

HANDLING OF COVARIATES IN STEPPED-WEDGE CLUSTER RANDOMIZED TRIALS: PROTOCOL FOR A METHODOLOGICAL REVIEW

Team (publication order to be determined): Pascale Nevins,^a Kendra Davis-Plourde,^b Jules Pereira Macedo,^c Yongdong Ouyang,^d Can Meng,^e Mary Ryan,^f Guangyu Tong,^{f,g} Xueqi Wang,^{f,h} Luis Ortiz-Reyes,^a Agnès Caille,^{c,i} Fan Li,^{f,g} Monica Taljaard.^{a,d}

^a Ottawa Hospital Research Institute, Clinical Epidemiology Program, Centre for Implementation Research, The Ottawa Hospital, Ottawa, ON, K1H 8L6, Canada.

^b Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA.

^c Université de Tours, Université de Nantes, INSERM, SPHERE U1246, Tours, France.

^d School of Epidemiology and Public Health, University of Ottawa, 600 Peter Morand Crescent, Room 101, Ottawa, Ontario, K1G 5Z3, Canada.

^e Yale Center for Analytical Sciences, Yale School of Public Health, New Haven, Connecticut, USA.

^f Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA.

^g Center for Methods in Implementation and Prevention Science, Yale University, New Haven, Connecticut, USA.

^h Section of Geriatrics, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA.

ⁱ INSERM CIC 1415, CHRU de Tours, Tours, France.

Word Count: 3918

Keywords: methodological review, stepped-wedge cluster randomized trials, covariate adjustment, baseline balance, restricted randomization

Corresponding Author

Monica Taljaard

Clinical Epidemiology Program

Ottawa Hospital, Civic Campus

1053 Carling Avenue, Civic Box 693

Admin Services Building, ASB 2-004

Ottawa, ON K1Y 4E9

613-798-5555 x18618 phone

Email: mtaljaard@ohri.ca

ABSTRACT

Background: Stepped-wedge cluster randomized trials (SW-CRTs), which randomize clusters to times at which they transition from control to intervention arms, are an increasingly popular design for evaluating health system and policy interventions. Reporting of SW-CRTs should ideally be guided by the 2018 CONSORT extension for SW-CRTs. Although substantial methodological development has taken place in recent years, several gaps remain in our knowledge about the best practices for the design, analysis, and reporting of SW-CRTs.

Objectives: We are undertaking a methodological systematic review of SW-CRTs published 2016-2022. Our primary objectives are to describe current methods and practices for handling covariates (number of and types of covariates) in the design and analysis of SW-CRTs; examine adherence to CONSORT guidelines around reporting of covariates including balance at baseline; and identify prevalent issues in SW-CRTs and gaps in the methodological literature.

Methods: We will identify primary reports of SW-CRTs published in English 2016-2022 from three sources: a previously published review covering 2020 to 2021, an existing database of pragmatic trials published 2014-2019, and an updated search capturing trials up to 2022. A data extraction form will be used to standardize the extraction of information about the trials.

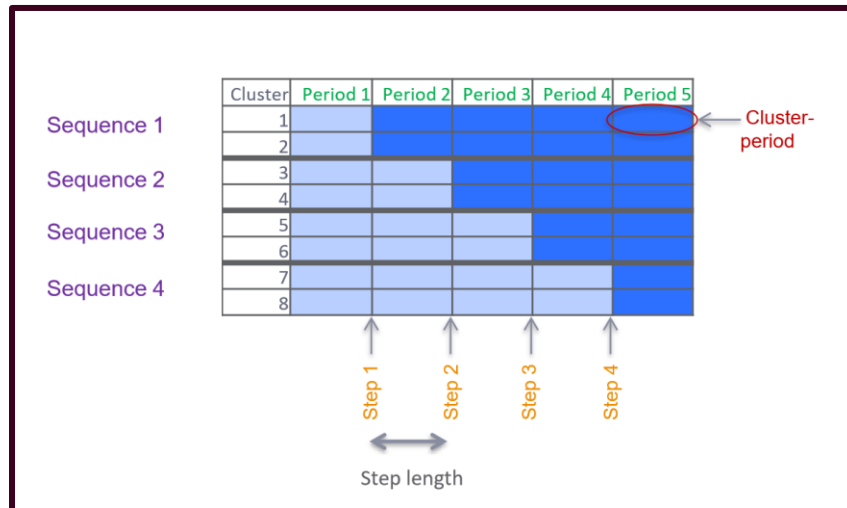
Analyses will be primarily descriptive in nature.

