



HHS Public Access

Author manuscript

J Clin Epidemiol. Author manuscript; available in PMC 2017 July 19.

Published in final edited form as:

J Clin Epidemiol. 2017 March ; 83: 108–109. doi:10.1016/j.jclinepi.2016.12.011.

“Cross-sectional” stepped wedge designs always reduce the required sample size when there is no time effect

Xin Zhou*,

Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, 677
Huntington Avenue, Boston, MA 02115, USA

Xiaomei Liao, and

AbbVie Inc., Data and Statistical Sciences, 1 North Waukegan Road, North Chicago, IL 60064,
USA

Donna Spiegelman

Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, 677
Huntington Avenue, Boston, MA 02115, USA

To the Editor

A recent article [1] suggested that the stepped wedge cluster-randomized trial (SWD) is more efficient than a parallel cluster-randomized trial (pCRD). Then, a subsequent letter [2] pointed out that the article [1] incorrectly applied the variance for a cross-sectional SWD to that corresponding to a cohort design and provided a counter example where a cross-sectional SWD requires a larger sample size than a pCRD. To review terms, in a “cross-sectional” SWD, only one measurement of the outcome is taken per subject and new subjects enroll at each new step, whereas in a “cohort” SWD, subjects enroll at the beginning of the study and have repeated measures taken over different steps. The authors of the original article [1] then revised their conclusion in a follow-up letter [3], agreeing that a “cross-sectional” SWD is not always more powerful than a pCRD. The conclusion was drawn from a variance formula which takes into account time effects, as given by Hussey and Hughes [4]. However, time effects, if they exist, do not need to be considered in a pCRD because randomization will, on average, achieve balance between the intervention groups with respect to such effects. Hence, a comparison between an SWD which incorporates time effects and a pCRD which does not might be unfair.

Many SWDs are conducted over a relatively short period, where, even if there were some modest background time trends operating over the long term, they would likely have little or no effect on outcome rates during the relatively short study period. For example, it is unlikely that the 1-year retention and 6-month viral suppression rates will change in any materially important way during the 3-year study period of the MaxART Early Access to ART in Swaziland SWD currently underway [5]. Similarly, 3 years was the duration of one

*Corresponding author. Tel.: 1-617-432-7991. stxzh@channing.harvard.edu (X. Zhou).

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2016.12.011>.

of the largest and most well-known examples of the use of the SWD design, the evaluation of former Harvard School of Public Health's Dean Julio Frenk's PROGRESA (Programa de Educación, Salud, y Alimentación) program, launched when Frenk served as Minister of Health for Mexico. PROGRESA was a conditional cash transfer program targeting the country's poorest families, providing them with financial incentives for uptake of recommended public health and nutrition practices and for keeping children in school, which was evaluated through a staggered roll-out of these policies among 495 communities [6]. Finally, soon to complete in Nepal, Tanzania, and Sri Lanka, an SWD to assess the effectiveness of postpartum intrauterine device insertion as a safe and acceptable means of preventing unintended pregnancy closely following a current birth, the FIGO study, is running for a 1-year enrollment period of a projected 64,000 pregnant women, each with 1.5 years of follow-up [7].

Correlations between intervention status and time are high in SWDs, for example, in FIGO, the empirical correlation between the intervention and time is 0.67, and in MaxART, it is 0.69. Considering a standard SWD with T steps as in Hussey and Hughes [4], it is shown in

the Appendix at www.jclinepi.com to this letter that the correlation is $\sqrt{\frac{T+1}{3(T-1)}}$, which is always greater than 0.58. Hence, time trends need to be very small, or the results discussed in the following will not apply, as confounding will otherwise lead to biased estimates.

Following the same notation as Hussey and Hughes [4], we consider a "cross-sectional" SWD with I clusters, T steps, and N participants sampled per cluster per step. Note that $N \times T$ is the cluster size, that is, the number of participants in each cluster, and $N \times T \times I$ is the total sample size of the study. Here, we provide a simplified version of the $Var(\hat{\theta})$ and the design effect for SWDs with a time effect and T steps fit as discrete time effects in Hussey and Hughes' model [4]. Simplifying Equation (8) in Hussey and Hughes, as shown in the Appendix at www.jclinepi.com to this letter,

$$Var(\hat{\theta}) = \frac{4 Var(Y)}{NTI} \frac{(T-1)(1+(NT-1)\rho)}{(T-2) \left(\frac{2}{3} + \frac{1}{3} N(T+1) \frac{\rho}{1-\rho} \right)}, \quad (1)$$

and its corresponding design effect is

$$DE_{sw}^* = \frac{(T-1)(1+(NT-1)\rho)}{(T-2) \left(\frac{2}{3} + \frac{1}{3} N(T+1) \frac{\rho}{1-\rho} \right)}.$$

