Recruitment and implementation challenges are common in stepped-wedge cluster randomized trials: results from a methodological review

Agnes Caille, Monica Taljaard, Floriane Le Vilain-Abraham, Alexis Le Moigne, Andrew Copas, Florence Tubach, Agnes Dechartres

▶ **To cite this version:**

Agnes Caille, Monica Taljaard, Floriane Le Vilain-Abraham, Alexis Le Moigne, Andrew Copas, et al.. Recruitment and implementation challenges are common in stepped-wedge cluster randomized trials: results from a methodological review. Journal of Clinical Epidemiology, 2022, 148, pp.93-103. 10.1016/j.jclinepi.2022.04.024 . inserm-03868094

HAL Id: inserm-03868094

https://inserm.hal.science/inserm-03868094

Submitted on 23 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recruitment and implementation challenges are common in stepped-wedge cluster randomized trials: results from a methodological review

Agnès CAILLE[1,2], Monica TALJAARD[3,4], Floriane LE VILAIN—ABRAHAM[2], Alexis LE MOIGNE[1], Andrew J COPAS[5], Florence TUBACH[1*], Agnès DECHARTRES[1*]

[1] Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP, Hôpital Pitié-Salpêtrière, Département de Santé Publique, Paris, France

[2] Université de Tours, Université de Nantes, INSERM, SPHERE U1246, Tours, France; INSERM CIC 1415, CHRU de Tours, Tours, France

[3] Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

[4] School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

[5]MRC Clinical Trials Unit at University College London, London, UK

* FT and AD contributed equally to this article

Correspondance to:

Agnès Caille, MD, PhD

Email: agnes.caille@med.univ-tours.fr

Phone: (33) 2 34 37 96 54

Word count: 3471

Abstract (199/200 words)

Objective: To explore challenges in recruitment and intervention implementation in recent stepped-wedge cluster randomized trials (SW-CRTs).

Study design and setting: We searched PubMed to identify primary reports of SW-CRTs (2019–2020). Two reviewers independently screened studies and extracted data from each report. Recruitment challenge was defined as planned number of clusters or participants not achieved, or any reported changes made to the design to address recruitment difficulties. Implementation challenge was defined as early, late or no implementation of the intervention in at least one cluster.

Results: Of 55 SW-CRTs, 18 (33%) had a recruitment challenge, 23 (42%) had none, and for 14 (26%) it was impossible to judge. At least one implementation challenge was present in 24 (44%), 8 (15%) had none, and for 23 (42%) it was impossible to judge. Of the 35 (64%) trials with recruitment or implementation challenges, 18 (72%) had one or more modifications of their design, most often a modification of the trial duration.

Conclusion: Investigators must be aware of the risks of recruitment or implementation challenges when considering use of a SW-CRT design. Mitigating strategies should be adopted when planning the trial. More transparent reporting of planned and actual design features is required.

Keywords: Stepped-wedge cluster randomized trial; recruitment; implementation; design; methodological review; reporting

Running title: Recruitment or implementation challenge in stepped-wedge cluster randomized trials

What is new?

- Recruitment and implementation challenges are frequent in stepped-wedge cluster randomized trials (SW-CRTs)
- Many SW-CRT reports do not permit an assessment of whether the planned schedule has been adhered to
- Reasons for implementation challenges are heterogeneous; some can be avoided, others are unpredictable
- The risks of recruitment and implementation challenges must be considered prior to initiating a SW-CRT. Once the decision has been made to adopt a SW-CRT, mitigating strategies should be put in place
- Transparent reporting of planned and realized design features is essential for correct interpretation of the results

1. Introduction

Cluster randomized trials (CRTs) are trials that randomize groups, such as hospitals, general practices, or geographical areas, rather than individuals [1]. The most common CRT design is the two-parallel group design in which randomization determines which clusters will receive the intervention and which will receive the control. A recent and increasingly used alternative CRT design is the stepped-wedge CRT (SW-CRT) in which all clusters receive the intervention by the end of the trial; clusters are randomized to sequences and these sequences determine the timing at which a cluster will start implementing the intervention [2]. In each cluster, measurements are repeatedly taken from initial time periods spent in the control condition and subsequent time periods spent in the intervention condition. Advantages over a classic two-parallel group CRT that have been highlighted in the literature include logistical benefits due to a staggering of intervention implementation over time, enhanced recruitment because all clusters will receive the intervention before the end of the trial, and under certain conditions, better statistical efficiency [3]. Although appealing at first view, SW-CRTs may be at higher risk of bias than parallel CRTs [4] and require advanced statistical methods to account for underlying secular trends [5].

