

Prevalence and Methodological Characteristics of Subgroup Analyses in Stepped Wedge Cluster Randomised Trials: Protocol for a Systematic Review

Évèhouéno Lionel Adisso

Université Laval

Monica Taljaard

University of Ottawa

Louis-Paul Rivest

Université Laval

Hervé Tchala Vignon Zomahoun

Université Laval

Pierre Jacob Durand

Université Laval

France Légaré (✉ france.legare@fmed.ulaval.ca)

Université Laval

Protocol

Keywords: Stepped wedge design, Cluster randomised trials, Subgroup analyses, Statistical interactions, Effect modifiers, Heterogeneity of treatment effect, Moderators of treatment effects

Posted Date: December 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-133475/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The stepped wedge cluster randomised trial is an increasingly common trial design. The design can be useful for informing real-world clinical decision-making, including decisions about the effectiveness of interventions in particular subgroups. However, there is little existing guidance about how to perform subgroup analyses in the stepped wedge design. We aim to determine the prevalence of subgroup analyses and describe statistical methods used to perform them in stepped wedge cluster randomised trials.

Methods: We will conduct a systematic review following the methodology recommended in the Cochrane Handbook for Systematic Reviews of Interventions. We report this protocol according to the PRISMA-P checklist. The protocol has been registered in the Open Science Framework. We will search for terms related to 'stepped wedge'. Sources will be PubMed, Embase, PsycINFO, Web of Science, CINAHL, Cochrane Library, and Current Controlled Trials Register up to 16 October 2020. Studies will be eligible if they are written in English, involve human participants and are primary or secondary reports of planned or completed stepped wedge cluster randomised trials. Two reviewers will first screen the titles and abstracts, then full texts, to select studies that should be included in the review. Disagreements will be solved by consensus through discussion with a third reviewer. We will extract data related to study characteristics including presence or absence of subgroup analyses, characteristics of subgroup variables examined, statistical methods used to perform subgroup analyses, and adherence to the most consistently recommendations suggested for subgroup analyses in general including in clinical trials. We will perform a qualitative synthesis of the extracted data.

Discussion: This protocol offers a reproducible and transparent procedure for a systematic review of the literature. It will provide a portrait of the frequency and types of subgroup analyses performed in stepped wedge cluster randomised trials. These results will inform the development of recommendations for subgroup analyses in such trials.

Systematic review registration: This protocol has been registered on Open Science Framework, Registration ID: <https://osf.io/2kwrz>.

Background

A stepped wedge cluster randomised trial (SW-CRT) is a relatively new pragmatic study design particularly used to evaluate interventions during their routine implementation in real-world settings [1, 2]. It is commonly used to evaluate health policy and service delivery interventions[3]. SW-CRT randomises clusters to a sequence of time-periods spent in the control condition followed by time-periods spent in the treatment condition. Usually, there is an initial period in which no clusters are exposed to the intervention [4]. Subsequently, at regular intervals ("periods"), one cluster (or a group of clusters) is randomised to cross from the control to the intervention group [5]. This process continues until all clusters have crossed over to the intervention (Fig. 1). Data collection continues throughout the study so that each cluster

contributes observations under both control and intervention conditions, improving its statistical efficiency when making comparisons [5]. As all clusters ultimately receive the intervention, the study has more social acceptability. These advantages, among others [6], may explain the increasing use of the SW-CRT design in recent years [7]. But the design also has many disadvantages. It has numerous methodological complexities such as potential confounding of time [8–10]; possible within-cluster contamination [11]; possible time varying treatment effect [12]; potential cluster-treatment heterogeneity [13]; and complex correlation structures [14–17]. This makes data analysis of SW-RCTs more complex than other designs [10]. Statistical methods for data analysis of SW-CRTs have been proposed [8, 18, 19] and others are still in development [9, 20]. The design can be useful to inform real-world clinical decision-making, including decisions about the effectiveness of interventions in particular subgroups [21–24]. For this reason, the CONSORT extension for the SW-CRTs recommends (item 18) to report results of subgroup analyses performed [25].

Subgroup analyses are used to examine whether the effect of one variable (e.g. exposure) on another (e.g. outcome) varies across strata (subgroups) of a third (e.g. demographic characteristics of patients) [26, 27]. Such analyses are performed to inform individualized treatment decisions [28, 29] or to investigate the consistency of the trial conclusions among different subpopulations [30, 31]. When appropriately performed, subgroup analyses can lead to more targeted clinical recommendations, better informed clinical decision making and improved patient care [32]. Existing recommendations for subgroup analyses, derived from methodological papers and systematic reviews, mainly focus on design, analysis, reporting and interpretation [33–35]. Regarding design, subgroup analyses should 1) be based on strong biological reasoning, previous empirical evidence or current scientific theory to reduce susceptibility to bias [36]; 2) be pre-specified (the plan should be laid out in the protocol) rather than post-hoc to reduce risk of spurious findings [37]; 3) include power calculation (only if the subgroup analyses are related to the primary trial objectives) to ensure the trial sample size is adequate to detect interaction [38]; 4) be measured prior to randomization so as not to be affected by treatment response; 5) stratify randomization based on subgroup variables to allow for balanced treatment assignments within subgroups; [35, 39, 40]; and 6) test a small number (≤ 5) of subgroup hypotheses to reduce false-positives [34, 40]. Regarding analysis, subgroup analyses depend on the statistical methods used to assess primary outcome effects [41, 42]. To determine whether there is a subgroup effect, it is recommended to 1) use a formal statistical test for the interaction [35, 41, 42]; 2) adjust for multiple subgroup hypothesis testing by applying correction (e.g. Bonferroni method) [42–45]; and 3) check the subgroups for comparability of prognostic factors [46]. Regarding reporting, it is suggested that studies should report 1) all the subgroup analyses performed [33]; and 2) the scale (additive or multiplicative) on which subgroup analyses were assessed [33, 47]. Regarding interpretation, experts strongly suggest that results of subgroup effects should be interpreted with caution to prevent mis- or overinterpretation [48–50]. From the 1980s to 2018, of all the methodological recommendations made, five basic recommendations have remained consistent and are frequently suggested for subgroup analyses in general (Table 1) [26, 34, 40, 42, 51, 52].

