

Inclusion of Misanalyzed Stepped Wedge Trials in Meta-Analysis: Findings from a Simulation Study

Santos Joseph Alvin Ramos⁽¹⁾, Di Tanna Gian Luca⁽¹⁾

(1) Department of Business Economics, Health and Social Care (DEASS), University of Applied Sciences and Arts of Southern Switzerland (SUPSI)

CORRESPONDING AUTHOR: Santos Joseph Alvin Ramos, joseph.santos@supsi.ch

INTRODUCTION

The use of the stepped wedge cluster randomized trial (SWT) design to assess the effect of interventions in real-world settings has gained considerable popularity in recent years [1]. In this design, all clusters begin in the control condition, and the intervention is sequentially and randomly rolled out until all clusters transition into the intervention condition [2]. The unique characteristics of this design—for example, the unidirectional crossover to the intervention, variation in the timing and duration of exposure across clusters, and repeated measurements within clusters—pose additional statistical challenges that must be addressed when defining models to evaluate the effects of interventions [3].

In the context of meta-analysis, the inclusion of outcome data from SWTs requires careful methodological consideration. While analytic approaches for incorporating data from parallel cluster randomized trials have been well-documented [4], these methods are frequently misapplied in practice, with many analyses failing to properly account for clustering [5]. Given the added complexity of SWT designs, the potential for analytical errors, such as model misspecification [6], is expected to be higher. Currently, no established methods exist for synthesizing evidence from SWTs within meta-analyses, underscoring an important methodological gap that warrants attention.

OBJECTIVES

This study aims to examine the effects of including misanalyzed SWT in meta-analysis through a series of simulations.

METHODS

RCT and SWT datasets were simulated. RCT datasets were generated with a balanced two-arm design (treatment and control), assuming a normally distributed treatment effect with a mean difference of 0 and a fixed sample size of 1000

participants per study. SWT datasets were generated based on a repeated cross-sectional design with 5 time points, 50 observations per cluster, an average treatment effect of 0, an error variance (σ^2) of 5, and a random cluster effect (τ^2) of 1. Eighteen data-generating scenarios were considered, varying in number of clusters (20, 40 or 60), random treatment effect ($\eta^2 = 1, 2$ or 3), and random time effect (absent ($\gamma^2 = 0$) or present ($\gamma^2 = 1$)). For simplicity, no correlation between random effects was included.

Each SWT dataset was analyzed using three linear mixed-effects models:

1. Unadjusted for time: $y \sim \text{treatment} + (1 \mid \text{cluster})$
2. Hussey and Hughes model: $y \sim \text{treatment} + \text{time} + (1 \mid \text{cluster})$
3. Extended model accounting for random treatment and random time effects, where appropriate: $y \sim \text{treatment} + \text{time} + (1 \mid \text{cluster}) + (0 + \text{treatment} \mid \text{cluster}) + (1 \mid \text{clustertime})$

For each model, the estimated treatment effect and standard error were extracted. Meta-analytic datasets were then constructed, each consisting of 3 RCTs and 1 SWT. To enable direct comparison across the three models, each SWT dataset was included in three separate but matched meta-analytic datasets per scenario. For each scenario, random-effects meta-analysis was done using the Sidik-Jonkman estimator of between-study heterogeneity. Model performance was assessed by calculating the mean of the pooled effect sizes, along with bias, model-based standard error, coverage, and the percentage of statistically significant results out of 1000 pooled effect sizes at $p < 0.05$.

RESULTS

The model unadjusted for time consistently yielded the highest percentage of statistically significant results (based on 1000 meta-analyses per scenario), followed by the Hussey and Hughes and the extended model (Figure 1). Out of 18 scenarios, the unadjusted model exceeded the alpha threshold of 5% in 9 scenarios (50%), whereas neither the

Hussey and Hughes model nor the extended model exceeded this threshold in any scenario. Compared to the extended model—considered the correctly specified model in the simulations—the model unadjusted for time yielded, on average, 3.5% more statistically significant results with a maximum difference of 5.8%. The Hussey and Hughes model produced a smaller difference, on average, 1.2% more statistically significant results with a maximum difference of 2.2%.

In terms of coverage, the extended model produced the highest coverage probabilities, while the unadjusted model yielded the lowest. In contrast, the pooled effect size and associated bias averaged over all iterations per scenario were 0 (Monte Carlo standard error ranging from 0 to 0.01), consistent with the treatment effect set during the data generation process.

