Considerations Before Selecting a Stepped-Wedge Cluster Randomized Trial Design for a Practice Improvement Study

ABSTRACT

PURPOSE Despite the growing popularity of stepped-wedge cluster randomized trials (SW-CRTs) for practice-based research, the design’s advantages and challenges are not well documented. The objective of this study was to identify the advantages and challenges of the SW-CRT design for large-scale intervention implementations in primary care settings.

METHODS The EvidenceNOW: Advancing Heart Health initiative, funded by the Agency for Healthcare Research and Quality, included a large collection of SW-CRTs. We conducted qualitative interviews with 17 key informants from EvidenceNOW grantees to identify the advantages and challenges of using SW-CRT design.

RESULTS All interviewees reported that SW-CRT can be an effective study design for large-scale intervention implementations. Advantages included (1) incentivized recruitment, (2) staggered resource allocation, and (3) statistical power. Challenges included (1) time-sensitive recruitment, (2) retention, (3) randomization requirements and practice preferences, (4) achieving treatment schedule fidelity, (5) intensive data collection, (6) the Hawthorne effect, and (7) temporal trends.

CONCLUSIONS The challenges experienced by EvidenceNOW grantees suggest that certain favorable real-world conditions constitute a context that increases the odds of a successful SW-CRT. An existing infrastructure can support the recruitment of many practices. Strong retention plans are needed to continue to engage sites waiting to start the intervention. Finally, study outcomes should be ones already captured in routine practice; otherwise, funders and investigators should assess the feasibility and cost of data collection.

INTRODUCTION

A s a burgeoning study design in health services research, stepped-wedge cluster randomized trials (SW-CRTs) can have advantages over parallel CRTs in terms of statistical power and offer a pragmatic approach to providing the intervention to all practices, which often aligns with practices’ priorities. The “CRT” in SW-CRT refers to clusters (eg, practices) being randomized to a sequence, which specifies the timing of crossover from one condition to another (ie, from control to intervention), as opposed to being randomized to study arms as in a parallel CRT. In other words, clusters are randomized to a sequence that determines when—not if—they receive the intervention, which makes the design appealing and relevant for quality improvement and practice transformation initiatives.

Figure 1 shows a sample SW-CRT design scheme. Traditionally, all clusters are recruited and enrolled at baseline and followed for the duration of the study. Outcomes are measured for every cell (ie, every time block for every cluster). Thus, all clusters participate in some way for the entire study period, at times only via data collection.

Despite the growing popularity of SW-CRTs for practice-based research, there are numerous considerations that should be made before selecting this design. Stepped-wedge cluster randomized trials come with challenges, many of which are not well documented, owing to lack of publications on their real-world application. Examining 2 systematic reviews of SW-CRT studies, only 14 SW-CRTs have been conducted in primary care settings, those study teams selected the design on the basis of resource constraints, methodologic preferences (eg, phased implementation,
all participants receive the intervention), or had no reason indicated. The largest of the studies included 72 practices, and only 1 was conducted in the United States. Our objective was to identify the advantages and challenges of the SW-CRT design for large-scale intervention implementations in primary care settings. We assessed SW-CRTs via EvidenceNOW (also known as Evidence-NOW: Advancing Heart Health), one of the largest practice improvement primary care studies funded by the Agency for Healthcare Research and Quality (AHRQ) to date and one of the largest collections of SW-CRTs. Other methodologic articles on SW-CRT have examined logistic, ethical, political, and statistical considerations across a broad range of settings. We examined the considerations for selecting SW-CRT for large-scale implementations, specifically in primary care settings.