More recommendations are related to standard randomised controlled trials (RCTs) than to cluster randomised trials (CRTs). Subgroup analyses are rare in CRTs, which limits their ability to shed light on the extent of benefits or risks for treatments tested [35]. Subgroup analyses are most straightforward in trials with a single measurement from each participant or cluster. However, multiple period trials such as SW-CRTs have different underlying modeling assumptions which can complicate the conduct of subgroup analyses [53]. Due to the inherent confounding of the treatment effect with time, analysis of a SW-CRT should always account for secular trends [9, 54]. Subgroup analysis in SW-CRTs could require introducing into the statistical model two interaction terms involving the subgroup variable: with treatment and with time [31, 41]. To the best of our knowledge, there is no recommendation about how to handle this issue. Recommendations might depend on whether the subgroup variable is a cluster-level or individual-level variable, and whether differences in the secular trend are anticipated across the subgroups. As a first step in developing methodological recommendations for subgroup analyses in SW-CRTs, it will be useful to know how they have been handled so far in published trials and protocols of SW-CRTs. Therefore, we aim to review reports of SW-CRTs to identify and describe the statistical methods used to perform subgroup analyses. Specific objectives are: 1) determine the prevalence of reporting subgroup analyses in SW-CRTs; 2) describe the characteristics of subgroup variables examined; 3) identify and describe statistical methods used to perform subgroup analyses in SW-CRTs; and 4) determine prevalence of adherence to the most consistently recommendations suggested for subgroup analyses in general including clinical trials.

Methods

We will conduct a systematic review. We report this protocol according to the PRISMA-P guidelines [55] and provide the Checklist in an additional file. We registered this protocol with the Open Science Framework (Registration ID: <https://osf.io/2kwrz>).

Eligibility criteria

Studies will be eligible if they are a) written in English, b) involve human participants, and c) are protocols, primary or secondary reports of planned or completed SW-CRT. Papers with their protocols will be considered as one completed original study. We will include only studies that use cluster randomisation with a minimum of two sequences and three periods or three sequences and two periods. The design may not have all clusters starting in control and ending in intervention, or may not have complete data [56]. We will not include any restriction on interventions, comparators, and outcomes. We will consider both healthcare and non-healthcare settings. We will exclude a) individually randomised trials; b) bi-direction cross-over; c) non-randomised stepped wedge designs; or d) trials retrospectively analysed as a stepped wedge design when the study was not originally designed as a stepped wedge [54]. We will also exclude systematic reviews, editorials, design manuscripts, and letters.

Information sources and search strategy

We will use an adaptation of three previously published search strategies of systematic reviews on stepped wedge design [7, 19, 57, 58]. We will search the following databases up to 16 October 2020: PubMed, Embase, PsycINFO, Web of Science, CINAHL, Cochrane Library and Current Controlled Trials Register. Our search terms are “stepped wedge”, “step wedge”, “experimentally staged introduction”, “one-way crossover”, “one directional crossover”, “one way cross-over”, “SW-CRT” as well as the 28 combinations of the terms “incremental”, “phased”, “staggered”, “staged”, “stepwise”, “step wise” and “delayed” with the terms “recruitment”, “introduction”, “implementation”, “intervention”. An information specialist will first perform the search strategy in PubMed, and then will translate it into the other databases. A second information specialist will revise the initial search strategy with the *Peer Review of Electronic Search Strategies (PRESS)* Tool [59]. Our search strategy in PubMed is described in Table 2. We will report on the search process following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [60]. The flow diagram is presented in Fig. 2. We will also search grey literature by contacting stepped wedge design experts of relevant networks such as the Ottawa Methods Centre (OMC) at the Ottawa Hospital Research Institute, Canada [61], the Pragmatic Clinical Trials Unit (PCTU) in London, UK [62] and the National Institute of Health, USA [63].

Data management

We will first merge the citations identified from seven electronic databases mentioned above in EndNote software. We will then identify and remove the duplicates. Unique citations will be considered for the selection process.

