averaged unit average of binary outcomes with incontinence was 32 % (or 0.32) with $\sigma = 0.14$ and correlation of repeated measures in the same unit was $\rho = 0.85$. For ulcers, the unit average was 9.5 % (or 0.095) with $\sigma = 0.0475$ and $\rho = 0.50$. We used this normative data to guide estimation of minimal detectable effect sizes δ_1 for BNR-SW designs and compare these to NR-DD and BR-SW designs for a

two sided $\alpha = 0.05$, $\beta = 0.20$ on intervention trials that would be conducted at $N = 30$ long-term care facilities lasting from 1.25 ($T = 6$) years. While ρ ranged from 0.50 to 0.85, for outcomes in New Jersey LTCF, we added $\rho = 0.00$ and 0.30 to provide insight into outcomes with lower ρ .

Suppose it is impossible to plan a randomized study and one must choose between a NR-DD and a BNR-SW. Table 1 presents the minimal effect size δ_1 that can be detected with fixed effect analysis of BNR-SW designs for $t = 1, 2$, and 3 from equations (10) and (12) and

optimal NR-DD for $b^* = \frac{T}{2}$. We do not consider random effects model here as it will be directionally biased for NR-DD with only two strata that are unbalanced with respect to proportions treated.

Thus, using ulcers ($\rho = 0.50$), with $T = 6$ and $t = 1$, the minimal detectable effect size from the BNR-SW fixed effects design is $\delta_1 = 0.529$ or $0.095 \pm 0.529 \times 0.0475$ which is 0.07 or 0.12. By contrast, the minimal detectable $\delta_1 = 0.590$ or $0.095 \pm 0.590 \times 0.0475$ which is 0.067 or 0.123 for an NR-DD. While the BNR-SW is preferable from this standpoint, one would have to consider if this benefit were enough if the SW design was more complicated to implement. For $t = 2$, the BNR-SW is less efficient than the NR-DD when $T = 6$.

We caution the reader on one point for interpreting Table 1. Often a large number of baselines with $t_0 > t$ measures is available from historical data where all units were untreated over a long longitudinal monitoring period meaning the t_s 's can only be balanced in the future with $t_0 > t_1 = \dots = t_s$. While the full details are beyond this paper, the Supplementary Appendix (available at: smm.sagepub.com) suggests that if this is the case, the NRDD approach becomes more favorable relative to the best possible NR-SW than what is seen in Table 1.

Now suppose a balanced SW design will be used, but the investigator wants to determine if randomization is worth the extra effort. Table 2 compares minimal detectable δ_1 for BR-SW (from equation (13)), BNR-SW designs analyzed as fixed effects (from equation (12)) and as random effects (calculated on computer using $\Lambda = (X' V^{-1} X)^{-1} \sigma^2$ with V based on equation (7)). We let σ and ρ be the same for the BNR-SW and BR-SW random/fixed effect designs since as Section 4.1 and Appendix 4 show any changes on σ and ρ from non-randomization in the fixed effects NR-SW formulation cancel out. We assume ρ_s is proportional to ρ as it seems reasonable that the level of differentiation between the non-randomized shifting strata intercepts will be proportional to the differentiation between persons. We choose $\rho_s = 0.1\rho$ for a small, $\rho_s = 0.25\rho$ for a noticeable, and $\rho_s = \rho$ as an extreme value for intra-stratum correlation. Note the result for $\rho_s = 0$ (no strata effects) is mathematically the same as BR-SW.

Thus, for example, with $T = 6$, $\rho = 0.85$, $t = 1$, the randomization benefit is barely noticeable with the minimal detectable δ_1 only dropping from 0.290 in a fixed effects BNR-SW down to 0.287 in a R-SW. However, for $\rho = 0$ when $T = 6$, $t = 1$, the minimal detectable δ_1 drops from 0.749 in a fixed effects BNR-SW down to 0.572 in the BRSW. The BR-SW performs better than both fixed and random effects BNR-SW, and the random effects BNR-SW is more powerful than fixed effect BNR-SW in terms of minimum detectable effect size

particularly as ρ_s decreases. However, for larger values of ρ ($\rho = 0.50$), as was seen in New Jersey LTCF outcomes, the differences between minimal detectable δ_1 between even BR-SW and fixed effects BNR-SW are very small. This suggests that if $\rho = 0.50$, the penalty for doing BNR-SW instead of BR-SW on the variance of estimated intervention effect may be ignorable and also that a BNR-SW design should be analyzed as fixed rather than random effects to avoid bias. For smaller ρ ($\rho = 0.30$), well below the range of what we saw for outcomes in New Jersey LTCF, the range between minimal detectable δ_1 from BR-SW and random/fixed effects NR-SW is larger. In these settings, it may be more important to fit BR-SW or use random effects analysis on BNR-SW to preserve power. However, other designs such as randomized parallel may be preferable to the SW if $\rho = 0.30$.²⁶