BACKGROUND, RATIONALE, AND OBJECTIVES

The defining feature of a cluster randomized trial (CRT) is that the units of randomization are intact groups such as medical practices, hospitals or even entire communities.¹ CRTs are preferred over individually randomized designs when interventions are delivered at the cluster level or when there is a substantial risk of contamination due to individuals in the same cluster allocated to different arms interacting.² Importantly, in this design, outcomes from participants within the same cluster are more similar than outcomes obtained from participants in different clusters: this reduces the effective sample size and thus, a CRT always requires a larger sample size than an individually randomized trial.¹

In recent years, a novel type of CRT design called a stepped-wedge cluster randomized trial (SW-CRT) has been gaining in popularity.³ Unlike parallel arm CRTs, in which some clusters are randomly allocated to the treatment and the rest to the control, in SW-CRTs all clusters typically receive both control and intervention conditions (see Figure 1). Clusters begin in the control condition and are randomized to a particular “sequence” which determines the time at which they will transition to the intervention condition.⁴ Thus, observations are repeatedly collected from each cluster over time, i.e., in both the control and intervention conditions in each “cluster-period”. These assessments may be taken on the same participants across all periods (called a “cohort” design, in which the majority (open-cohort) or all (closed-cohort) participants are identified at baseline and outcomes are repeatedly assessed on the same individuals) or different participants across the periods (a “cross-sectional” design, in which the recruitment of new participants occurs in each period, either at a single point or continuously over time, e.g., as they arrive at a clinical location).⁵

Figure 1. Diagram of a typical stepped-wedge cluster randomized trial roll out. Each of eight clusters is randomly assigned to one of the four “sequences” which determines the time (the “step”) at which it will cross over from control (light blue) to intervention (dark blue) condition. The trial timeline is delineated into equal-length blocks of time called periods. Here, all clusters begin in the control condition and end in the intervention condition although this is not required.



The SW-CRT design is often chosen over a parallel CRT as it is believed to facilitate cluster recruitment by allowing eventual intervention exposure in all clusters, when it is viewed as beneficial to expose all clusters to the intervention, or for logistical reasons such as the practicality of staggered implementation across the clusters. However, the need to carefully adhere to the scheduled timing of implementation, along with the need for complex statistical analyses, can raise challenges.^{3,3,6,7} Key methodological considerations for SW-CRTs include that the analysis must always account for a time effect since the intervention is confounded with time, and complex correlation structures.^{4,8} As in a parallel arm CRT, within-cluster correlation (the similarity between participants in the same cluster) must be taken into account in SW-CRTs. An additional factor to be considered in SW-CRTs is the possibility of changes in the strength of the correlation between participant outcomes over the course of different periods.⁹ A variety of

methods of analysis have been used to account for these complexities, including in generalized linear mixed-models (GLMM) and Generalized Estimating Equations (GEE).¹⁰ In addition to the fixed time effect, these models may include additional effects as covariates, motivated by the desire to control for potential confounding or to improve power or efficiency. These covariates should include those used in any form of restricted randomization; the choice of all covariates should be determined *a priori*.

The CONSORT Extension for Stepped-Wedge Trials, published in 2018, contains reporting guidelines specific to SW-CRTs.¹¹ In particular, the CONSORT extension emphasizes the clear reporting of key design features such as the number of clusters, sequences, and periods; the timing of outcome assessments and recruitment; and whether the SW-CRT is a cross-sectional or cohort design, as these all have implications for the sample size calculation and analysis. Further, the CONSORT extension notes that the method of random allocation, and whether this occurred at a single time point or multiple, as well as "details of any constrained randomisation or stratification" should be reported.¹¹ It also distinguishes several possible ways of reporting balance on baseline characteristics in SW-CRTs, as balance can be presented by condition, by sequence/cluster, by period, or by some overlap of these three. The definitions of "baseline" may also differ between designs: in a cross-sectional design, it may refer to characteristics of each participant included at their time of presentation to the trial; in cohort designs, it may refer instead to measures taken pre- or during the first control period on all included participants. Per the CONSORT extension, the assumed correlation structures (within-cluster and between-period) and whether time effects were adjusted for should also be reported.