Designing a SW-CRT is complex and requires many elements to be specified in advance such as the number of sequences, clusters and periods [6]. Once started, a SW-CRT is like a race against time; clusters must adhere to the planned schedule, i.e., they must comply with the timing of implementation of the intervention and attain the target sample size in each period. However, several challenges in the implementation of the intervention and recruitment have been reported in SW-CRTs [7]. To reach the planned sample size or accommodate unanticipated problems in intervention delivery, changes in the trial design are sometimes decided during the trial such as extension of trial duration or postponement of intervention implementation in some clusters, which can have implications for clusters yet to receive the

intervention [8]. While these challenges are not specific to the SW design, any changes in the timing of a SW-CRTs affect control and intervention observations unevenly and may therefore alter the results and their interpretation. A previous methodological review of 35 SW-CRTs found that only 69% recruited their targeted number of participants and 43% reported difficulties during the study conduct such as cluster dropout or delayed intervention [9]. Nevertheless, this review did not explore recruitment challenges in detail and to our knowledge, implementation challenges have never been systematically assessed in SW-CRTs.

The aim of this study was to describe recruitment and implementation challenges in recent SW-CRTs and to assess whether and how such challenges were accommodated in the trial design and analysis strategy. We also sought to investigate factors associated with recruitment or implementation challenges.

2. Methods

2.1. Search Strategy

We searched MEDLINE via PubMed to identify eligible SW-CRTs. The search algorithm, implemented on September 23, 2020, was based on previously published electronic search strategies [9,10] and used several synonyms to describe the SW design (Supplementary Appendix A).

2.2. Eligibility criteria

We included full reports of SW-CRTs conducted in humans and published in English between January 1, 2019 and September 23, 2020. We restricted the search to this period to focus on trials published after the CONSORT extension for SW-CRTs in November 2018 [6]. We only included primary reports of completed trials: we excluded research letters, protocols, secondary or subgroup analysis papers and methods papers. To qualify as a SW-CRT, the

design had to use cluster randomization and have a minimum of two sequences and three periods. Non-randomized or quasi-experimental designs and pilot or feasibility studies (as stated by the authors) were excluded. We also excluded designs randomizing fewer than five clusters, even if not described as pilot or feasibility studies, as inferences that can be drawn from such trials would be limited. Finally, SW-CRTs with more than one evaluated intervention were excluded as implementation issues become more complex in this case.

2.3. Selection of articles

Identified references were saved and managed using Zotero 5.0. Duplicates were removed. Two reviewers (AC and MT) independently screened the titles and abstracts of the identified references to assess eligibility. If necessary, full-text articles were searched and screened. Any reference not meeting eligibility criteria was excluded and the reason for exclusion was recorded. Any discrepancies in the eligibility of a study were resolved by discussion, with the help of a third reviewer (AD) if needed to reach a consensus.

2.4. Data extraction and management

Two reviewers independently extracted the data in a random computer-generated order (AC and one of ALM, FLVA, and MT). We used a data extraction form tested and revised using 4 trials. Any discrepancies in data extraction were resolved by discussion between the two reviewers or with the help of a third reviewer to reach a consensus (AC and two of ALM, FLVA, and MT). We attempted to access the protocol and any cited secondary analysis paper to collect complete information on the trials as planned and realized. When no protocol was publicly available, we emailed the corresponding author to request the protocol. We collected and managed extracted data using Airtable (Airtable, San Francisco, California) [11].

We extracted the following characteristics for each selected trial (definitions of methodological elements used in data extraction are provided in Table 1):

- General characteristics: journal, publication year, location of study recruitment, type of clusters, any reported rationale for the SW-CRT design, any reported prior pilot or feasibility study;
- Design characteristics: timing of randomization (at a single time point, in batches or unclear), type of SW-CRTs design (cross sectional, closed cohort or open cohort), whether there was prospective recruitment of participants, source of outcome data collection (exclusively routinely collected data or not), planned and actual number of participants (or observations), number of sequences, number of clusters, number of periods, duration of the trial, complete or incomplete design, and allowance for a transition period;
- Intervention condition: typology of the intervention (targeted at the organization of health care or health delivery service, at health care professionals, direct participant therapeutic intervention, participant health promotion or education intervention), level of the intervention target (cluster, individual or both levels);
- Control condition (usual care or other);
- Results on the primary outcome (positive or negative).

Table 1. Definitions of stepped-wedge cluster randomized trial key methodological elements

Variables	Definition
Types of SW-CRT design	Depending on whether measurements taken in the different periods within a cluster come from the same or different participants, the design is classified as closed cohort, open cohort or cross-sectional type: **Closed cohort design**: all participants are recruited at the beginning of the trial, they are repeatedly assessed over multiple measurement points and cannot join the trial as it is ongoing. **Open cohort design**: almost all participants are recruited at the beginning of the trial, they are repeatedly assessed over multiple measurement points and can join or leave the trial as it is ongoing. **Cross-sectional design**: measurements come from different participants at each measurement point.
Complete design	Measurements are taken in each cluster-period
Incomplete design	Measurements are deliberately omitted in some cluster-periods
Transition period	Time to embed the intervention, with no performed measurement
Primary outcome	Main or primary outcome specified by the authors; if not specified, we used the outcome reported for sample size calculation. In case no primary outcome or no sample size calculations were reported or no unique primary outcome could be identified, we considered the first outcome mentioned in the Methods section of the manuscript
Positive result on the primary outcome	Statistically significant difference in favor of the intervention condition
Negative result on the primary outcome	Non significant difference or statistically significant difference in favor of the control condition