7 Conclusion

This paper presents generalized stepped-wedge designs expanded to non-randomized settings. An orthogonalized framework for estimating GLS variance and corresponding power for intervention effects is developed assuming compound symmetry for intra-unit correlation of repeated measures. With the above orthogonal coding for intervention effect, we showed the following properties. First, for any given SW design, randomized unit allocation achieves greater power than does non-randomized allocation analyzed using fixed effects due to the added IDP penalty term in the denominator of the NR-SW variance estimate (i.e. equation (8) versus equation (9)). Second, for any otherwise equivalent R-SW and NR-SW fixed effect design, the GLS power estimate is invariant to t_0 (or t_S) conditional on the sum $t_0 + t_S (= E)$ because observations falling in those two edges contribute equally to the GLS variance.

To further investigate the optimal design in terms of power, we focused on the balanced SW design that simplifies variance formulas to equations (12) and (13). For both BR-SW and BNR-SW using fixed effects for a given T , the power increases as step size (t) decreases and thus the optimal design for both BR-SW and BNR-SW is with $t^* = 1$. In a more comprehensive investigation of optimality for the R-SW, Lawrie et al.²⁷ also observed optimization at $t^* = 1$, but as their analysis allowed n_s to vary (we did not), they observed having $\{n_1 = n_S\} > \{n_2 = \dots = n_{S-1}\}$ maximized efficiency.

For power approximation in balanced SW designs, the advantage of random allocation decreases as ρ increases and becomes ignorable when $\rho = 0.50$. Therefore, for ρ in this range, as was the case for our illustrative example of New Jersey LTCF, we believe a BNR-SW design even analyzed using fixed effects may achieve very similar power as does the comparable BR-SW design. However, potential biases from differential secular trends in nonrandomized designs need to be considered.³³ For small values of ρ (i.e. $\rho < 0.30$), we suggest researchers should be cautious to use NR-SW instead of R-SW and perhaps not use any SW design at all. However, in this range of ρ , use of random effects rather than fixed effects analyses to model the non-randomized strata effects considerably improves power in the BNR-SW design. Thus, further research into whether the strata effects could be modeled as random effects in a NR-SW design may be warranted.

In non-randomized settings with fixed T and N , we discovered that the optimal BNR-SW ($t^* = 1$) is always better than optimal NR-DD in terms of lower variance for the intervention effect estimate. More specifically, for all t the relative efficiency of BNR-SW to NR-DD increases as T increases, but the ratio of variance eventually converges to 0.75 as T gets larger. For any fixed T , as the step size t increases, the advantage of BNR-SW to NR-DD gets smaller and eventually reverses to favor NR-DD.

Several limitations should be mentioned. We assumed a constant intervention effect across unit and time, which could be extended by modeling an interaction term of intervention and time and/or including intervention heterogeneity into the covariance structure. While compound symmetry might be a usable approximation if the intra-unit repeated measure correlation does not change or decays slowly over time,³¹ there may be cases where time decay is too large for CS to be reasonable. In such cases, extension to Toeplitz decay covariance structures may yield a better power estimation. Analytical-based methods get much more complicated in random effects NR-SW models due to the additional level of intra-stratum correlation; it is unclear if simple variance estimates can be achieved for this setting. As is often done for stepped-wedge studies, we assumed normality or that sample size was large enough for the central limit theorem to hold. Future work on simulation-based methods may be promising to address all of the above issues because they provide flexible alternatives in power and sample size calculation that can deal with specific features of given studies at hand.³⁴