Despite increasing interest in SW-CRTs, methodological gaps remain in designing, conducting, and reporting on this design. Previous reviews have been published on many reporting and

methodological aspects of SW-CRTs, including their quality of reporting,¹²^{Error! Bookmark not defined.}^{13,14} ethical conduct,^{15,16} rationale for use,^{6,17} sample size calculation,^{10,10,12,18} methods surrounding unequal cluster sizes,¹⁸ use of feasibility studies,¹⁹ recruitment and implementation,⁷ and statistical methods of analysis.^{10,14,20} However, few of these have been published since the CONSORT extension's publication in 2018, and to our knowledge, none have examined the use of covariates in the design and analysis of SW-CRTs.

Previous reviews of parallel individual and cluster randomized trials have found marked differences between guidance and use of covariates.^{21,22} In CRTs, methods such as stratification, matching, covariate-constrained randomization, or minimization are commonly used,²³ but no reviews have explicitly examined the extent to which restricted randomization techniques are adopted in SW-CRTs, which often have a limited number of clusters.¹²^{Error! Bookmark not defined.}²⁰ While some methods of allocation have been evaluated or advocated for in SW-CRTs,^{24,25} guidance is not yet widespread for how to best achieve balance at baseline in this design. Though a review by Grayling et al. examined the prevalence of reporting baseline data in SW-CRTs, the ways in which this information was presented in the trials was not discussed.¹³ A previous review noted the variety of ways in which baseline data are presented in a small sample of 10 SW-CRTs before 2014.¹⁴ Further, though the literature has long recommended against significance testing of balance at baseline,²⁶ a review of cluster randomized trials has shown that the practice continues to be prevalent.²⁷ Finally, though reviews have found marked discrepancies in methods of analysis and use of design effects (adjustment for time effects and clustering) in SW-CRTs in the past,^{10,18} information on the use of additional covariates is limited to specific characteristics such as cluster size,¹⁸ or to a small selection of trials.¹⁴ An updated review is needed to inform future methodological guidance.

Motivated by these gaps in the literature, the primary objectives of this review are to:

- 1) Describe current methods to promote balance in the randomization including the specific methods used and the number and types of characteristics balanced in the design of SW-CRTs;
- 2) Describe current practices for reporting balance at baseline on both cluster- and individual-level characteristics (e.g., by what combination of sequence, period, and condition; with or without significance testing) and how this varies across trial designs;
- 3) Describe current methods of analysis used in SW-CRTs including how covariates restricted in the randomization are accounted for, prevalence of adjustment for prespecified covariates in the analysis (individual- and/or cluster-level covariate adjustment) and methods for handling missing values on covariates, and compare these to existing recommendations.

METHODS

ELIGIBILITY CRITERIA

Inclusion criteria:

- Primary reports of completed stepped-wedge CRTs conducted on humans;
- Must have at least five *independent* clusters randomized;
- Must have a minimum of two sequences and three periods;
- No more than two conditions (but multiple phases are acceptable);
- Published in English between January 2016 and March 2022.

Exclusion criteria:

- Individually randomized;
- Protocols, pilot studies, feasibility studies;
- Non-primary reports (e.g., secondary analyses, subgroup analyses, exploratory analyses);
- Wait-list designs with fewer than 2 periods;
- Not involving humans;
- Not health research (e.g., medical education trials);
- Not randomized (although trials that include a small number of non-randomized clusters is acceptable (e.g., when a site involved in a pilot study is included)).

SEARCH STRATEGY:

We aim to achieve a representative, if not comprehensive, sample of SW-CRTs published over the past 7 years. We chose 7 years as we aimed to include a reasonably large sample of published trials, whilst still keeping the project manageable. This also enables us to consider the reporting of trials published both in the years pre- and post- the publication of the CONSORT extension for SW-CRTs. To efficiently locate eligible trials for this review, trials were identified from 3 sources:

- 1) SW-CRTs identified in a recently published review of implementation challenges in SW-CRTs by Agnès Caille and colleagues.⁷ The search was implemented in PubMed and covered January 1, 2019 to September 23, 2020.
- 2) An updated search using the same terms used by Caille et al. and implemented in PubMed to cover September 24, 2020 to March 4, 2022.
- 3) SW-CRTs identified in a previously established database of 4336 primary reports of pragmatic trials, covering January 2014 to April 2019.²⁸ This was done to increase the

date range of our review to include trials published before the CONSORT extension for SW-CRTs. It was also done to improve the efficiency of our searches as the database included only primary trial reports. Moreover, there was overlap between the search terms used to establish this database²⁹ and the search terms used by Caille et al.