We classified a trial as having a recruitment challenge if (i) it did not reach its planned number of clusters, and/or (ii) it did not reach its planned number of participants with a 10% allowed margin (i.e. less than 90% of the target sample size recruited), and/or (iii) it clearly reported that design changes were made in response to recruitment difficulties. We considered only the total number of participants because the number of participants per cluster-period was seldom reported. We defined implementation challenge as any of (i) early, (ii) delayed or (iii) no implementation of the intervention in one or more clusters. Implementation challenges were either self-reported by the authors or identified by the reviewers, especially by comparing the diagrams in the protocol and report of the trial. We defined modification of the trial design as any of: deviation from (i) the planned number of sequences, (ii) the planned number of clusters, (iii) the planned number of periods, or (iv) the planned duration (allowing

for a one-month margin), or (v) change from a complete to an incomplete design and vice-versa, or (vi) addition or withdrawal of a transition period. We considered modifications that were made after the trial initiation in trials for which we identified recruitment or implementation challenges but also in other trials as such modifications can be related to unreported challenges. In case of identified recruitment or implementation challenges, we extracted whether a reason was reported and whether this challenge was accounted for in the analysis strategy (in the primary analysis or any sensitivity analysis).

2.5. Data analysis

Categorical data were summarized using frequency and percentage and quantitative data using mean and standard deviation or median and interquartile range, as appropriate. We described the characteristics of trials with and without recruitment or implementation challenge without performing any statistical tests of association because of the small sample sizes. We considered the following trial characteristics that we thought *a priori* might be associated with challenges: number and type of clusters, availability of a protocol, previous pilot study, trial design, number of participants, prospective recruitment, allowance for a transition period, type and level of experimental intervention, method for data collection. We also explored whether recruitment or implementation challenges were associated with modifications of the planned design and with positive results for the primary outcome. Statistical analyses were performed with SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

3. Results

3.1. Search results

Among the 562 references identified from PubMed, three were duplicates and 440 were excluded as ineligible based on title and abstract. Of the 119 references assessed on full-text, we excluded 64 that did not meet our eligibility criteria (Fig. 1).

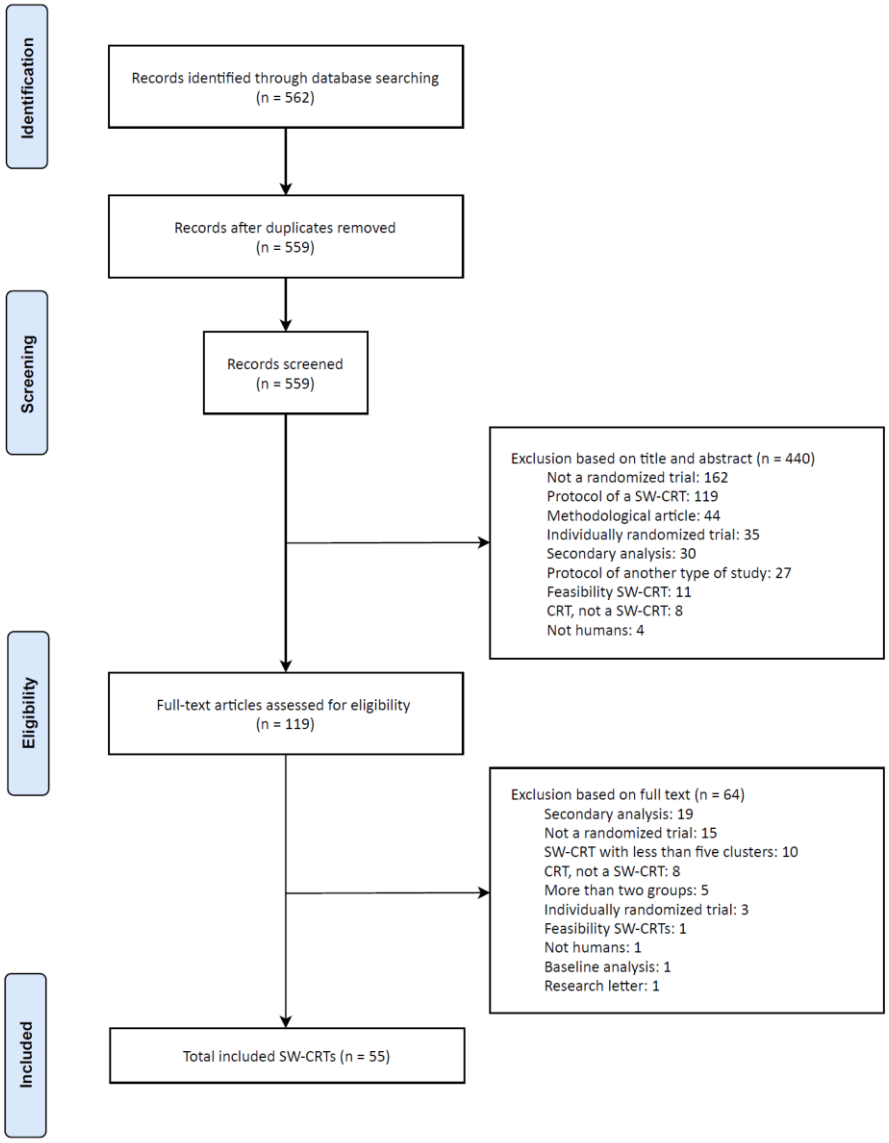

Figure 1. Flow diagram of the selection process

3.2. Characteristics of included studies

The 55 included SW-CRTs were published in 42 journals with impact factor ranging from 1.85 to 79.32; their general characteristics are reported in Table 2.