**METHODS**

**Study Setting**

EvidenceNOW was designed to improve cardiovascular health care delivery in the United States, aiming to increase adoption of the “ABCS” cardiovascular disease prevention and treatment guidelines: Aspirin use by high-risk individuals, Blood pressure control, Cholesterol management, and Smoking cessation. The goal was to ensure that small to medium-sized primary care practices implement the latest evidence to decrease their patients’ cardiovascular disease risk and live longer, healthier lives. The AHRQ awarded 8 grants; 1 to a national evaluator (Evaluating System Change to Advance Learning and Take Evidence to Scale [ESCALATES]) and 7 thirty-six–month grants to regional cooperatives to study the use of external practice facilitation for implementing cardiovascular disease guidelines. Whereas each cooperative designed its own intervention, all used facilitation as a core implementation strategy, enrolled >200 primary care practices, and provided the intervention to all practices (see Supplemental Table 1 for intervention components by cooperative). In the program announcement, the AHRQ encouraged, but did not require, cooperatives to use the SW-CRT design; ultimately, the SW-CRT design was used by 4 of the 7 cooperatives. The 3 that did not use the SW-CRT design had regional coverage of >1 state. Table 1 provides an overview of the cooperatives including their study design selections.

**Study Design and Sample**

To identify the advantages and challenges of using SW-CRT design, we used the rapid assessment process, an “intensive, team-based qualitative inquiry using triangulation, iterative data analysis, and additional data collection.” We conducted semistructured interviews with all 8 grantees. We sent an e-mail invitation to the principal investigators and encouraged them to invite relevant team members (purposive snowball sampling);
all grantee principal investigators agreed to participate, with some electing to be interviewed alone, and others inviting up to 2 team members to join.

Interview guides (Supplemental Appendix 1) asked each grantee to share what worked, what challenges they experienced with their study design, and lessons learned from using their design. Participants from ESCALATES were asked to reflect on experiences harmonizing data from the different study designs across cooperatives. Each interview had 1 primary interviewer (A.M.N. [female] or M.P.B. [male]), with the other present to ask clarifying questions. Interviews were conducted by telephone or video conferencing, lasted approximately 30 minutes, and were audiorecorded with permission. The final sample comprised 17 key informants across the 8 grantees.

Analysis
We used rapid qualitative analysis techniques,21 which start with team members debriefing after each interview and populating a structured template (Supplemental Appendix 2) that corresponded with central topics of the interview guide. During the debriefing process, team members assessed data saturation, finding no new themes after interviewing the 6th grantee.21,23 Next, data were aggregated into a matrix (Supplemental Appendix 2) to compare preliminary themes across grantees. Finally, we reviewed and discussed the matrix at multiple meetings to determine themes as advantages or challenges of SW-CRT design. Results were shared with grantees for participant checking and feedback; grantees confirmed that their perspectives were captured accurately and completely.

RESULTS
All interviewees reported that SW-CRTs can be highly effective for large-scale intervention implementations. A key design strength of SW-CRTs is that all sites receive the intervention. Interviewees noted that if an intervention is expected to provide a benefit with minimal risk, it is “unethical not to do the intervention for all” (North Carolina cooperative). Under that shared belief, the 3 cooperatives that did not select the SW-CRT design selected the parallel CRT design