CONCLUSIONS

The inclusion of SWT data analyzed using misspecified models in meta-analyses can lead to inflated false-positive findings and potentially misleading conclusions about the effect of interventions. Researchers doing meta-analysis that include SWTs should exercise caution in evaluating the appropriateness of the underlying analytic methods to ensure valid and reliable inferences.

REFERENCES

1. Lung T, Si L, Hooper R, et al. Health Economic Evaluation Alongside Stepped Wedge Trials: A Methodological Systematic Review. *Pharmacoeconomics*. 2021;39(1):63-80.
2. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*. 2007;28(2):182-91.
3. Li F, Hughes JP, Hemming K, et al. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*. 2021;30(2):612-39.
4. Higgins JPT, Eldridge S, Li T (editors). Chapter 23: Including variants on randomized trials. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor. *Cochrane Handbook for Systematic Reviews of Interventions* version 65 (updated August 2024): Cochrane; 2024.
5. Santos JAR, Riggi E, Di Tanna GL. Warnings on the inclusion of cluster randomized trials in meta-analysis: results of a simulation study. *BMC Medical Research Methodology*. 2025;25(1):133.
6. Voldal EC, Xia F, Kenny A, et al. Model misspecification in stepped wedge trials: Random effects for time or treatment. *Stat Med*. 2022;41(10):1751-66.

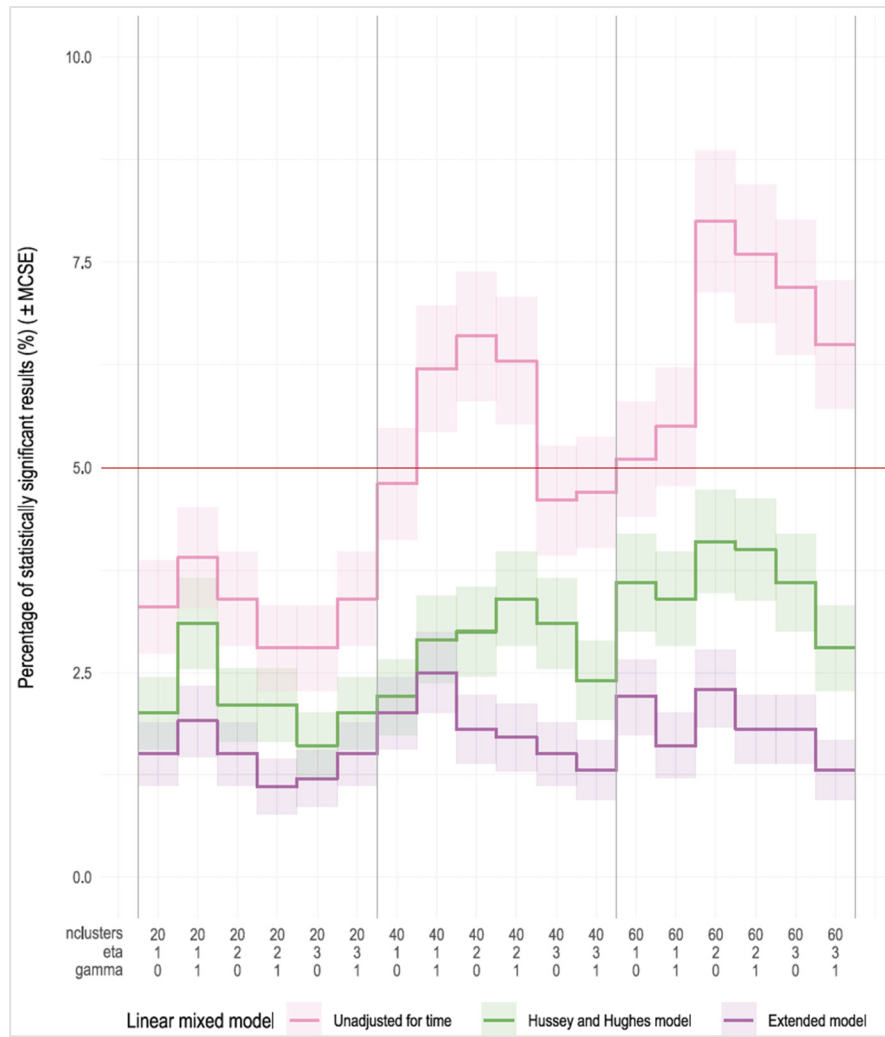


Figure 1. Percentage of statistically significant results across scenarios