Table 2. Characteristics of stepped-wedge cluster randomized trials included in the review

GENERAL CHARACTERISTICS	N=55
Publication year	
2019	30 (54.5)
2020	25 (45.5)
Location of study recruitment[a]	
North America	15 (27.3)
Europe	12 (21.8)
Asia	10 (18.2)
Oceania	9 (16.4)
Africa	7 (12.7)
Central America/South America/Caribbean	4 (7.3)
Type of clusters	
Hospitals or hospital wards	31 (56.4)
Primary care practices	12 (21.8)
Geographical areas (e.g., villages)	5 (9.1)
Nursing homes	4 (7.3)
Other[b]	3 (5.4)
Rationale for the stepped-wedge design reported	
Yes[c]	47 (85.5)
Logistical/Practical reasons	26 (55.3)
Desire that all clusters receive the intervention	25 (53.2)
Statistical reasons (e.g., power considerations)	24 (51.1)
Ethical/Equity reasons	13 (27.7)
Opportunistic, intervention to be implemented anyway	7 (14.9)
Facilitate recruitment of clusters/participants	4 (8.5)
Other	3 (6.4)
No	8 (14.5)
Available protocol	
Yes	44 (80.0)
No	11 (20.0)
Prior pilot study	
Yes	18 (32.7)
No or unclear	37 (67.3)
Realized design features	
Number of sequences	5 [4;7]
Number of clusters	11 [7;18]
Number of periods[e]	6 [5;10]
Number of participants[f]	7635 [1023;32194]
Allowed for a transition period	10 (18.2)
Complete design[g]	48 (90.6)
RANDOMIZATION, DESIGN, OUTCOME DATA COLLECTION	
Timing of randomization	
At a single time point	47 (85.5)
In batches	3 (5.5)
Unclear	5 (9.1)
Trial design	
Cross sectional	46 (83.6)
Continuous recruitment	42 (91.3)
Fixed time point recruitment	4 (8.7)
Open cohort	5 (9.1)
Closed cohort	4 (7.3)
Prospective recruitment of participants	

Yes	24 (43.6)
No	31 (56.4)
Exclusively routinely collected data	
Yes	26 (47.3)
No	29 (52.7)
INTERVENTION AND CONTROL CONDITIONS	
Type of experimental intervention[i]	
Targeted at the organization of health care or health delivery service	26 (47.3)
Targeted at health care professionals	19 (34.5)
Direct participant therapeutic intervention	7 (12.7)
Participant health promotion or educational intervention	3 (5.5)
Level at which intervention is delivered	
Cluster level	38 (69.1)
Individual level	7 (12.7)
Both cluster and individual levels	10 (18.2)
Type of control group	
Usual care	52 (94.5)
Other[h]	3 (5.5)
RESULTS	
Results for the primary outcome	
Positive	20 (36.4)
Negative	35 (63.6)

Data are expressed as number and percentage, n (%) or median [interquartile range]. Percentages may not total 100% due to rounding.

[a]One trial was performed over 3 locations: Africa, South America, Asia, so does not sum to 55.

[b]Other cluster types are mental healthcare service providers (n=1), surgeons (n=1), reception centers for asylum seekers (n=1).

[c]A trial can report multiple reasons, so does not sum to 47.

[d]Only the most prominent was considered.

[e]Two cases were missing.

[f]Four cases were missing.

[g]A complete design implies measurement in every cluster-period of the study. Two cases were missing.

[h]In two trials, control group consisted of minimal application of the experimental intervention and in one trial, it was an attention control intervention.

A rationale for the stepped-wedge design was provided in 47 studies (86%). Notably, logistical reasons were reported in 26 (55%) and desire that all clusters receive the intervention in 25 (53%). A prior pilot study was mentioned in 18 (33%). A protocol was available for 44 studies (80%) (including seven obtained after contacting the authors). More than half of the studies randomized hospitals or hospital wards and the vast majority randomized clusters at a single time point. The most frequent design for inclusion of participants was cross-sectional (n=46, 84%) and 24 (44%) studies had a prospective recruitment of participants. Routinely collected data were exclusively used in nearly half (26, 47%). Interventions were delivered only at the cluster level in 38 studies (69%). Results from the main analysis of the primary outcome were negative for 35 (64%) of the studies: either non-significant (n=31) or significant in favor of the control group (n=4).