SCREENING:

The previously published review by Caille et al. identified 55 trials; no screening was required as the same inclusion/exclusion criteria applied in our review.

The updated search yielded 561 potential titles and abstracts for screening. Each title and abstract were independently screened by both PN and MT in Covidence,³⁰ resulting in 125 records for potential inclusion. Once 8 duplicates with the review by Caille et al. were removed, 117 articles were imported into the study database. The full text was screened independently by two reviewers (PN and YO). Sixty-five met the criteria for this review based on full text review.

Within the database of 4336 primary reports of pragmatic trials previously described,²⁸ application of the search terms by Caille et al. identified 92 potential SW-CRTs. Of these, 9 were duplicates with the already identified 55, leaving 83 unique trials. The full text reports were screened independently by two reviewers (PN and YO); yielding 66 which fulfilled inclusion criteria. To keep the sample size for this review manageable, we excluded trials published in 2014 or 2015, leaving 46 for an overall total of 166 trials.

SAMPLE SIZE

We aimed to include at least 160 trials in this review. This sample size was considered acceptable based on yielding a two-sided 95% confidence interval around an estimated proportion with a total width no greater than 0.155 (i.e., in the most conservative case when the proportion is 0.5).

DATA EXTRACTION ELEMENTS

An extraction form was developed to standardize the capture of data elements of interest. Basic trial characteristics included identification as a pragmatic trial, region of recruitment (classified as North America, South or Central America, Europe, Asia, Middle East, Africa, or Australia/Oceania), type of intervention (classified as targeted at the health care organization, targeted at health care professionals, participant education/health promotion, or direct participant therapeutic intervention), and type of cluster (classified as hospitals or hospital wards, primary care, nursing homes, geographical areas, schools or workplaces, or other).

The data elements pertaining to trial design and statistical methodology were organized in four domains. Domain 1 included design characteristics of the SW-CRTs including type of design (e.g., cross-sectional, closed cohort, open cohort), size (number of clusters and participants), type and number of primary outcomes. Domain 2 included characteristics of randomization, including whether any restricted randomization was used, and the type and number of variables used to balance characteristics during allocation. Domain 3 included information relating to reporting of balance at baseline, in particular how characteristics (either cluster- or individual-level) were presented and whether any imbalances were described. Domain 4 included information about the analysis, focusing on the primary analysis of the primary outcome: the statistical methods used,

whether analysis adjusted for time and types of correlation structures assumed, whether and how covariates were handled in the analysis. If a primary outcome was not clearly reported or if multiple outcomes were present, reviewers were instructed to choose the outcome in the power or sample size calculation, or else the first outcome listed under outcomes or the one presented more prominently. If a primary analysis for the primary outcome was not clearly identified, reviewers were instructed to choose the main analysis reported in the abstract, or otherwise the first analysis presented for the primary outcome.

The extraction form was pilot tested on eight trials. All nine reviewers completed the pilot test as part of calibration and training and participated in consensus discussions to review discrepancies and refine the form.

DATA EXTRACTION

Basic trial characteristics will be extracted independently by either one or two trained reviewers (PN and/or LOR). Journal impact factors in the year of publication of each trial will be obtained from Journal Citation Reports;³¹ or, when unavailable, from the SCImago Journal and Country Rank. Nine reviewers (a senior supervisory statistician, 6 PhD-trained statisticians, a PhD candidate in biostatistics, and an MSc candidate in biostatistics) will proceed with data extraction for the design and statistical methodology domains. Two reviewers will be randomly allocated to each trial and will independently extract information from the full text report, any supplementary material, and any available trial protocols. Reviewers will proceed in batches of 4 trials per week and consensus discussions will be held weekly to rectify any discrepancies between reviewers. Two senior supervisory statisticians (FL and AC) will further extract information from 5 trials

apiece and hold a consensus discussion with one of the above reviewers who also independently extracted for these trials. FL, AC, or MT will be consulted in cases where consensus cannot be reached.