Study selection process

We will complete the screening in four steps. First, we will randomly select a small sample (10%) of unique records identified. Two reviewers will independently screen these studies using our inclusion/exclusion criteria. We will assess the agreement between screeners to ensure that the eligibility screening process is reproducible and reliable. We will describe the observed proportions of articles where pairs of screeners agree or disagree on their eligibility decisions and calculate Kappa statistics [64]. Disagreements will be resolved through discussion. If necessary, we will modify instructions and re-test another sample to improve the agreement between screeners [65]. Second, once agreement is acceptable, the two reviewers will independently screen the remaining titles and abstracts. Third, we will obtain full-text articles for all the studies that passed the title and abstract screening and confirm eligibility. In addition, we will also examine references lists of relevant systematic reviews and eligible papers for additional papers which the search strategy would have missed. Papers with their protocols will be considered as one completed original study. The studies found not to meet the eligibility criteria will be excluded and we will document the reason for exclusion. Finally, from full text screening, we will identify SW-CRTs in which subgroup analyses were performed. We will email study authors when relevant information for the selection decision is missing or unclear. Discrepancies between the two reviewers will be resolved by consensus through discussion with a third reviewer.

Data extraction

We will develop a data extraction form (an Excel spreadsheet) by examining items reported in other relevant systematic reviews on subgroup analyses or on SW-CRTs [19, 40, 54, 66, 67] and developing new items specific to our review. The data extraction form will be pilot tested on a sample of 10% trials (for which full-text screening is complete) to refine the extraction items and to ensure that data will be collected consistently. Two reviewers will extract data from these trials and discrepancies will be identified and resolved through discussion. The data collection form will be improved if necessary. Once consensus is reached, the form will be finalized, and the trials will be randomly divided among the reviewers with two reviewers assigned to each trial. Reviewers will meet periodically (e.g. after every five trials have been completed) to review discrepancies and come to a consensus. If consensus cannot be obtained, a third investigator will resolve differences. Kappa statistics will be calculated to determine the agreement between the independent reviewers. We will extract data in multiple domains related to our objectives. First, we will extract information on **study characteristics**: first author, title, year of publication, type of article (e.g. primary report or secondary report of a completed trial), country, setting (e.g. non-health care), type of design (e.g. cross sectional), number of clusters, number of sequences, number of periods, number of participants per step, step length, trial duration, presence of subgroup analyses. Second, we will extract **characteristics of subgroup variables examined**: number of subgroup variables, whether subgroup variables were defined at cluster-level or individual-level, type of each subgroup variable (e.g. categorical), number of categories, how subgroup variables were measured during data collection phase (e.g. continuous), how the cut-off was justified (if categorised). We will extract detailed information about the **statistical methods used to perform subgroup analyses in SW-CRTs**. Third, we will consider a) methods used for the outcome on which subgroup analyses were based: outcome variable (e.g. primary outcome), type of outcome variable (e.g. continuous), unit of analysis (e.g. individual-level-analysis), distribution (e.g. binomial), statistical models (e.g. Generalised Linear Mixed Model), association measure (e.g. risk ratio), assumptions about the correlation structure (e.g. model suggested by Hussey et al. [17]), whether time was modelled as discrete or continuous, whether the analysis adjusted for time, whether there was any interaction between time and treatment, whether the effect of the intervention on the outcome of interest was statistically significant. We will then consider b) statistical methods specific to subgroup analyses: methods used to assess subgroup effect (e.g. test for interaction), scale on which interaction was assessed (e.g. multiplicative), whether there was a significant test for interaction, which interaction terms specifically were included in the statistical model, what statistical methods were used to adjust for multiple subgroup hypotheses testing, whether authors claimed a subgroup effect (reported treatment effect by subgroup instead of overall treatment effect) [40]. Fourth, we will examine **adherence to the most consistently recommendations suggested for subgroup analyses in general including in clinical trials** (Table 1): whether a rationale for subgroup analyses was provided, type of subgroup analysis (e.g. prespecified), whether a formal test for interaction was used to assess subgroup effect, whether multiple subgroup analyses were performed and, whether a correction for multiplicity was applied when multiple subgroup analyses were performed.

Quality assessment

As our study is a systematic review of statistical methods used to perform subgroup analyses in SW-CRTs, we will not assess quality of trials. In addition, the quality of subgroup analyses in SW-CRTs will not be assessed since there is no methodological tool designed for this purpose.

Data synthesis

Kappa statistic will be performed to determine the agreement between reviewers for the main items. We will consider agreement as “fair” when kappa statistic values are between 0.40 and 0.59, “good” when values are between 0.60 and 0.74, and “excellent” when values are greater or equal to 0.75[68]. Our results will be presented separately according whether they are protocols or completed original studies. We will perform a descriptive analysis in four steps. First, we will describe the characteristics of all included SW-CRTs. We will use median (interquartile range) to describe continuous variables and frequency (percentage) to describe categorical variables. We will then identify the number of completed original research (or protocols) that performed (or planned to perform) subgroup analyses, overall and in each category of study. Second, we will describe characteristics of subgroup variables examined. Third, we will identify and describe statistical methods used to perform subgroup analyses in SW-CRTs. Fourth, we will determine for each of the five items (Table 1), the prevalence of adherence of the included trials to the most consistently suggested recommendations for subgroup analyses in general including in clinical trials. Analyses will be performed using version 9.4 of SAS software.