3.3. Recruitment and implementation challenges

Thirty-five trials (64%) had at least one of recruitment or implementation challenge. Seven trials (13%) had both recruitment and implementation challenges. Among the 35 trials with recruitment or implementation challenge, 18 (51%) had a modification in their design.

3.3.1. Frequency and description of recruitment challenges

A recruitment challenge was identified in 18 (33%) trials; under-recruitment of clusters occurred in one trial, under-recruitment of participants in 13 trials and four trials reported a design adaptation to address recruitment issues (Table 3). In 14 (26%) trials, we had insufficient data to judge whether there were recruitment challenges. Among the 31 trials for which we had the number of participants in the control and intervention conditions reported separately, under-recruitment of participants occurred more frequently in the intervention condition, n=15 (48%) than in the control condition, n=12 (39%). Additional details are

provided in Appendix B and C. Reasons for recruitment challenges were reported in 7 trials; in 3 trials it was because of fewer than expected eligible participants (Appendix D).

Table 3. Description of recruitment and implementation challenges in included stepped-wedge cluster randomized trials

Characteristics	N=55
Recruitment or implementation challenge	35 (63.6)
Both recruitment and implementation challenges	7 (12.7)
Any recruitment challenge[a]	
Yes	18 (32.7)
None identified	23 (41.8)
Insufficient information to identify any recruitment challenges	14 (25.5)
Type of recruitment challenges encountered	
Under-recruitment of clusters	1
Under-recruitment[b] of participants	13
Design adaptation to address recruitment issues	4
Any implementation challenge[c]	
Yes	24 (43.6)
No, clearly reported that there were no implementation challenges	8 (14.6)
Insufficient information to identify any implementation challenges	23 (41.8)
Type of implementation challenges encountered[d]	
Early implementation	8
Delayed implementation	18
No implementation	2

[a]Recruitment challenge was defined as a planned number of clusters not achieved, a planned number of participants not achieved (less than 90% of the target number of participants recruited), or reported changes made to the design to achieve the planned number of participants.
[b]Recruitment or identification in case there was no prospective recruitment of participants
[c]Implementation challenge includes early, late or no implementation of the intervention in clusters that did not drop out of the study.
[d]A trial can report multiple implementation challenges, so does not sum to 24.

3.3.2. Modifications of the design and adaptation of the analysis strategy in trials with recruitment challenges

Besides the four trials already classified as having a recruitment challenge based on reported modifications of the trial design, six other trials had modifications of their design. The most common modification was change of the trial duration (n=8), either extension of the trial duration (n=6) or shortening of the trial duration (n=2). We also identified modification of the number of clusters (n=3) and periods (n=4) (Appendix D). In one trial, investigators implemented a rule after trial initiation to allow each cluster to move to their next period once

70% of their target sample size for the ongoing period was attained, leading to variable cluster-period lengths and unpredictable trial duration [12].

An adaptation of the analysis strategy was reported in four trials; three performed sensitivity analyses and one trial adapted the main analysis by excluding the period affected by inclusion issues. In one trial, sensitivity analyses to assess the impact of recruitment challenges qualitatively modified the results (from negative to positive)[13].

3.3.3. Factors associated with recruitment challenges

Trial characteristics according to the presence or absence of recruitment challenges are described in Table 4. All trials with a recruitment challenge had a cross-sectional design and more than half recruited participants prospectively. The median planned sample size was greater in trials with recruitment challenges. A pilot study was performed in 44% of the trials with recruitment challenges as compared to 17% of the trials without recruitment challenges. Trials with a recruitment challenge more often allowed for a transition period (33%) than trials without recruitment challenges (5%).

Table 4. Characteristics of stepped-wedge cluster randomized trials included in the review according to the presence or absence of recruitment or implementation challenges

Characteristics	Recruitment challenges present?[a]		Implementation challenges present?[b]	
	Yes (n=18)	No (n=23)	Yes (n=24)	No (n=8)
Type of clusters				
Hospitals or hospital wards	12 (66.7)	13 (56.5)	14 (58.3)	3 (37.5)
Primary care practices	4 (22.2)	1 (4.4)	5 (20.8)	0
Nursing homes	0	3 (13.0)	0	3 (37.5)
Geographical areas	0	5 (21.7)	3 (12.5)	1 (12.5)
Other	2 (11.1)	1 (4.4)	2 (8.3)	1 (12.5)
Available protocol				
Yes	14 (77.8)	21 (91.3)	22 (91.7)	7 (87.5)
No	4 (22.2)	2 (8.7)	2 (8.3)	1 (12.5)
Prior pilot study				
Yes	8 (44.4)	4 (17.4)	9 (37.5)	3 (37.5)
No or unclear	10 (55.6)	19 (82.6)	15 (62.5)	5 (62.5)
Number of clusters	11 [7;19]	12 [6;16]	11.5 [6.5;17.0]	8.5 [6.5;11.0]
Number of participants, as planned	3200 [915;14000][c]	1800 [640;32400]	3000 [960;32400][d]	1780 [680;6763]
Trial design				
Cross sectional	18 (100.0)	17 (73.9)	21 (87.5)	5 (62.5)
Open cohort	0	3 (13.0)	0	2 (25.0)
Closed cohort	0	3 (13.0)	3 (12.5)	1 (12.5)
Prospective recruitment of participants				
Yes	10 (55.6)	13 (56.5)	13 (54.2)	2 (25.0)
No	8 (44.4)	10 (43.5)	11 (45.8)	6 (75.0)
Allowed for a transition period, as realized				
Yes	6 (33.3)	1[c] (4.5)	4 (16.7)	3 (37.5)
No	12 (66.7)	21 (95.5)	20 (83.3)	5 (62.5)
Type of experimental intervention				
Targeted at the organization of health care or health delivery	9 (50.0)	8 (34.8)	8 (33.3)	6 (75.0)
service	5 (27.8)	9 (39.1)	12 (50.0)	2 (25.0)
Targeted at health care professionals	2 (11.1)	5 (21.7)	2 (8.3)	0
Direct participant therapeutic intervention	2 (11.1)	1 (4.4)	2 (8.3)	0
Participant health promotion or educational intervention				