All data will be captured in Airtable, an online collaborative spreadsheeting service.³²

ANALYSIS

The analysis will be primarily descriptive in nature. Counts and frequencies will be presented to describe categorical variables. The range, mean and standard deviation, and/or median and interquartile range will be used to describe continuous variables. Proposed tables for presenting results are summarized as an Appendix, below. Figures may be used in place of tables where appropriate.

We will report results overall and consider variation across types of stepped-wedge design, trial size, or journal impact factor. A stratified analysis by year of publication (before or after the publication of the 2018 CONSORT extension for stepped-wedge trials) will also be considered for relevant items. Analyses will be conducted in R. We will use the results from this review to identify prevalent issues in SW-CRTs and gaps in the methodological literature.

PROJECTED TIMELINE

- Conceptualization and main development of extraction form: February to May 2022
- Extraction form pilot testing: May 2022
- Data extraction: June to August 2022
- Analysis and draft report: August to September 2022
- Final report and submission: by December 2022

ETHICS

Ethical approval is not required for this review of previously published trials.

FUNDING

This work was supported by the Canadian Institutes of Health Research through the Project Grant competition (competitive, peer-reviewed), award number PJT-153045. MT is supported by the National Institute of Aging (NIA) of the National Institutes of Health under Award Number U54AG063546, which funds NIA Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in the design of the study, in the writing of the protocol, or in the decision to submit the protocol for registration.

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS TO THE PROTOCOL

Conceptualization and methodology developed by Monica Taljaard, Fan Li, Pascale Nevins.

Original protocol draft by Pascale Nevins, Monica Taljaard. Project administration by Pascale Nevins. Supervision by Monica Taljaard, Fan Li, Agnès Caille. All authors reviewed and approved the protocol before submission.

ACKNOWLEDGEMENTS

We are grateful to summer student Eric Tran at the Ottawa Hospital Research Institute for his contributions to database management.

APPENDIX: TABLES

TABLE 1. GENERAL TRIAL AND PUBLICATION CHARACTERISTICS OF INCLUDED STEPPED-WEDGE CRTs (N=XX)

Characteristic	Frequency (%)
Publication Year 2016 2017 2018 2019 2020 2021 2022	
Country or region of study recruitment* North America South and Central America Europe Asia Australia or New Zealand Middle East Africa	
Type of experimental intervention Targeted at the healthcare organization Targeted at healthcare professionals Direct participant therapeutic intervention Participant health promotion or educational intervention Unclear or Other	
Self-identify as pragmatic Yes No	
Journal Impact Factor Mean (SD) Median (Q1 to Q3) Min, Max Missing	
Type of cluster used in randomization Hospitals or hospital wards Primary care Nursing homes Geographical areas Schools or Workplaces Other	

* Multiple selections possible

TABLE 2. DESIGN CHARACTERISTICS OF INCLUDED STEPPED-WEDGE CRTs (N=XX)

Characteristic	Frequency (%)
Number of clusters randomized Mean (SD) Median (Q1 to Q3) Min, Max Not reported	
Number of sequences Mean (SD) Median (Q1 to Q3) Min, Max Not reported	
Type of stepped-wedge trial Cross-sectional with continuous recruitment Cross-sectional with fixed time-point recruitment Open cohort Closed cohort	
Complete or incomplete design Complete Incomplete	
Number of primary outcomes One primary outcome Two or more co-primary outcomes or multivariate outcome subscales No outcome defined as primary	
Type of primary outcome Continuous Binary Ordinal or multinomial Time-to-event Count Rate	

* Multiple selections possible

TABLE 3. RANDOMIZATION METHODS USED (N=XX)

Characteristic	Frequency (%)
Strategy Unrestricted Stratification Matching Minimization Covariate-constrained randomization Other or mixture Unclear or not specified	
Timing of randomization Single time point Sequentially or in batches	
Number of inherently cluster-level characteristics balanced during allocation 0 1 >1 Unclear Range	
Type of cluster-level characteristics balanced during allocation* Cluster size Cluster location Other (e.g., ...)	
Number of inherently individual-level characteristics used in randomization 0 1 >1 Unclear Range	
Justification provided for restricting randomization by chosen variable(s) Yes, for at least one No	