Discussion

This protocol offers a reproducible and transparent procedure for a systematic review of the literature. We aim to determine how often subgroup analyses are performed in SW-CRTs and what statistical methods are used. We hope this research protocol will achieve the following:

First, this review will fill gaps in the literature on both subgroup analyses (rationale) and SW-CRTs (methods). Second, it will provide information on the characteristics of subgroup variables examined in SW-CRTs. Third, it will help researchers to perform such analyses, as recommended in the extension of the CONSORT guidelines for SW-CRTs [24, 25].

We hypothesize that subgroup analyses are rare in SW-CRTs, as they are rare in health-related CRTs in general [35]. We also hypothesize that the prevalence of adherence to the first two of the five most consistently recommendations suggested in reference to subgroup analysis, i.e. justifying the subgroup analyses and specifying them *a priori* (Table 1), will be very low, as Moreira et al. found [42]. Finally, by reviewing the statistical methods used to perform subgroup analyses in SW-CRTs, [24] this review will lay the foundation for development of specific recommendations for such analyses.

Potential limitations of our study relate to the selection of articles. First, we will not systematically search the grey literature (unpublished articles). We plan to contact authors of relevant papers, but the rate of response in this situation could be low [69, 70]. In addition, data from unpublished studies can introduce bias [71]. They may reflect an unrepresentative sample of all unpublished studies and be of lower

methodological quality than published studies. Furthermore, as the selection criteria focus strongly on methodology, we feel there is little chance that a validated new methodology of statistical methods for performing subgroup analyses will be hidden in unpublished studies. Second, our results will depend mainly on information reported in trials publication. Authors may have conducted subgroup analyses but failed to report them. We planned to contact authors, but due to the low rate of response [69, 70], the prevalence of reporting subgroup analyses in SW-CRTs may be underestimated. Third, we could also miss some studies that report in languages other than English. However, since English remains the preferred language of scientific communication [72, 73], we anticipate that a small number of studies will be missed.

Abbreviations

SW-CRT: Stepped wedge cluster randomised trial

RCT: randomised controlled trial

CRT: Cluster randomised trial

SWD: Stepped wedge design

CONSORT: Consolidated Standards of Reporting Trials

CINAHL: Cumulative Index of Nursing and Allied Health Literature

PsychINFO: Psychological Information

PRISMA: Preferred reporting items for systematic review and meta-analysis

PRISMA-P: Preferred reporting items for systematic review and meta-analysis protocols

PRESS: Peer Review of Electronic Search Strategies

OMC: Ottawa Methods Centre

PCTU: Pragmatic Clinical Trials Unit

SAS: Statistical Analysis System

Declarations

Ethics approval and consent to participate

The ethical approval is not required because we will only use literature as data source. Also there is no requirements of an informed consent.

Consent for publication

Our manuscript contains no individual person's data in any form.

Availability of data and materials

Data and materials used at this step are available in our protocol. We provided an additional file for the PRISMA-P checklist. All data and materials that will be used during the review, will be available from the corresponding author.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work is part of a larger project primarily funded by the Canadian Institutes of Health Research (Grant number: 201403MOP-325236-KTR-CFBA-19158) and is also supported by the *CIUSSS de la Capitale-Nationale* (in-kind contribution included in the CIHR grant). France Légaré holds a Tier 1 Canada Research Chair in Shared Decision Making and Knowledge Translation. Évèhouéno Lionel Adisso received a doctoral award (Code: 290563 Adisso Lionel) from the Fonds de recherche du Québec – Santé (FRQS). The funders had no role in developing the protocol.

Authors' contributions

Évèhouéno Lionel Adisso made substantial contributions to the study conception, design and analysis, drafted the paper and substantively revised it; Monica Taljaard, made substantial contributions to the study conception, design and analysis and substantively revised the paper; Hervé Tchala Vignon Zomahoun made substantial contributions to the study conception, design and analysis, and substantively revised the paper; Louis-Paul Rivest made substantial contributions to the study conception and design; Pierre Jacob Durand made substantial contributions to the study conception and design; France Légaré made substantial contributions to the study conception, design and analysis, substantively revised the paper and is the guarantor of the review.

Acknowledgments

We thank Frédéric Bergeron at Université Laval and Nathalie Rhéault at the Quebec SPOR-SUPPORT Unit for their development of the search strategy. We also thank Maude Downey and Louisa Blair for their editorial support.