<p>| Table 2. Advantages and Challenges of Using SW-CRTs |</p>
<table>
<thead>
<tr>
<th>Adamant</th>
<th>Description</th>
<th>Representative Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentivized recruitment</td>
<td>All sites receive the intervention and serve as their own control.</td>
<td>“To ask a site to engage in research for a year without receiving resources doesn’t seem possible.” (New York City)</td>
</tr>
<tr>
<td>Staggered resource allocation</td>
<td>Resources (eg, personnel) can be allocated over a longer period.</td>
<td>“You can shift resources from one sequence to another, which eases workforce logistical concerns.” (ESCALATES)</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Power can be higher than parallel cluster randomized controlled trials under certain conditions.</td>
<td>“Stepped-wedge designs potentially have the power advantage over alternative designs, though the trade-off is the time issue.” (New York City)</td>
</tr>
<tr>
<td><strong>Challenge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-sensitive recruitment</td>
<td>Recruitment must occur at all sites up front.</td>
<td>“Recruitment challenges were the main reason we chose the parallel cluster randomized trial and not a stepped-wedge design.” (Midwest)</td>
</tr>
<tr>
<td>Retention</td>
<td>Sites might drop out owing to a long lag time between recruitment and the start of the intervention.</td>
<td>“By the time all partnership agreements were signed and sites were randomized, 47 had dropped out of the study.” (Northwest)</td>
</tr>
<tr>
<td>Randomization requirements and practice preferences</td>
<td>It might be difficult to randomize sites into sequences, given that real-world practice priorities are often changing.</td>
<td>“The real-world environment does not really respect the randomization. Stepped-wedge designs are complicated by the fact that they have defined start and stop dates. That’s not how quality improvement works.” (North Carolina)</td>
</tr>
<tr>
<td>Achieving treatment schedule fidelity</td>
<td>It might be difficult to deliver the intervention as prescribed (eg, sites might cross-talk across sequences).</td>
<td>“We held weekly meetings with all facilitators to deliberately talk about cross-contamination and staying with the timeline.” (Oklahoma)</td>
</tr>
<tr>
<td>Intensive data collection</td>
<td>Sites might have difficulty contributing data for specified outcome measures for every time block of the implementation timeline.</td>
<td>“Practice burden is usually greater for SW-CRTs (compared to other designs) in terms of measurement because every practice has to report every measure for every time block.” (Southwest)</td>
</tr>
<tr>
<td>Hawthorne effect</td>
<td>Sites might modify their behavior before the intervention begins.</td>
<td>“We anticipated that the sites would start preparing (before the intervention started).” (Midwest)</td>
</tr>
<tr>
<td>Temporal trends</td>
<td>Effect of intervention might be confounded by underlying temporal trends.</td>
<td>“Ideally, we would use the stepped-wedge design in a scenario where there aren’t significantly different covariates across clusters” (Southwest) and “when an outcome isn’t already expected to be improving.” (New York City)</td>
</tr>
</tbody>
</table>

ESCALATES = Evaluating System Change to Advance Learning and Take Evidence to Scale; SW-CRT = stepped-wedge cluster randomized trial.
or the 2×2 factorial design, which also allow for delivery of
the intervention to all sites. However, interviewees recom-
mended carefully weighing the advantages and challenges of
SW-CRT design (Table 2) before selecting this design, given
its numerous challenges, because deviations from the study
design might introduce bias into the analyses.

Advantages
The advantages of SW-CRT design were threefold: (1) incen-
tivized recruitment, (2) staggered resource allocation, and (3)
statistical power.

Incentivized Recruitment
Cooperatives each aimed to recruit 200-250 primary care
practices. As described by the New York City cooperative,
the guarantee that all study sites would receive the interven-
tion was an important incentive for practices to enroll. This
guarantee became important for recruiting many practices,
especially ones with whom the cooperative did not have an
existing relationship.

Staggered Resource Allocation
The SW-CRT design allows resources to be allocated over a
longer period, a key advantage for large-scale implementa-
tions for which there might be limited resources. Owing to the
staggered intervention start and end dates, resources, including
the implementation team, can be shifted from one sequence
to another, which eases workforce logistical concerns. In com-
parison, activities of parallel CRTs are condensed into a short
time frame and are thus more resource intense.

Statistical Power
The SW-CRT design can have a power advantage over
alternative designs, such as the parallel CRT design, when
the intracluster correlation is larger. Intracluster correlation
is larger when outcomes for a practice are more similar than
those across practices. Grantees acknowledged that this might
be difficult to determine beforehand, especially owing to
recruitment challenges (described below) and if there are elec-
tronic health record (EHR) inconsistencies across practices.

Challenges
Challenges of the SW-CRT design included (1) time-sensitive
recruitment, (2) retention, (3) randomization requirements
as opposed to practice preferences, (4) achieving treatment
schedule fidelity, (5) intensive data collection, (6) the Haw-
thorne effect, and (7) temporal trends.