In conclusion, researchers have recognized the usefulness of stepped-wedge designs in recent years.^{2,12,25,35} While considerable development has been made into deriving variance estimates of the intervention effect, this process is still at the beginning for randomized designs and to our knowledge has not been explored for the nonrandomized setting. We developed an orthogonalized least squares framework for both R-SW and NR-SW in which number of steps (S), length of step sizes (t_s), and number of units switched at each step (n_s) can be varied. We then focused on balanced settings to obtain insights on optimal designs in terms of power. While BR-SW always achieved lower variance for the intervention effect estimate than did BNR-SW designs, the differences are small for $\rho = 0.50$. Compared to the traditional NR-DD design, optimal BNR-SW ($t^* = 1$) is more efficient. As the length of step size t increases in BNR-SW, the advantage over NR-DD gets smaller and eventually reverses. Further, as the Supplementary Appendix (available at: smm.sagepub.com) implies, a large number of historical untreated baseline time points may further shift the advantage to the NR-DD approach. Future work perhaps including structured simulations may help to clarify numerous unresolved issues.

Acknowledgments

This work was undertaken as part of Yirui Hu's PhD Dissertation. We thank both reviewers for helpful comments that have improved the paper.

Funding

This work was supported by NIH Grants 6U01AI035004, 6U01AI096299-06 and R01NR014632-01A1.

References

1. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodology*. 2006; 6:54.
2. Mdege N, Man M, Brown C, et al. Systematic review of stepped wedge cluster randomised trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol*. 2011; 64:936–948. [PubMed: 21411284]
3. Viera AJ, Garrett JM. Preliminary study of a school-based program to improve hypertension awareness in the community. *Fam Med*. 2008; 40:264–270. [PubMed: 18382839]
4. Liddy C, Hogg W, Russell G, et al. Improved delivery of cardiovascular care (IDOCC) through outreach facilitation: study protocol and implementation details of a cluster randomized controlled trial in primary care. *Implement Sci*. 2011; 6:110. [PubMed: 21952084]
5. Gambia Hepatitis Study Group. The gambia hepatitis intervention study. *Cancer Res*. 1987; 47:5782–5787. [PubMed: 2822233]
6. Husaini BA, Reece MC, Emerson JS, et al. A church-based program on prostate cancer screening for African American men: reducing health disparities. *Ethnicity Dis*. 2008; 18(2 Suppl. 2):179–184.
7. Hughes J, Goldenberg RL, Wilfert CM, et al. Design of the HIV prevention trials network (HPTN) protocol 054: a cluster randomized crossover trial to evaluate combined access to nevirapine in developing countries. *UW Biostat Work Paper Ser*. 2003 Working Paper 195.
8. Somerville M, Basham M, Foy C, et al. From local concern to randomised trial: the Watcombe housing project. *Health Expect*. 2002; 5:127–135. [PubMed: 12031053]
9. Ciliberto MA, Sandige H, Ndekha MJ, et al. Comparison of home-based therapy with ready-to-use therapeutic food with standard therapy in the treatment of malnourished Malawian children: a controlled, clinical effectiveness trial. *Am J Clin Nutr*. 2005; 81:864–870. [PubMed: 15817865]
10. Winani S, Wood S, Coffey P, et al. Use of a clean delivery kit and factors associated with cord infection and puerperal sepsis in Mwanza, Tanzania. *J Midwifery Womens Health*. 2007; 52:37–43. [PubMed: 17207749]
11. Allegri M, Pokhrel S, Becher H, et al. Step-wedge cluster-randomised community-based trials: an application to the study of the impact of community health insurance. *Health Res Policy Syst*. 2008; 6:10. [PubMed: 18945332]
12. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007; 28:182–191. [PubMed: 16829207]
13. Margaret AH, Dean S, Stephen S. Quasi-experimental designs in practice-based research settings: design and implementation considerations. *JABFM*. 2011; 24(5):589–596. [PubMed: 21900443]
14. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014; 312(22):2401–2402. [PubMed: 25490331]
15. Athey S, Imbens GW. Identification and inference in nonlinear difference-in-differences models. *Econometrica*. 2006; 74:431–497.
16. Moher D, Schultz KF, Altman DG. The CONSORT Group. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001; 357:1191–1194. [PubMed: 11323066]
17. Pearson D, Torgerson D, McDougall C, et al. Parable of two agencies, one of which randomizes. *Ann Am Acad Pol Soc Sci*. 2010; 628:11–29.
18. Meldrum ML. A brief history of the randomized controlled trial. From oranges and lemons to the gold standard. *Hematol Oncol Clin North Am*. 2000; 14:745–760. [PubMed: 10949771]
19. West SG, Duan N, Pequegnat W, et al. Alternatives to the randomized controlled trial. *Am J Public Health*. 2008; 98:1359–1366. [PubMed: 18556609]
20. Abadie A. Semiparametric difference-in-differences estimators. *Rev Econ Stud*. 2005; 72:1–19.
21. Brand SL, Musgrove A, Jeffcoate WJ, et al. Evaluation of the effect of nurse education on patient-reported foot checks and foot care behaviour of people with diabetes receiving haemodialysis. *Diabet Med*. 2016; 33(2):204–207. [PubMed: 26042333]