References

1. Mdege, N.D., et al., Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *Journal of clinical*

- epidemiology, 2011. **64**(9): p. 936-948.
2. Prost, A., et al., Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. *Trials*, 2015. **16**(1): p. 351.
 3. Taljaard, M., et al., Inadequacy of ethical conduct and reporting of stepped wedge cluster randomized trials: Results from a systematic review. *Clinical Trials*, 2017. **14**(4): p. 333-341.
 4. Hemming, K., et al., The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *Bmj*, 2015. **350**: p. h391.
 5. Brown, C.A. and R.J. Lilford, The stepped wedge trial design: a systematic review. *BMC medical research methodology*, 2006. **6**(1): p. 54.
 6. Zhan, Z., et al., Strengths and weaknesses of a stepped wedge cluster randomized design: its application in a colorectal cancer follow-up study. *Journal of Clinical Epidemiology*, 2014. **67**(4): p. 454-461.
 7. Martin, J., et al., Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ open*, 2016. **6**(2): p. e010166.
 8. Hemming, K., M. Taljaard, and A. Forbes, Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials*, 2017. **18**(1): p. 101.
 9. Nickless, A., et al., Mixed effects approach to the analysis of the stepped wedge cluster randomised trial—Investigating the confounding effect of time through simulation. *PloS one*, 2018. **13**(12).
 10. Hughes, J.P., T.S. Granston, and P.J. Heagerty, Current issues in the design and analysis of stepped wedge trials. *Contemporary clinical trials*, 2015. **45**: p. 55-60.
 11. Kotz, D., et al., Use of the stepped wedge design cannot be recommended: a critical appraisal and comparison with the classic cluster randomized controlled trial design. *Journal of clinical epidemiology*, 2012. **65**(12): p. 1249-1252.
 12. Davey, C., et al., Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. *Trials*, 2015. **16**(1): p. 358.
 13. Hemming, K., M. Taljaard, and A. Forbes, Modeling clustering and treatment effect heterogeneity in parallel and stepped-wedge cluster randomized trials. *Statistics in medicine*, 2018. **37**(6): p. 883-898.
 14. Girling, A.J. and K. Hemming, Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statistics in medicine*, 2016. **35**(13): p. 2149-2166.
 15. Hooper, R., et al., Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in medicine*, 2016. **35**(26): p. 4718-4728.
 16. Kasza, J., et al., Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Statistical methods in medical research*, 2019. **28**(3): p. 703-716.
 17. Hussey, M.A. and J.P. Hughes, Design and analysis of stepped wedge cluster randomized trials. *Contemporary clinical trials*, 2007. **28**(2): p. 182-191.
 18. Twisk, J., et al., Different methods to analyze stepped wedge trial designs revealed different aspects of intervention effects. *Journal of clinical epidemiology*, 2016. **72**: p. 75-83.

19. Barker, D., et al., Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC medical research methodology*, 2016. **16**(1): p. 69.
20. Martin, J.T., K. Hemming, and A. Girling, The impact of varying cluster size in cross-sectional stepped-wedge cluster randomised trials. *BMC medical research methodology*, 2019. **19**(1): p. 123-123.
21. Thabane, A., et al., Reporting quality of stepped wedge design randomized trials: a systematic review protocol. *Clinical epidemiology*, 2016. **8**: p. 261.
22. Wang, M., et al., The reporting quality of abstracts of stepped wedge randomized trials is suboptimal: A systematic survey of the literature. *Contemporary clinical trials communications*, 2017. **8**: p. 1-10.
23. Hemming, K., et al., Quality of stepped-wedge trial reporting can be reliably assessed using an updated CONSORT: crowd-sourcing systematic review. *Journal of clinical epidemiology*, 2019. **107**: p. 77-88.
24. Hemming, K., M. Taljaard, and J. Grimshaw, Introducing the new CONSORT extension for stepped-wedge cluster randomised trials. *Trials*, 2019. **20**(1): p. 1-4.
25. Hemming, K., et al., Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *bmj*, 2018. **363**: p. k1614.
26. Cook, D.I., V.J. Gebski, and A.C. Keech, Subgroup analysis in clinical trials. *Medical Journal of Australia*, 2004. **180**(6): p. 289.
27. VanderWeele, T.J. and J.M. Robins, Four types of effect modification: a classification based on directed acyclic graphs. *Epidemiology*, 2007. **18**(5): p. 561-568.
28. Sacks, F.M., et al., The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *New England Journal of Medicine*, 1996. **335**(14): p. 1001-1009.
29. Grouin, J.-M., M. Coste, and J. Lewis, Subgroup analyses in randomized clinical trials: statistical and regulatory issues. *Journal of biopharmaceutical statistics*, 2005. **15**(5): p. 869-882.
30. Jackson, R.D., et al., Calcium plus vitamin D supplementation and the risk of fractures. *New England Journal of Medicine*, 2006. **354**(7): p. 669-683.
31. Tanniou, J., et al., Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. *BMC medical research methodology*, 2016. **16**(1): p. 20.
32. Kasenda, B., et al., Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. *bmj*, 2014. **349**: p. g4539.
33. Knol, M.J. and T.J. VanderWeele, Recommendations for presenting analyses of effect modification and interaction. *International journal of epidemiology*, 2012. **41**(2): p. 514-520.
34. Dijkman, B., B. Kooistra, and M. Bhandari, How to work with a subgroup analysis. *Canadian Journal of Surgery*, 2009. **52**(6): p. 515.
35. Starks, M.A., et al., Assessing heterogeneity of treatment effect analyses in health-related cluster randomized trials: A systematic review. *PloS one*, 2019. **14**(8): p. e0219894.
36. Burke, J.F., et al., Three simple rules to ensure reasonably credible subgroup analyses. *Bmj*, 2015. **351**: p. h5651.