Time-Sensitive Recruitment
Interviewees agreed that time-sensitive recruitment was the
most influential factor on their study design selection, given
that all practices need to be recruited up front for random-
ization. The SW-CRT design does not allow for staggered
recruitment; staggered recruitment prevents full randomiza-
tion. The funding for EvidenceNOW cooperatives was for
36 months. Recruitment was very challenging because of this
short time frame, in addition to the large volume of sites, par-
ticularly given that smaller practices tend to be independent,
making them difficult to reach. The Oklahoma cooperative
did not have an existing network from which to recruit and
ultimately extended their initial 3-month recruitment period
by 8 months. The New York City cooperative benefited from
partnering with large practice networks that had existing
relationships and communication and data infrastructures that
allowed them to identify and contact eligible practices.

Retention
The SW-CRT design involves lags between when sites are
recruited, randomized, and receive the intervention, leading
to site-retention challenges. Some cooperatives experienced
attrition between recruitment and randomization; the North-
west cooperative reported that 47 sites dropped out by the
time all partnership agreements were signed and sites random-
ized. Others lost sites randomized to later sequences, which
involved waiting more than a year before starting the inter-
vention. Grantees reflected that retention was a critical step
after recruitment. Cooperatives with recruitment networks in
place were able to shift efforts from recruitment to retention.

Randomization Requirements and Practice Preferences
The SW-CRT design has strict randomization requirements;
all practices must be enrolled before randomization, and
practices are assigned to staggered start dates. However,
practice priorities might not always align with the random-
ization schedule. The Northwest cooperative learned from
prior experience with SW-CRTs that sites often want to
start sooner rather than later or would not join unless they
received an early intervention. This was one reason why that
cooperative chose the 2×2 factorial design, which allows all
sites to begin the intervention at the same time. The North
Carolina cooperative had a different experience, in which
sites wanted to start later than when they were assigned,
owing to staffing or EHR changes. Discounting sites’ prefer-
ces put the cooperatives at risk of losing sites; however,
accounting for preferences subjected the study to unequal
distribution of site characteristics (eg, sites that start early
differ from those that start late).

Achieving Treatment Schedule Fidelity
There is risk of cross-contamination between sites in different
phases of the study (eg, across sequences), especially if sites
are from the same network or geographic region. The Vir-
iginia cooperative, which used SW-CRT design, opted to ran-
domize groups of practices as a block to contain any cross-
talk within sequences. There was also the risk that facilitators
working across multiple sequences were delivering the inter-
vention to sites that were in the control period. For example,
in New York City facilitators continued to visit sites in the
control period to deliver other programs that the network
leadership was implementing. The Oklahoma cooperative
attempted to decrease cross-contamination by strengthening training and quality control.

Intensive Data Collection
Interviewees reported that many sites had difficulty contributing data for every time block of the implementation timeline on the specified cardiovascular disease outcome measures. Complete data are necessary to adjust for underlying temporal trends. The Southwest cooperative referred to this as the measurement burden. In comparison, a parallel CRT does not require measurements across multiple time blocks and has a shorter time frame. Some sites did not have the technical capacity to pull quarterly data, and others did not have a systematic way to extract measures. An optimal condition might be one in which researchers have access to the data at the beginning as well as the ability to collect data from practices retrospectively via EHR data pulls.

Hawthorne Effect
In SW-CRTs, all sites are introduced to the intervention before their intervention starts, in some cases more than a year in advance. This might lead to the Hawthorne effect, which is when study subjects modify their behavior when made aware that they are being observed. The North Carolina cooperative might have experienced the effect more acutely than others, owing to its institutional policies, which required contracts be signed up front specifying the outcome measures of interest. Thus, sites knew which measures would be observed.

Temporal Trends
In SW-CRT design, more clusters receive the intervention toward the end of the study than in its early stages. Thus, the effect of the intervention might be confounded by an underlying temporal trend, especially if an outcome is already expected to improve over time. This consideration is particularly challenging for large-scale primary care studies, in which there can be variation across sites related to the breadth of the study and recruitment delays.