22. Davey C, Boulay M, Hargreaves JR. Strengthening nonrandomized studies of health communication strategies for HIV prevention. *J Acquir Immune Defic Syndr*. 2014; 66(Suppl 3):S271–S277. [PubMed: 25007196]
23. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am J Pub Health*. 2004; 94:423–432. [PubMed: 14998806]
24. Sedgwick P. Randomised controlled trials: balance in baseline characteristics. *BMJ*. 2014; 349:g5721. [PubMed: 25238747]
25. Woertman W, de Hoop E, Moerbeek M, et al. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol*. 2013; 66:752–758. [PubMed: 23523551]
26. Hemming K, Taljaard MJ. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. *Clin Epidemiol*. 2015 Sep 5. Epub ahead of print.
27. Lawrie J, Carlin JB, Forbes AB. Optimal stepped wedge designs. *Stat Probab Lett*. 2015; 99:210–214.
28. Hausman JA. Specification tests in econometrics. *Econometrica*. 1978; 46:1251–1271.
29. Snijders TAB. Fixed and random effect. *Encyclop Stat Behav Sci*. 2005; 2:664–665.
30. Aitken AC. On least-squares and linear combinations of observations. *Proc Royal Soc Edinburgh*. 1934; 55:42–48.
31. Fisher RA. Applications of “Student’s” distribution. *Metron*. 1925; 5:90–104.
32. [accessed 17 June 2016] Centers for Medicare and Medicaid Services Five Star Quality Rating System. www.cms.gov/medicare/provider-enrollment-and-certification/certificationandcompliance/fsqrs.html
33. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*. 1992; 11:1685–1704. [PubMed: 1485053]
34. Baio G, Copas A, Ambler G, et al. Sample size calculation for a stepped wedge trial. *Trials*. 2015; 16:354. [PubMed: 26282553]
35. de Hoop, EO. PhD Thesis. Netherlands: Thesis Radboud University Medical Centre; 2015. Efficient designs for cluster randomized trials with small numbers of clusters. <http://repository.ubn.ru.nl/bitstream/handle/2066/134179/134179.pdf> [accessed 17 June 2016]

Appendix 1

Conversion of cluster-randomized designs to our setting

We should caution those readers who are familiar with cluster-randomized trials (CRT) that while our design has level i =unit (denoted as cluster in those papers) and j =time, it does not have a level k nested within i and j as our examples do not have data down to such a level. This differs from cluster-randomized SW designs,^{12,25–27} which do have a level k (person-visit measure) from m randomly chosen independent patients of cluster i at time j . In such cluster-randomized designs, Y_{ijk} denotes the outcome of the k th person of cluster i from time j ; however, $\bar{Y}_{ij}$ the average outcome of all m persons at i, j is the functional outcome used in those papers and corresponds to Y_{ij} used here.

Further our notation for ρ is for within i intra-class correlation of Y_{ij} and $Y_{ij'}$, which would correspond to the intra-class correlation of $\bar{Y}_{ij}$ $\bar{Y}_{ij'}$. This differs from the intra-class correlation used in cluster-randomized stepped-wedge designs taken down to the nested level k which are for the correlation of repeat measures Y_{ijk} and $Y_{ijk'}$ within the same i , because our response here is a single measure (Y_{ij}) as opposed to the average of m independent observations for a cluster i at time j ($\bar{Y}_{ij}$). However, if $\tilde{\rho}$ denotes the “ ρ ” used in those “cluster-randomization notation” papers for intra-class correlation of Y_{ijk} and $Y_{ijk'}$,

then $\frac{\tilde{\rho}}{\tilde{\rho} + \frac{1-\tilde{\rho}}{m}}$ will be the intra-class correlation of $\bar{Y}_{ij}, \bar{Y}_{ij}'$ from those designs. Thus, to apply our formulas to such cluster-randomized trials, just substitute for ρ in our given formula

$\frac{\tilde{\rho}}{\tilde{\rho} + \frac{1-\tilde{\rho}}{m}}$ where $\tilde{\rho}$ is the “ ρ ” in those cluster-randomized design papers.