37. Schühlen, H., Pre-specified vs. post-hoc subgroup analyses: are we wiser before or after a trial has been performed? 2014, Oxford University Press.
38. Fletcher, J., Subgroup analyses: how to avoid being misled. *Bmj*, 2007. **335**(7610): p. 96-97.
39. Rothwell, P.M., Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet*, 2005. **365**(9454): p. 176-186.
40. Sun, X., et al., Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *Bmj*, 2012. **344**: p. e1553.
41. Wang, R., et al., Statistics in medicine—reporting of subgroup analyses in clinical trials. *New England Journal of Medicine*, 2007. **357**(21): p. 2189-2194.
42. Moreira Jr, E.D., Z. Stein, and E. Susser, Reporting on methods of subgroup analysis in clinical trials: a survey of four scientific journals. *Brazilian journal of medical and biological research*, 2001. **34**(11): p. 1441-1446.
43. Follmann, D., Subgroups and interactions. 2004: Chapman & Hall/CRC.
44. Schulz, K.F. and D.A. Grimes, Multiplicity in randomised trials II: subgroup and interim analyses. *The Lancet*, 2005. **365**(9471): p. 1657-1661.
45. Yusuf, S., et al., Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *Jama*, 1991. **266**(1): p. 93-98.
46. Sormani, M.P. and P. Bruzzi, Reporting of subgroup analyses from clinical trials. *Lancet Neurology*, 2012. **11**(9): p. 747.
47. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern epidemiology*. 2008: Lippincott Williams & Wilkins.
48. Lagakos, S.W., The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine*, 2006. **354**(16): p. 1667.
49. Spears, M., N. James, and M.R. Sydes, 'Thursday's child has far to go'—interpreting subgroups and the STAMPEDE trial. 2017, Oxford University Press.
50. Richardson, M., P. Garner, and S. Donegan, Interpretation of subgroup analyses in systematic reviews: a tutorial. *Clinical Epidemiology and Global Health*, 2019. **7**(2): p. 192-198.
51. Schandelmaier, S., et al., A systematic survey identified 36 criteria for assessing effect modification claims in randomized trials or meta-analyses. *Journal of clinical epidemiology*, 2019. **113**: p. 159-167.
52. Pincus, T., et al., Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC medical research methodology*, 2011. **11**(1): p. 14.
53. Kasza, J. and A.B. Forbes, Information content of cluster–period cells in stepped wedge trials. *Biometrics*, 2019. **75**(1): p. 144-152.
54. Beard, E., et al., Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials*, 2015. **16**(1): p. 353.

55. Moher, D., et al., Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Revista Espanola de Nutricion Humana y Dietetica*, 2016. **20**(2): p. 148-160.
56. Taljaard, M. Stepped Wedge Cluster Randomized Designs for Disease Prevention Research. Mind the gap. NIH webinar series. Available https://prevention.nih.gov/sites/default/files/2018-07/Taljaard_Stepped_Wedge.pdf. 2018.
57. Kristunas, C., T. Morris, and L. Gray, Unequal cluster sizes in stepped-wedge cluster randomised trials: a systematic review. *BMJ open*, 2017. **7**(11).
58. Eichner, F.A., et al., Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size. *Journal of clinical epidemiology*, 2019. **107**: p. 89-100.
59. McGowan, J., et al., PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of clinical epidemiology*, 2016. **75**: p. 40-46.
60. Stovold, E., et al., Study flow diagrams in Cochrane systematic review updates: an adapted PRISMA flow diagram. *Systematic reviews*, 2014. **3**(1): p. 54.
61. Ottawa methods center (OMC) – The Ottawa hospital research institute. Research design an methodology. <http://ohri.ca/ottawamethodscentre/Team.aspx>.
62. Pragmatic clinical trials unit (PCTU) - Queen Mary University of London. <https://www.qmul.ac.uk/pctu/>.
63. National institute of Health. Education and training. Group and Pragmatic and Group-Randomized Trials in Public Health and Medicine. Office of disease prevention. <https://prevention.nih.gov/education-training/pragmatic-and-group-randomized-trials-public-health-and-medicine>, USA.
64. Altman, D.G., *Practical statistics for medical research* Chapman and Hall. London and New York, 1991.
65. Collaboration for environmental evidence. Section 6 Eligibility screening. <https://www.cafonline.org/charityprofile/Collaboration-for-Environmental-Evidence/CCRegNo1157607>. March 2019.
66. Donegan, S., et al., Exploring treatment by covariate interactions using subgroup analysis and meta-regression in cochrane reviews: a review of recent practice. *PloS one*, 2015. **10**(6).
67. Brankovic, M., et al., Understanding of interaction (subgroup) analysis in clinical trials. *European journal of clinical investigation*, 2019. **49**(8): p. e13145.
68. Higgins JPT, T.J., Chandler J, Cumpston M, Li T, Page MJ, Welch VA, *Cochrane Handbook for Systematic Reviews of Interventions* version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook. 2020.
69. Young, T. and S. Hopewell, Methods for obtaining unpublished data. *Cochrane Database of Systematic Reviews*, 2011(11).

70. Manca, A., et al., Non-corresponding authors in the era of meta-analyses. *Journal of clinical epidemiology*, 2018. **98**: p. 159-161.
71. Egger, M., et al., How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess*, 2003. **7**(1): p. 1-76.
72. Van Weijen, D., The language of (future) scientific communication. *Research trends*, 2012. **31**(November).
73. Corcoran, J., English as the International Language of Science: A Case Study of Mexican Scientists' Writing for Publication. 2015.
74. Légaré, F., et al., Implementing shared decision-making in interprofessional home care teams (the IPSDM-SW study): protocol for a stepped wedge cluster randomised trial. *BMJ open*, 2016. **6**(11).

Tables

Table 1.