Level at which intervention is targeted				
Cluster level	11 (61.1)	14 (60.9)	16 (66.7)	6 (75.0)
Individual level	3 (16.7)	4 (17.4)	3 (12.5)	0
Both cluster and individual levels	4 (22.2)	5 (21.7)	5 (20.8)	2 (25.0)
Exclusively routinely collected data				
Yes	7 (38.9)	7 (30.4)	12 (50.0)	4 (50.0)
No	11 (61.1)	16 (69.6)	12 (50.0)	4 (50.0)
Modification of the design				
Yes	10 (55.6)	8 (34.8)	14 (58.3)	3 (37.5)
No	8 (44.4)	15 (65.2)	10 (41.7)	5 (62.5)

Data are expressed as number and percentage, n (%) or median [interquartile range]. Percentages may not total 100% due to rounding.

[a]14 trials in which there was insufficient information to identify any recruitment challenge were excluded. Recruitment covers recruitment and identification in case there was no prospective recruitment of participants.

[b]23 trials in which there was insufficient information to identify any implementation challenges were excluded.

[c]Information was missing for one trial

[d]Information was missing for five trials

3.3.4. Frequency and description of implementation challenges

Implementation challenges were identified in 24 (44%) trials: 18 had delayed implementation, 8 had early implementation and two had no implementation of the intervention in some clusters (Table 3). In 23 trials (42%), we had insufficient information to judge whether there were implementation challenges and in eight (15%) it was clearly reported that there were no implementation challenges. Reasons for implementation challenges were reported in 18 trials with three reporting multiple reasons (Appendix D). The main reasons for delayed implementation were logistical or technical issues (6 trials), issues with staff — staff turnover, strike or implementation planned during holidays (5 trials), lower than expected recruitment (4 trials) and approval issues (2 trials). In two trials, intervention was implemented early because it became standard of care during the trial. In one trial, some clusters refused to implement the intervention and in another, the intervention was not rolled out in three clusters because of technical issues.

3.3.5. Modifications of the design and adaptation of the analysis strategy in trials with implementation challenges

Among the 24 trials with implementation challenges, 14 reported a modification of their design such as extension of the trial duration (n=7), shortening of the trial duration (n=4), addition of periods (n=4), addition of clusters (n=2) (Appendix D).

Five trials with implementation challenges adapted their analysis strategy: in two the main analysis was performed using an "as implemented" strategy; in three trials, sensitivity analyses were performed to assess the impact of implementation challenges on the intervention effect and found consistent results with the main analysis (one positive and two negative results).

In three trials, the authors had anticipated the risk of implementation challenges at the planning stage of the trial: in two trials, they planned to perform sensitivity analyses in case of deviation from the planned implementation schedule; in one they planned to conduct an analysis based on the randomization schedule regardless of the actual date of implementation. Among these three trials, only one actually reported implementation challenges [14].

3.3.6. Factors associated with implementation challenges

Trials with implementation challenges more often took place in a healthcare setting -hospitals, wards or practices- (79%) than trials without implementation challenges (38%) (Table 4). They more often had prospective recruitment of participants (54% vs. 25%) and less often allowed for a transition period (17% vs. 38%).

3.4. Other modifications of the design

Seven trials without recruitment or implementation challenges had at least one modification of the planned design. The most frequent modification was deviation from the planned duration of the study, which was identified in five of those studies and modification in the number of periods in three. Additional details are provided in Appendix D.

4. Discussion

Summary of key findings and comparison with other studies

To our knowledge, this is the first methodological review to assess recruitment and implementation challenges in recent SW-CRTs. We found that nearly two-thirds of SW-CRTs had recruitment or implementation challenges. More than half of trials with challenges modified their planned design — most often the trial duration. In some trials, recruitment difficulties led to design modifications and deviation from the implementation schedule.