Appendix 2

Orthogonal coding for intervention effect

With the orthogonal coding for intervention effect shown below, we are able to simplify the GLS component and calculate the variance of $\hat{\theta}$.

$$Z_1^* = \left(\underbrace{0, \dots, 0}_{t_0}, \underbrace{\frac{N-n_1}{N}, \dots, \frac{N-n_1}{N}}_{t_1}, \underbrace{\frac{N-\sum_{s=1}^2 n_s}{N}, \dots, \frac{N-\sum_{s=1}^2 n_s}{N}}_{t_2}, \dots, \underbrace{\frac{N-\sum_{s=1}^{S-1} n_s}{N}, \dots, \frac{N-\sum_{s=1}^{S-1} n_s}{N}}_{t_S} \right)$$

Table 3

Summary of orthogonalized coding of Z_s^* ($s = 1, \dots, S$) at each step.

Transition stratum	Ordered time periods that timepoint j falls into					
	$(1, j_1)$ contains t_0 timepoints	$(j_1 + 1, j_2)$ contains t_1 timepoints	$(j_2 + 1, j_3)$ contains t_2 timepoints	...	$(j_{S-2} + 1, j_{S-1})$ contains t_{S-2} timepoints	$(j_{S-1} + 1, j_S)$ contains t_{S-1} timepoints
n_S units in S th stratum/step	0	$-\frac{n_1}{N}$	$-\sum_{s=1}^2 n_s$	...	$-\frac{\sum_{s=1}^{S-2} n_s}{N}$	$-\frac{\sum_{s=1}^{S-1} n_s}{N}$
n_{S-1} units in $S-1$ th stratum/step	0	$-\frac{n_1}{N}$	$-\sum_{s=1}^2 n_s$	...	$-\frac{\sum_{s=1}^{S-2} n_s}{N}$	$\frac{N - \sum_{s=1}^{S-1} n_s}{N}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_2 units in 2nd stratum/step	0	$-\frac{n_1}{N}$	$\frac{N - \sum_{s=1}^2 n_s}{N}$	...	$\frac{N - \sum_{s=1}^{S-2} n_s}{N}$	$\frac{N - \sum_{s=1}^{S-1} n_s}{N}$
n_1 units in 1st stratum/step	0	$\frac{N - n_1}{N}$	$\frac{N - \sum_{s=1}^2 n_s}{N}$	...	$\frac{N - \sum_{s=1}^{S-2} n_s}{N}$	$\frac{N - \sum_{s=1}^{S-1} n_s}{N}$

$$\begin{aligned}
Z_2^* &= \left(\underbrace{0, \dots, 0}_{t_0}, \underbrace{-\frac{n_1}{N}, \dots, -\frac{n_1}{N}}_{t_1}, \underbrace{\frac{N - \sum_{s=1}^2 n_s}{N}, \dots, \frac{N - \sum_{s=1}^2 n_s}{N}}_{t_2}, \dots, \underbrace{\frac{N - \sum_{s=1}^{S-1} n_s}{N}, \dots, \frac{N - \sum_{s=1}^{S-1} n_s}{N}}_{t_s} \right) \\
&\dots \\
Z_S^* &= \left(\underbrace{0, \dots, 0}_{t_0}, \underbrace{-\frac{n_1}{N}, \dots, -\frac{n_1}{N}}_{t_1}, \underbrace{-\frac{\sum_{s=1}^2 n_s}{N}, \dots, -\frac{\sum_{s=1}^2 n_s}{N}}_{t_2}, \dots, \underbrace{-\frac{\sum_{s=1}^{S-1} n_s}{N}, \dots, -\frac{\sum_{s=1}^{S-1} n_s}{N}}_{t_s} \right) \\
Z^* &= (\underbrace{Z_1^*, \dots, Z_1^*}_{n_1}, \underbrace{Z_2^*, \dots, Z_2^*}_{n_2}, \dots, \underbrace{Z_S^*, \dots, Z_S^*}_{n_S}).
\end{aligned}$$

The advantage of this orthogonal coding is to set cross-products that involve the time periods and the intervention in $X' V^{-1} X$ to 0, which makes it easier to invert the cross-product matrix for the purpose of finding the variance of the intervention effect estimate as is done in Appendix 3.