Above, θ is the treatment effect, Y is the continuous outcome with $Var(Y) = \sigma_e^2 + \tau^2$, where σ_e^2 and τ^2 are the within- and between- cluster variances, respectively, and $\rho = \tau^2 / (\sigma_e^2 + \tau^2)$ is the intraclass correlation (ICC). Liao et al. [8] derived a variance formula for the estimated intervention effect in an SWD that applies when time is unrelated to the outcome and thus can be ignored in the analysis. Using the same notation as mentioned previously, but

simplifying Liao et al.'s formula, as shown in the Appendix at www.jclinepi.com to this letter,

$$\text{Var}(\tilde{\theta}) = \frac{4\text{Var}(Y)}{NTI} \frac{1+(NT-1)\rho}{1+\frac{2}{3}N(T+1)\frac{\rho}{1-\rho}}, \quad (2)$$

where the tilde is used to distinguish this estimator from the previous one in Equation (1). Thus, the design effect for a cross-sectional SWD with no time effect is

$$DE_{SW} = \frac{1+(NT-1)\rho}{1+\frac{2}{3}N(T+1)\frac{\rho}{1-\rho}}.$$

Noting that the design effect of a pCRD is $DE_P = 1 + (NT - 1) \cdot \rho$, for fixed cluster size NT , it is easy to see that $DE_{SW} < DE_P$. Thus, when there is no time effect, the SWD always requires smaller sample size. We used these design effect formulas in the following illustrative example.

For example, consider a study that would require 950 participants under individual randomization. Next, consider several cluster-randomized alternatives, with an ICC of 0.01:

1. Under a pCRD with 100 participants per cluster, the design effect is 1.99. The minimum total sample size required would then be 1,891 participants, which would be distributed across 19 clusters, for a total sample size of 1,900.
2. Under an SWD **with no time effect**, with 100 participants per cluster as mentioned previously and, for example, five steps, there will be 20 participants per cluster per step, leading to a design effect of 1.10 and a minimum total sample size of 1,045. Because the number of clusters needs to be a multiple of 4, with five steps, this design will have a minimum of 12 clusters, rounding up to a total sample size of 1,200.
3. Under an SWD **with time effects**, with five steps and 20 participants per cluster per step as mentioned previously, the design effect is 2.48, using the formula as mentioned previously, a simplification of Hussey and Hughes [4]. The minimum total sample size will then be 2,356, and rounding up to obtain a number of clusters which is a multiple of 4, there will be a minimum of 24 clusters, and a total sample size is 2,400.

In conclusion, when there is no effect of time on the outcome, the SWD is always more efficient than the pCRD. This is most likely to occur in SWDs of short duration, as is often the case, but caution is needed as the correlation between time and intervention is typically high in SWDs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This work was supported by the grants NIH/R01AI112339 and NIH/DPES025459.

References

1. Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol.* 2013; 66:752–8. [PubMed: 23523551]
2. Hemming K, Girling A. The efficiency of stepped wedge vs. cluster randomized trials: stepped wedge studies do not always require a smaller sample size. *J Clin Epidemiol.* 2013; 66:1427–8. [PubMed: 24035495]
3. de Hoop E, Woertman W, Teerenstra S. The stepped wedge cluster randomized trial always requires fewer clusters but not always fewer measurements, that is, participants than a parallel cluster randomized trial in a cross-sectional design. In Reply *J Clin Epidemiol.* 2013; 66:1428. [PubMed: 24021612]
4. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials.* 2007; 28:182–91. [PubMed: 16829207]
5. Walsh F, Bärnighausen T, Delva W, Fleming Y, Khumalo G, Lejeune C, et al. Impact of early initiation versus national standard of care of antiretroviral therapy in Swaziland’s public sector health system: a stepped-wedge randomized implementation study. 2016 Under review.
6. Paul Schultz T. School subsidies for the poor: evaluating the Mexican Progresa poverty program. *J Development Econ.* 2004; 74(1):199–250.
7. Canning D, Shah I, Pearson E, Pradhan E, Karra M, Senderowicz L, et al. Institutionalizing postpartum intrauterine device (IUD) services in Sri Lanka, Tanzania, and Nepal: study protocol for a cluster-randomized stepped-wedge trial. *BMC Pregnancy Childbirth.* 2016; 16:362. [PubMed: 27871269]
8. Liao XM, Zhou X, Spiegelman D. A note on “Design and analysis of stepped wedge cluster randomized trials”. *Contemp Clin Trials.* 2015; 45:338–9. [PubMed: 26394003]