A previous review found that 33% of SW-CRTs did not reach their prespecified sample size, defined as less than 100% of the planned number of participants [9]. Our definition of under-recruitment was more permissive, allowing for a 10% margin. Using the same definition as the previous study, the prevalence of under-recruitment in our review was 43%, thus showing no improvement in recent SW-CRTs [9]. Moreover, our definition of recruitment challenge was broader than only under-recruitment of participants and included under-recruitment of clusters and design modifications in response to recruitment difficulties. The same review found that 43% of SW-CRTs had one or more difficulties during the study roll out, data collection or analysis but it did not specifically focus on implementation challenges [9]. Another article examining six case studies reported that staggered implementation of SW-CRTs raises new practical challenges of adhering to the planned schedule [7].

Our results also highlight that planned and actual designs are poorly reported, preventing the identification of recruitment or implementation challenges as well as other changes made after trial initiation. Although the extension of the CONSORT statement for SW-CRTs clearly recommends the reporting of changes to methods after trial commencement, it seems that there is still room for improvement [6]. We identified seven studies with no identified recruitment or implementation challenges but with some modifications of their designs; such modifications could reflect challenges that were not reported. Due to poor reporting of reasons in most of the trials, we were unable to judge whether recruitment challenges were preventable. One or multiple reasons were identified in 75% of trials with implementation challenges, some preventable (such as holiday periods) but others completely unpredictable (such as strike or concurrent major reform). Of note, SW-CRTs included in this review were not impacted by the corona pandemic (the most recent ended in August 2019) but such major event certainly challenged ongoing trials at that time.

Our small sample sizes did not allow us to test whether trial characteristics were associated with recruitment or implementation challenges. Nevertheless, while the use of a feasibility study prior to a SW-CRT has been proposed as a solution to avoid practical difficulties [15], we observed more recruitment challenges in trials with a prior pilot study than in those without. Our analysis cannot rule out the possibility that pilot studies were more frequently performed in highly complex trials prone to recruitment challenges. The use of a transition period was more frequently reported in trials with identified recruitment challenges but less often in trials with implementation challenges. Contrary to our hypothesis, we observed similar prevalence of prospective recruitment and exclusive use of routinely collected data among trials with and without recruitment challenges.

When there are implementation challenges, interpretation of the results is complex. The analysis strategy should be prespecified, including how deviations from the planned implementation schedule will be handled. Delayed implementation of the intervention and extension of the trial duration can also happen in a two-parallel group CRT but such deviations are more problematic in a SW-CRT because it affects unevenly intervention and control observations [16]. A first simulation study has shown that early implementation of the intervention can lead to biased intervention effect estimates which may be addressed by using models incorporating fixed or random group-by-time effects [17]. Implementation and recruitment challenges are also likely to impact power but further work is needed to explore how such challenges can affect the results of SW-CRTs.

Recommendations for future studies

Although several trials did anticipate practical challenges in their protocol, trialists may not fully appreciate the associated logistical complexities, need for time-constrained recruitment and likelihood of deviations from the implementation schedule. We suggest some

recommendations for the planning and reporting of future SW-CRTs (Table 5). These recommendations for planning and reporting can be used by trialists as well as journal editors to appraise SW-CRT protocol and results, respectively.

Limitations

Our study has several limitations. The search strategy might have missed some SW-CRTs. However, our aim was not to be exhaustive but to provide information on a sample of recent SW-CRTs. Due to poor reporting of planned and actual design in many trials, it was difficult to identify recruitment or implementation challenges and we had to add a category for trials with insufficient information. Our definition of implementation challenges focused on deviations from the planned timing and did not include elements related to fidelity of the delivered intervention. Indeed, core components of the intervention were hardly ever reported so it was impossible to judge whether clusters were fully exposed to the intervention or not. As we included trials published close to the publication of the CONSORT extension for SW-CRTs [6], most of the included trials were probably designed beforehand and one would expect some improvements in future trials. Finally, sample sizes were small making it difficult to explore factors associated with recruitment or implementation challenges.

Conclusions

Recruitment or implementation challenges are frequent in SW-CRTs. The theoretical advantages of a SW-CRT might be compromised by their organizational and logistical time-constrained requirements. Our practical recommendations may help researchers to enhance the design and reporting of future SW-CRTs.

Table 5. Recommendations for planning and reporting future SW-CRTs

Stage	Recommendations
Planning	- Write a precise roadmap for implementation of the intervention in clusters, especially when the study team has to implement the intervention in several clusters simultaneously, and ensure that the planned implementation schedule is practicable - Do not underestimate the likelihood of recruitment and implementation challenges during major holidays and other events - Obtain all necessary REC and gatekeeper approvals for all clusters before the beginning of the trial - Before randomization, obtain agreement — preferably in writing — from all participating clusters about the implementation schedule including the required lead time to prepare for implementation - Plan and pre-specify the analysis strategy to explain how any possible deviations from the implementation and recruitment schedule will be handled in the primary and secondary analyses
Reporting	- Clearly report the planned and actual number of participants (per condition and per cluster-period, when appropriate) - Clearly report the planned and actual schedule for implementation of the intervention - Rather than merely state that the analysis was conducted on an intention-to-treat basis, clearly report the analysis population and explain how deviations from the planned schedule were handled - In case of recruitment or implementation challenges or modification of the planned design, report the reasons to allow an assessment of implications for risks of bias

REC, research ethics committee

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare they have no competing interests.