Appendix 3

Derivation of $\text{Var}(\hat{\theta})$

The goal is to find the last element of $(X' V^{-1} X)^{-1} \sigma^2$ which is $\text{Var}(\hat{\theta})$. Because of the orthogonal coding, we are able to compute $(X' V^{-1} X)^{-1}$ by applying the inverse of partitioned matrix twice.

$$X' V^{-1} X = \begin{bmatrix} \frac{n_1 T}{[1+(T-1)\rho]} & \dots & 0 & 0 & \dots & 0 & I'_{\{i \in A_1\}} V^{-1} Z^* \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \frac{n_S T}{[1+(T-1)\rho]} & 0 & \dots & 0 & I'_{\{i \in A_S\}} V^{-1} Z^* \\ 0 & \dots & 0 & \frac{2N}{(1-\rho)} & \dots & \frac{2N}{(1-\rho)} & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \frac{N}{(1-\rho)} & \dots & \frac{2N}{(1-\rho)} & 0 \\ I'_{\{i \in A_1\}} V^{-1} Z^* & \dots & I'_{\{i \in A_S\}} V^{-1} Z^* & 0 & \dots & 0 & \text{OTD} \end{bmatrix}$$

For the first inverse of portioned matrix, $(X' V^{-1} X)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ where

$$A_{11} = \begin{bmatrix} \frac{n_1 T}{[1+(T-1)\rho]} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{n_S T}{[1+(T-1)\rho]} \end{bmatrix}$$

$$A_{12}=A'_{21}=\begin{bmatrix} 0 & \cdots & 0 & I'_{\{i \in A_1\}} V^{-1} Z^* \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & I'_{\{i \in A_S\}} V^{-1} Z^* \end{bmatrix},$$

and

$$A_{22}=\begin{bmatrix} \frac{2N}{(1-\rho)} & \cdots & \frac{N}{(1-\rho)} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{N}{(1-\rho)} & \cdots & \frac{2N}{(1-\rho)} & 0 \\ 0 & \cdots & 0 & \text{OTD} \end{bmatrix}$$

We then apply the inverse of the partitioned matrix to B_{22} as below.

$$B_{22}=(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}=\begin{bmatrix} \frac{2N}{(1-\rho)} & \cdots & \frac{N}{(1-\rho)} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{N}{(1-\rho)} & \cdots & \frac{2N}{(1-\rho)} & 0 \\ 0 & \cdots & 0 & \text{OTD-IDP} \end{bmatrix}^{-1}$$

For the “nested” inverse of portioned matrix,

$$B_{22}=\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1}=\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

where

$$C_{11}=\begin{bmatrix} \frac{2N}{(1-\rho)} & \cdots & \frac{N}{(1-\rho)} \\ \vdots & \ddots & \vdots \\ \frac{N}{(1-\rho)} & \cdots & \frac{2N}{(1-\rho)} \end{bmatrix}$$

$$C_{12}=C'_{21}=\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

and

$$C_{22}=[\text{OTD-IDP}]$$

It turns out that

$$D_{12}=D'_{21}=\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Therefore,

$$\text{Var}(\hat{\theta})=D_{22}\sigma^2=(C_{22}-C_{21}C_{11}^{-1}C_{12})^{-1}\sigma^2=\frac{1}{OTD-IDP}\sigma^2$$