*Multiple selections possible

TABLE 4. REPORTING OF BALANCE AT BASELINE, OVERALL AND BY SW-CRT DESIGN

Characteristic	Frequency (%)			
	Overall (N=XX)	Cross-sectional	Open cohort	Closed cohort
Balance on cluster-level characteristics* Not reported By condition By sequence or cluster By period By condition and sequence/cluster By condition and period By sequence/cluster and period Other				
Balance on individual-level characteristics* Not reported By condition By sequence or cluster By period By condition and sequence/cluster By condition and period By sequence/cluster and period Other				
Balance on baseline value of primary outcome* Not reported or not applicable By condition By sequence or cluster By period By condition and sequence/cluster By condition and period By sequence/cluster and period Other				
Significance testing of baseline balance Yes Simple test Other Unclear or not specified No				
Baseline imbalances identified or reported in the text? Yes No				

* Multiple selections possible

TABLE 5. CHARACTERISTICS OF THE PRIMARY ANALYSIS OF THE PRIMARY OUTCOME (N=XX)

Characteristic	Frequency (%)
Number of clusters included in analysis Mean (SD) Median (Q1, Q3) Range Missing	
Sample size Mean (SD) Median (Q1, Q3) Range Missing	
Statistical method used GEE GLMM Fixed-effects GLM Cox or Accelerated failure time model Simple/Naïve analysis Other Unclear	
Time or period effects adjusted for? Yes No	
Within-cluster correlation accounted for? Yes No	
Between-period correlation accounted for? Yes No	
Method for degrees of freedom reported? Yes No	
Absolute or relative treatment effects (N=ZZ with non-continuous outcome) Only absolute Only relative Both absolute and relative	
Primary results Positive (i.e., statistically significant) Negative	

TABLE 6. COVARIATE ADJUSTMENT IN ANALYSES OF THE PRIMARY OUTCOME (N=XX)

Characteristics	Frequency (%)
Covariates included in the analysis Yes, primary analysis Yes, non-primary analysis No, no covariates in any analysis of primary outcome	
Treatment-by-covariate interactions included in any analyses? Yes No	
Both adjusted and unadjusted analyses presented Yes, and they differ in significance Yes, but they do not differ in significance No	
Number of cluster-level covariates in primary analysis 0 1 >1 Range	
Number of individual-level covariates in primary analysis 0 1 >1 Range	
Covariates used in restricted randomization adjusted in the primary analysis? (N=YY with restricted randomization) Yes, some or all All identical in scale and level Change in at least one in terms of scale and/or level Unclear No, none	
Adjustment for baseline measure of primary outcome? Yes No or not applicable	
Handling of continuous covariates* Simple linear terms Dichotomization or categorization Other Not specified	
Covariates adjusted in the primary analysis prespecified? Clearly prespecified and included Clearly prespecified, but some omitted Clearly chosen post hoc Mixture (some prespecified, some post hoc) Unclear	
Rationale for covariate adjustment? Chance imbalance/confounding To improve precision of treatment effect To account for missing data under MAR assumption Other No information	

<p>Method for handling missing data on covariates?</p> <p>Complete case analysis or no method specified</p> <p>Missing indicator method</p> <p>Single imputation</p> <p>Multiple imputation</p> <p>Other</p> <p>No missing data reported</p>	
<p>Missing covariates noted as a barrier to adjustment</p> <p>Yes</p> <p>No</p>	