CRediT authorship contribution statement

Agnès Caille: Conceptualization, Formal analysis, Data curation, Writing - original draft. **Monica Taljaard**: Conceptualization, Data curation, Writing - original draft. **Floriane Le Vilain -- Abraham**: Data curation, Writing – review and editing. **Alexis Le Moigne**: Data curation, Writing – review and editing. **Andrew J Copas**: Conceptualization, Writing – review and editing. **Florence Tubach**: Conceptualization, Supervision, Writing - original draft. **Agnès Dechartres**: Conceptualization, Data curation, Supervision, Writing - original draft.

References

[1] Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.

[2] Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015;350:h391.

[3] Campbell MJ, Hemming K, Taljaard M. The stepped wedge cluster randomised trial: what it is and when it should be used. Med J Aust 2019;210:253-254.e1. https://doi.org/10.5694/mja2.50018.

[4] Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? Int J Epidemiol 2020;49:1043–52. https://doi.org/10.1093/ije/dyaa077.

[5] Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007;28:182–91. https://doi.org/10.1016/j.cct.2006.05.007.

[6] Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ 2018;363:k1614. https://doi.org/10.1136/bmj.k1614.

[7] Prost A, Binik A, Abubakar I, Roy A, De Allegri M, Mouchoux C, et al. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. Trials 2015;16:351. https://doi.org/10.1186/s13063-015-0837-4.

[8] Adrion C, Weiss B, Paul N, Berger E, Busse R, Marschall U, et al. Enhanced Recovery after Intensive Care (ERIC): study protocol for a German stepped wedge cluster randomised controlled trial to evaluate the effectiveness of a critical care telehealth program on process quality and functional outcomes. BMJ Open 2020;10:e036096. https://doi.org/10.1136/bmjopen-2019-036096.

[9] Eichner FA, Groenwold RHH, Grobbee DE, Oude Rengerink K. Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size. J Clin Epidemiol 2019;107:89–100. https://doi.org/10.1016/j.jclinepi.2018.11.013.

[10] Hemming K, Carroll K, Thompson J, Forbes A, Taljaard M, SW-CRT Review Group. Quality of stepped-wedge trial reporting can be reliably assessed using an updated CONSORT: crowd-sourcing systematic review. J Clin Epidemiol 2019;107:77–88. https://doi.org/10.1016/j.jclinepi.2018.11.017.

[11] Airtable. Available at: n.d. [https://airtable.com/] (accessed July 30, 2021).

[12] Schwarze ML, Buffington A, Tucholka JL, Hanlon B, Rathouz PJ, Marka N, et al. Effectiveness of a Question Prompt List Intervention for Older Patients Considering Major Surgery: A Multisite Randomized Clinical Trial. JAMA Surg 2019. https://doi.org/10.1001/jamasurg.2019.3778.

[13] Raphaelis S, Frommlet F, Mayer H, Koller A. Implementation of a nurse-led self-management support intervention for patients with cancer-related pain: a cluster randomized phase-IV study with a stepped wedge design (EvANtiPain). BMC Cancer 2020;20:559. https://doi.org/10.1186/s12885-020-06729-0.

[14] Lenguerrand E, Winter C, Siassakos D, MacLennan G, Innes K, Lynch P, et al. Effect of hands-on interprofessional simulation training for local emergencies in Scotland: the THISTLE stepped-wedge design randomised controlled trial. BMJ Qual Saf 2020;29:122–34. https://doi.org/10.1136/bmjqs-2018-008625.

[15] 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;19:12. https://doi.org/10.1186/s12874-019-0658-3.
[16] Neal B, Wu Y, Feng X, Zhang R, Zhang Y, Shi J, et al. Effect of Salt Substitution on Cardiovascular Events and Death. N Engl J Med 2021;385:1067–77. https://doi.org/10.1056/NEJMoa2105675.
[17] Rennert L, Heo M, Litwin AH, Gruttola VD. Accounting for confounding by time, early intervention adoption, and time-varying effect modification in the design and analysis of stepped-wedge designs: application to a proposed study design to reduce opioid-related mortality. BMC Med Res Methodol 2021;21:53. https://doi.org/10.1186/s12874-021-01229-6.
[18] Skinner EH, Lloyd M, Janus E, Ong ML, Karahalios A, Haines TP, et al. The IMPROVE-GAP Trial aiming to improve evidence-based management of community-acquired pneumonia: study protocol for a stepped-wedge randomised controlled trial. Trials 2018;19:88. https://doi.org/10.1186/s13063-017-2407-4.
[19] Lloyd M, Karahalios A, Janus E, Skinner EH, Haines T, De Silva A, et al. Effectiveness of a Bundled Intervention Including Adjunctive Corticosteroids on Outcomes of Hospitalized Patients With Community-Acquired Pneumonia: A Stepped-Wedge Randomized Clinical Trial. JAMA Intern Med 2019;179:1052–60. https://doi.org/10.1001/jamainternmed.2019.1438.