Most consistently recommendations suggested and rationale for subgroup analyses in general including in clinical trials

Most consistently recommendations suggested for subgroup analyses	Rationale for the recommendation
1) Subgroup analyses should be defined based on evidence, theory, or biological reasoning	The first step in evaluating a subgroup analysis is to determine its logical sense [34]. Credibility is higher if there is a compelling causal rationale explaining the subgroup effect, and lower if not (biologic rationale, clinical rationale, other mechanism)[26, 34, 40, 51, 52].
2) A priori specification of subgroup analyses	Subgroup analyses that are performed to test hypotheses generated before the study has started should be clearly distinguished from those identified after the main trial analyses are performed [34]. Credibility is higher if investigators stated a hypothesis prior to performing the study, lower if an explanation arose post hoc (confirmatory vs. exploratory; hypothesis testing vs. hypothesis generating)[26, 34, 40, 42, 51, 52].
3) An explicit test for interactions should be used to assess subgroup effect	To determine whether treatment efficacy differs between subgroups, it is recommended to use a formal test for interaction[26, 34, 42]. Credibility is higher if an interaction test suggests a small likelihood for a chance finding (rather than compatibility with chance or not interaction test at all) (test of homogeneity, test of heterogeneity)[40, 51, 52]
4) Authors limit the number of the subgroup analyses to be performed	Authors should perform a small number of subgroup analyses [26, 34, 42]. Credibility is higher if only a small number of subgroup effect have been tested [40, 51, 52].
5) Indicate potential effect on type I errors (false positives) due to multiple subgroup analyses and report methods used to address this effect	The greater the number of simultaneous subgroup analyses performed, the greater the probability of a false-positive finding caused [34]. Therefore, the significance of within-subgroup treatment effects should be adjusted for multiplicity[26, 42]. Credibility is higher if investigators accounted formally or informally for multiplicity [40, 51, 52].

Table 2.

Search Strategy in PubMed

Concepts	Search strategy keywords	Number
Step Wedge design	("step wedge*" [Title/Abstract] OR "stepped wedge*" [Title/Abstract] OR "one-way crossover" [Title/Abstract] OR "one way cross-over" [Title/Abstract] OR "One directional crossover" [Title/Abstract] OR SW-RCT [Title/Abstract])	#1
	("incremental recruitment" [Title/Abstract] OR "incremental introduction" [Title/Abstract] OR "incremental implementation" [Title/Abstract] OR "incremental intervention" [Title/Abstract] OR "phased recruitment" [Title/Abstract] OR "phased introduction" [Title/Abstract] OR "phased implementation" [Title/Abstract] OR "phased intervention" [Title/Abstract] OR "staggered recruitment" [Title/Abstract] OR "staggered introduction" [Title/Abstract] OR "staggered implementation" [Title/Abstract] OR "staggered intervention" [Title/Abstract] OR "staged recruitment" [Title/Abstract] OR "staged introduction" [Title/Abstract] OR "staged implementation" [Title/Abstract] OR "staged intervention" [Title/Abstract] OR "stepwise recruitment" [Title/Abstract] OR "stepwise introduction" [Title/Abstract] OR "stepwise implementation" [Title/Abstract] OR "stepwise intervention" [Title/Abstract] OR "step wise recruitment" [Title/Abstract] OR "step wise introduction" [Title/Abstract] OR "step wise implementation" [Title/Abstract] OR "step wise intervention" [Title/Abstract] OR "delayed recruitment" [Title/Abstract] OR "delayed introduction" [Title/Abstract] OR "delayed implementation" [Title/Abstract] OR "delayed intervention" [Title/Abstract])	#2
Total	#1 OR #2	#3