*Multiple selections possible

REFERENCES

- ¹ Donner A, Klar N. Design and analysis of cluster randomization trials in health research London. Arnold Publishers Limited. 2000
- ² Taljaard M, Goldstein CE, Giraudeau B, Nicholls SG, Carroll K, Hey SP, Brehaut JC, Jairath V, London AJ, Eldridge SM, Grimshaw JM, Fergusson DA, Weijer C. Cluster over individual randomization: are study design choices appropriately justified? Review of a random sample of trials. *Clin Trials*. 2020 Jun;17(3):253-263. doi: 10.1177/1740774519896799. Epub 2020 May 5. PMID: 32367741.
- ³ Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015 Feb 6;350:h391. doi: 10.1136/bmj.h391. PMID: 25662947.
- ⁴ Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007 Feb;28(2):182-91. doi: 10.1016/j.cct.2006.05.007. Epub 2006 Jul 7. PMID: 16829207.
- ⁵ Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*. 2015 Aug 17;16:352. doi: 10.1186/s13063-015-0842-7. PMID: 26279154; PMCID: PMC4538756.
- ⁶ Prost A, Binik A, Abubakar I, Roy A, De Allegri M, Mouchoux C, Dreischulte T, Ayles H, Lewis JJ, Osrin D. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. *Trials*. 2015 Aug 17;16:351. doi: 10.1186/s13063-015-0837-4. PMID: 26278521; PMCID: PMC4538739.
- ⁷ Caille A, Taljaard M, Le Vilain-Abraham F, Le Moigne A, Copas AJ, Tubach F, Dechartres A. Recruitment and implementation challenges were common in stepped-wedge cluster randomized trials: results from a methodological review. *J Clin Epidemiol*. 2022 Apr 25:S0895-4356(22)00113-5. doi: 10.1016/j.jclinepi.2022.04.024. Epub ahead of print. PMID: 35483552.
- ⁸ Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*. 2021 Feb;30(2):612-639. doi: 10.1177/0962280220932962. Epub 2020 Jul 6. PMID: 32631142; PMCID: PMC7785651.
- ⁹ Kasza J, Hemming K, Hooper R, Matthews J, Forbes AB. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res*. 2019 Mar;28(3):703-716. doi: 10.1177/0962280217734981. Epub 2017 Oct 13. PMID: 29027505.
- ¹⁰ Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Med Res Methodol*. 2016 Jun 6;16:69. doi: 10.1186/s12874-016-0176-5. PMID: 27267471; PMCID: PMC4895892.
- ¹¹ Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge S, Campbell MJ, Lilford RJ, Weijer C, Forbes AB, Grimshaw JM. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018 Nov 9;363:k1614. doi: 10.1136/bmj.k1614. PMID: 30413417; PMCID: PMC6225589.
- ¹² Martin J, Taljaard M, Girling A, Hemming K. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ Open*. 2016 Feb 4;6(2):e010166. doi: 10.1136/bmjopen-2015-010166. PMID: 26846897; PMCID: PMC4746455.
- ¹³ Grayling MJ, Wason JM, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. *Trials*. 2017 Jan 21;18(1):33. doi: 10.1186/s13063-017-1783-0. PMID: 28109321; PMCID: PMC5251280.
- ¹⁴ Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, Fielding KL. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. *Trials*. 2015 Aug 17;16:358. doi: 10.1186/s13063-015-0838-3. PMID: 26278667; PMCID: PMC4538923.