Where the full expansions of OTD, IDP and OTD-IDP are

$$OTD=\frac{1}{(1-\rho)}\sum_{l=1}^{S-1}t_l\frac{\left(\sum_{h=1}^ln_h\right)\left(N-\sum_{h=1}^ln_h\right)}{N}-\frac{\rho}{[1+(T-1)\rho](1-\rho)}\times\sum_{s=1}^Sn_s\left[\sum_{l=1}^{S-1}t_l\frac{N-\sum_{h=1}^ln_h}{N}-\sum_{l=1}^{S-1}t_lI_{\{s>1\}}\right]^2$$

$$IDP=\sum_{s=1}^S\frac{[1+(T-1)\rho]}{n_sT}(X'_{\alpha_s}V^{-1}X_{Tx})^2=\frac{1}{[1+(T-1)\rho]T}\times\sum_{s=1}^Sn_s\left[\sum_{l=1}^{S-1}t_l\frac{N-\sum_{h=1}^ln_h}{N}-\sum_{l=1}^{s-1}t_lI_{\{s>1\}}\right]^2$$

$$OTD-IDP=\frac{1}{(1-\rho)}\left\{\sum_{l=1}^{S-1}t_l\frac{\left(\sum_{h=1}^ln_h\right)\left(N-\sum_{h=1}^ln_h\right)}{N}-\frac{1}{T}\times\sum_{s=1}^Sn_s\left[\sum_{l=1}^{S-1}t_l\frac{N-\sum_{h=1}^ln_h}{N}-\sum_{h=1}^{s-1}t_h\right]^2\right\}$$

Note that in the R-SW design from equation (3), IDP=0 and

$$X'V^{-1}X=\begin{bmatrix} \frac{NT}{[1+(T-1)\rho]} & 0 & \cdots & 0 & 0 \\ 0 & \frac{2N}{(1-\rho)} & \cdots & \frac{N}{(1-\rho)} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \frac{N}{(1-\rho)} & \cdots & \frac{2N}{(1-\rho)} & 0 \\ 0 & 0 & \cdots & 0 & OTD \end{bmatrix}$$

$$\text{Therefore, } \text{Var}(\hat{\theta}_{R-SW})=\frac{1}{OTD}\sigma^2.$$

Appendix 4

Variance of $\hat{\theta}_{NR-SW}$ in fixed effects model is invariant to absorption from σ^2 into strata α_s dispersion

For the randomized model in equation (3)

$$Y_{ij} = \mu_i + \beta_j^* + \theta Z_{ij}^* + \varepsilon_{ij}$$

where the mean displacement of unit i is $\mu_i \sim N(\alpha_0, \tau^2)$, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. In this R-SW design, the total variance decomposition $\sigma_R^2 = \tau^2 + \sigma_e^2$, and the correlation of repeated measures from the same unit is $\rho_R = \frac{\tau^2}{\tau^2 + \sigma_e^2}$, thus $1 - \rho_R = \frac{\sigma_e^2}{\tau^2 + \sigma_e^2}$.

A non-randomized stepped-wedge (NR-SW) to the same setting can be presented in equation (4) as

$$Y_{ij} = \alpha_s + (\mu_i - \alpha_s) + \beta_j^* + \theta Z_{ij}^* + \varepsilon_{ij}$$

where $\alpha_s \sim N(\alpha_0, \sigma_s^2)$, $(\mu_i - \alpha_s) \sim N(0, \tau^2 - \sigma_s^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_e^2)$; α_s , μ_i are independent. The variance with μ_i treated as random effects (RE) decomposes into $\sigma_{RE-NR}^2 = \sigma_s^2 + (\tau^2 - \sigma_s^2) + \sigma_e^2 = \tau^2 + \sigma_e^2$ (which is the same total variance as σ_R^2). Now the within unit factor μ_i has been decomposed into “stratum mean” and “unit mean about stratum mean” random factors $\alpha_s + (\mu_i - \alpha_s)$. When fitting a random effects model to the

NR-SW design, the within unit repeated measure correlation is $\rho_{RE-NR} = \frac{\tau^2}{\tau^2 + \sigma_e^2}$ (the same value as ρ_R), but there is also a within strata correlation of measures of different units in the same stratum of $\rho_s = \frac{\sigma_s^2}{\tau^2 + \sigma_e^2}$.