Figures

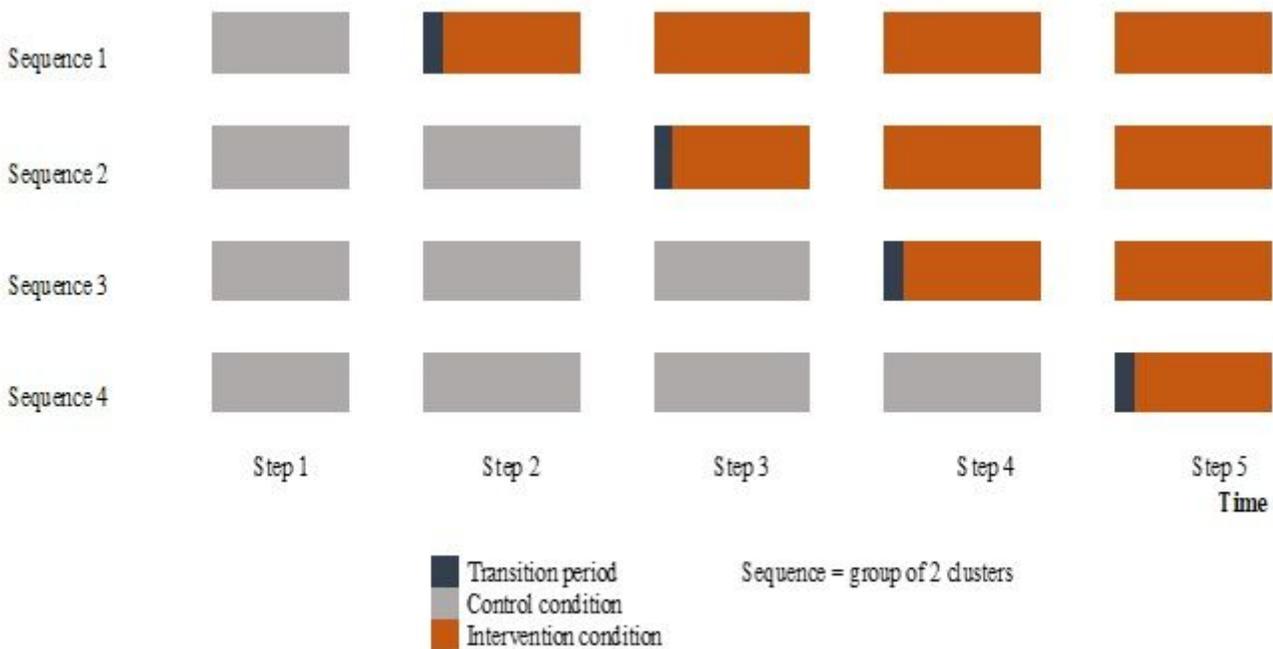


Figure 1

An example of stepped wedge cluster randomised trial with 4 sequences and 5 periods (steps) [74]

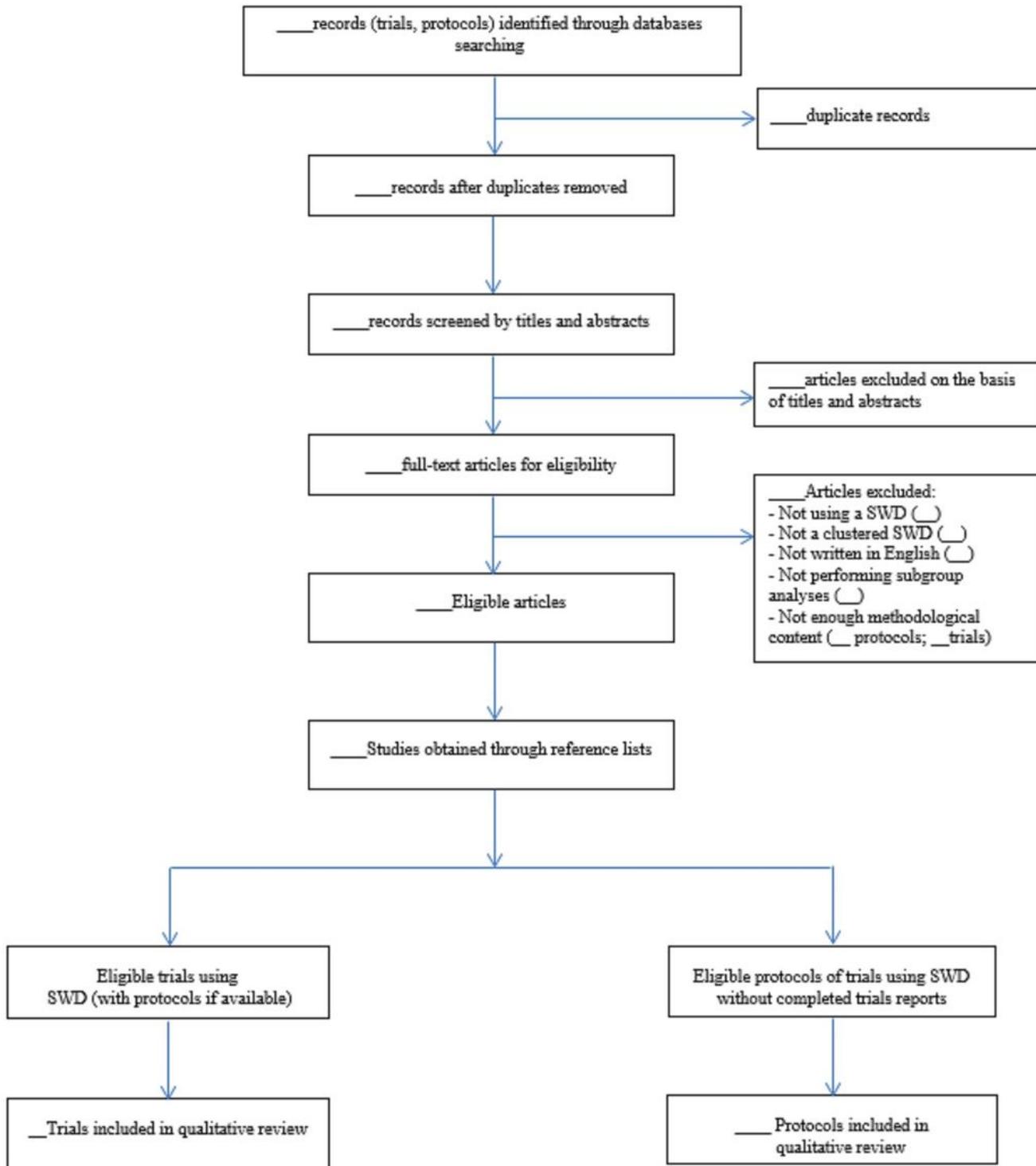


Figure 2

The preferred Reporting Items for Systematic reviews and Meta-analysis (PRISMA) flow diagram for article selection (Abbreviation: SWD, stepped wedge design)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMAPChecklist2015.docx](#)