- ¹⁵ Federico CA, Heagerty PJ, Lantos J, O'Rourke P, Rahimzadeh V, Sugarman J, Weinfurt K, Wendler D, Wilfond BS, Magnus D. Ethical and epistemic issues in the design and conduct of pragmatic stepped-wedge cluster randomized clinical trials. *Contemp Clin Trials*. 2022 Apr;115:106703. doi: 10.1016/j.cct.2022.106703. Epub 2022 Feb 15. PMID: 35176501.
- ¹⁶ Taljaard M, Hemming K, Shah L, Giraudeau B, Grimshaw JM, Weijer C. Inadequacy of ethical conduct and reporting of stepped wedge cluster randomized trials: Results from a systematic review. *Clin Trials*. 2017 Aug;14(4):333-341. doi: 10.1177/1740774517703057. Epub 2017 Apr 8. PMID: 28393537.
- ¹⁷ Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol*. 2011 Sep;64(9):936-48. doi: 10.1016/j.jclinepi.2010.12.003. Epub 2011 Mar 16. PMID: 21411284.
- ¹⁸ Kristunas C, Morris T, Gray L. Unequal cluster sizes in stepped-wedge cluster randomised trials: a systematic review. *BMJ Open*. 2017 Nov 15;7(11):e017151. doi: 10.1136/bmjopen-2017-017151. PMID: 29146637; PMCID: PMC5695383.
- ¹⁹ Kristunas CA, Hemming K, Eborall H, Eldridge S, Gray LJ. The current use of feasibility studies in the assessment of feasibility for stepped-wedge cluster randomised trials: a systematic review. *BMC Med Res Methodol*. 2019 Jan 10;19(1):12. doi: 10.1186/s12874-019-0658-3. PMID: 30630416; PMCID: PMC6327386.
- ²⁰ Beard E, Lewis JJ, Copas A, Davey C, Osrin D, Baio G, Thompson JA, Fielding KL, Omar RZ, Ononge S, Hargreaves J, Prost A. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials*. 2015 Aug 17;16:353. doi: 10.1186/s13063-015-0839-2. PMID: 26278881; PMCID: PMC4538902.
- ²¹ Ciolino JD, Palac HL, Yang A, Vaca M, Belli HM. Ideal vs. real: a systematic review on handling covariates in randomized controlled trials. *BMC Med Res Methodol*. 2019 Jul 3;19(1):136. doi: 10.1186/s12874-019-0787-8. PMID: 31269898; PMCID: PMC6610785.
- ²² Wright N, Ivers N, Eldridge S, Taljaard M, Bremner S. A review of the use of covariates in cluster randomized trials uncovers marked discrepancies between guidance and practice. *J Clin Epidemiol*. 2015 Jun;68(6):603-9. doi: 10.1016/j.jclinepi.2014.12.006. Epub 2014 Dec 30. PMID: 25648791; PMCID: PMC4425474.
- ²³ Ivers NM, Halperin IJ, Barnsley J, Grimshaw JM, Shah BR, Tu K, Upshur R, Zwarenstein M. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials*. 2012 Aug 1;13:120. doi: 10.1186/1745-6215-13-120. PMID: 22853820; PMCID: PMC3503622
- ²⁴ Chaussee EL, Dickinson LM, Fairclough DL. Evaluation of a covariate-constrained randomization procedure in stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2021 Jun;105:106409. doi: 10.1016/j.cct.2021.106409. Epub 2021 Apr 21. PMID: 33894362.
- ²⁵ Lew RA, Miller CJ, Kim B, Wu H, Stolzmann K, Bauer MS. A method to reduce imbalance for site-level randomized stepped wedge implementation trial designs. *Implement Sci*. 2019 May 3;14(1):46. doi: 10.1186/s13012-019-0893-3. PMID: 31053157; PMCID: PMC6500026.
- ²⁶ Senn S. Testing for baseline balance in clinical trials. *Stat Med*. 1994 Sep 15;13(17):1715-26. doi: 10.1002/sim.4780131703. PMID: 7997705.
- ²⁷ Wright N, Ivers N, Eldridge S, Taljaard M, Bremner S. A review of the use of covariates in cluster randomized trials uncovers marked discrepancies between guidance and practice. *J Clin Epidemiol*. 2015 Jun;68(6):603-9. doi: 10.1016/j.jclinepi.2014.12.006. Epub 2014 Dec 30. PMID: 25648791; PMCID: PMC4425474.
- ²⁸ Nicholls SG, Carroll K, Hey SP, Zwarenstein M, Zhang JZ, Nix HP, Brehaut JC, McKenzie JE, McDonald S, Weijer C, Fergusson DA, Taljaard M. A review of pragmatic trials found a high degree of diversity in design and scope, deficiencies in reporting and trial registry data, and poor indexing. *J Clin Epidemiol*. 2021 Sep;137:45-57. doi: 10.1016/j.jclinepi.2021.03.021. Epub 2021 Mar 28. PMID: 33789151; PMCID: PMC8996736.
- ²⁹ Taljaard M, McDonald S, Nicholls SG, Carroll K, Hey SP, Grimshaw JM, Fergusson DA, Zwarenstein M, McKenzie JE. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked

sensitivity. *J Clin Epidemiol*. 2020 Aug;124:75-84. doi: 10.1016/j.jclinepi.2020.05.003. Epub 2020 May 11. PMID: 32407765.

³⁰ Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia, 2022. Available at www.covidence.org.

³¹ 2016-2022 Journal Impact Factors, Journal Citation Reports. Clarivate Analytics, 2022.

³² Airtable, Formagrid, Inc., CA, USA, 2022. Available at <https://airtable.com/>.