However, for the NR-SW fixed effects (FE) model, α_s the strata effects are treated as fixed eliminating σ_s^2 from both the overall variance and the within unit correlation. The within unit factor now is $(\mu_i - \alpha_s)$. Thus, for the NR-SW fixed effects model applied to this setting, the overall variance of an observation is $\sigma_{FE-NR}^2 = (\tau^2 - \sigma_s^2) + \sigma_e^2$ and the within unit mean

variance correlation is $\rho_{FE-NR} = \frac{\tau^2 - \sigma_s^2}{\sigma_{FE-NR}^2} = \frac{\tau^2 - \sigma_s^2}{(\tau^2 - \sigma_s^2) + \sigma_e^2}$, and

$$(1 - \rho_{FE-NR}) = \frac{\sigma_e^2}{(\tau^2 - \sigma_s^2) + \sigma_e^2}.$$

This leads to a potential awkwardness in Tables 1 and 2, where different NR-SW fixed effects designs are compared to each other and to the R-SW and NR-SW random effects designs in that both the variance and the intra-class correlations are changed from $\sigma^2 = \sigma_R^2$ and $\rho = \rho_R$ to $\sigma^2 = \sigma_{RE-NR}^2$ and $\rho = \rho_{RE-NR}$ by the absorption of variance into the stratum specific intercepts in the NR-SW designs. However, $Var(\hat{\theta}_{NR-SW})$ in equation (8) and $Var(\hat{\theta}_{BNR-SW})$ in equation (12) for fixed effects NR-SW only depend on σ^2 and ρ through the product $(1 - \rho)\sigma^2$. To that end, this product is unchanged by application of the fixed effects NR-SW design in that $(1 - \rho_{FE-NR})\sigma_{FE-NR}^2 = (1 - \rho_R)\sigma_R^2 = \sigma_e^2$. Thus, the parameters for σ^2 and ρ from the “pre-non-randomized” study design population can be used in

equation (8) and equation (12) no matter what the impact of the fixed effects NR-SW on the final σ^2 and ρ .

A special case of this principle is when $n_s = 1$ for all s . In this case, μ_j and α_s are unidentifiable and the maximization process sets $\mu_j = \alpha_s$ which results in $\sigma_s^2 = \tau^2$ or the estimated target $\sigma_{FE-NR}^2 = \sigma_e^2$; $\rho_{FE-NR} = 0$. But again,

$(1 - \rho_{FE-NR})\sigma_{FE-NR}^2 = (1 - \rho_R)\sigma_R^2 = \sigma_e^2$ so equation (12) can be applied with normative parameters from the randomized design.

Figure 1.
Overview for general stepped-wedge designs.

Figure 2.
Ratio of $\text{Var}(\text{R-SW})/\text{Var}(\text{NR-SW})$ with optimal $t^*=1$.

Figure 3.
Ratio of variance for fully balanced NR-SW to optimal NR-DD (both fixed effect designs).

Table 1

Minimal detectable effect size (δ_1) for $T = 6$ in a Study of 30 LTCF for non-randomized design.

$T = 6$	BNR-SW(fixed effects)		NR-DD ^a $b^* = \frac{T}{2}$
	$t = 1 (S = 5)$	$t = 2 (S = 2)$	
$\rho = 0.85^b$	0.290	0.343	0.323
$\rho = 0.50^c$	0.529	0.626	0.590
$\rho = 0.30$	0.626	0.741	0.699
$\rho = 0.00$	0.749	0.886	0.835

^aFixed effects by default as a random effects model on a difference-in-differences design is structurally biased.

^bThis ρ was observed for incontinence over seven quarters at New Jersey LTCF.

^cThis ρ was observed for ulcers over seven quarters at New Jersey LTCF.

Table 2
Minimal detectable effect size (δ_1) for $T = 9$ in a study of 30 LTCF in balanced stepped-wedge designs.

Non-randomized step wedge						
$T = 6N = 30$	t	Analyzed using random effects				
		Analyzed using fixed effects	Analyzed using random effects			Randomized step wedge
			$\rho_s = \rho$	$\rho_s = 0.25\rho$	$\rho_s = 0.1\rho$	
$\rho = 0.85^1$	$t^* = 1$	0.290	0.289	0.289	0.288	0.287
	$t = 2$	0.343	0.343	0.343	0.342	0.341
$\rho = 0.50^2$	$t^* = 1$	0.529	0.524	0.517	0.511	0.504
	$t = 2$	0.626	0.625	0.621	0.616	0.605
$\rho = 0.30$	$t^* = 1$	0.626	0.613	0.596	0.584	0.572
	$t = 2$	0.741	0.727	0.727	0.717	0.694
$\rho = 0$	$t^* = 1$	0.749	NA	NA	NA	0.572
	$t = 2$	0.886	NA	NA	NA	0.723