DISCUSSION
This study reports lessons learned from EvidenceNOW on the advantages and potential challenges associated with the SW-CRT design for large-scale intervention implementations in primary care settings. Overall, EvidenceNOW grantees considered SW-CRTs attractive for large-scale primary care research because it guarantees that all practices receive the intervention. Our findings suggest that recruitment is a major challenge for large-scale primary care studies, particularly when a study spans multiple states or lacks established networks from which to recruit. The guarantee that all practices receive the intervention is appealing because it might decrease barriers to recruitment among practices that do not value research engagement, especially when they are not guaranteed to receive intervention resources. From the implementer’s perspective, another advantage of the design is the opportunity to deliver the intervention in steps and over a longer period compared with other study designs, making it less resource intense. From the statistical standpoint, this design can be well powered under certain conditions. These reported advantages of SW-CRT design are consistent with earlier reviews.1,4,15 We extended the literature by identifying advantages that persist in large-scale primary care SW-CRTs. It is worth noting that not all of the advantages reported here are unique to large-scale SW-CRTs.

Ethics alone was not the presiding consideration for grantees selecting the SW-CRT design. As noted, 3 cooperatives selected alternate designs that allotted delivery of the intervention to all practices. Joag et al reported that the strongest arguments for selecting SW-CRT design are often political and logistical rather than ethical.16 As was the case in the present study, SW-CRT design was recommended by the funder, which might have affected grantees’ design selection. Cooperatives that deviated from using SW-CRT design did so to mitigate logistical challenges.

The reported challenges of SW-CRT design for use in large-scale primary care studies were related primarily to the long time frame of SW-CRTs, resulting in challenges with site retention, the heavy burden of data reporting, the Hawthorne effect, and possible confounding with temporal trends. In addition, SW-CRTs require that all sites be randomized at the start of the study. This creates burdens on practices not being able to choose when to start the intervention and on the study team to retain sites while they wait to receive the intervention. It is also possible that the perceived value of participating in the study is discounted over time, resulting in practices dropping out.

To address these challenges, EvidenceNOW grantees made recommendations for recruitment and retention strategies including increasing the recruitment budget, engaging stakeholders early to align research goals with practices’ priorities, and maintaining consistent communication.24-27 Our findings also suggest that implementers consider using data already routinely collected by the practice, which might mitigate the Hawthorne effect while making participation less onerous to the practice. During site selection, implementers should consider whether a practice has the capacity at the start to generate data needed for the trial; if not, allocate resources from the research budget so any burden associated with modifying data infrastructure and collection does not fall on the practice. Practices’ EHR functionalities might also hinder the intensive data collection process28; long-term solutions might require systemic advancements in EHR functionalities. Finally, to mitigate confounding from temporal trends, implementers might consider using fewer sequences, using an external comparison group, or collecting an associated baseline covariate to help understand sources of variance.17

The above-reported challenges of SW-CRT design resonate with the literature in primary care1,4,15 and other fields.29,30 However, grantees did not report challenges with
changes in data quality among practices in a long control period, as reported by Handley et al. It is possible that this was not experienced by grantees because the outcome measures were ones already captured by the practices. The grantees reported possible data quality issues owing to suspected Hawthorne effects, which is a related but novel finding.

Finally, statistical analysis of data generated by SW-CRT design is complicated by the partial confounding of intervention effects with time as well as clustering of observations (eg, repeated measures on individual patients within primary care practices). Major analytic approaches to address these complexities include mixed-effects regression models, generalized estimating equations, and robust nonparametric methods. Other complexities include delayed onset of intervention effects (the full effect is not observed in the first intervention period) or intervention-effect heterogeneity across sites or time (eg, sites with intervention onset later in calendar time experience smaller effects than sites with earlier onset, owing to factors external to the trial). Challenges such as changes in intervention effects over time might be more likely in SW-CRTs because they generally take longer than alternatives. These complexities should be weighed when designing SW-CRTs and considering alternatives.

Limitations
The present study has limitations. We can only make conclusions for studies that enroll primary care practices as the unit of enrollment and randomization. However, we believe the identified themes are high level and might apply broadly to health services and organizational research. The qualitative data interpretation might have been influenced by investigator bias. We took steps to minimize bias and confirm accuracy by checking interpretation of findings across all grantees. However, some bias might persist given their position as grantees. Alternative approaches that ensure confidential 1-on-1 interviews might have resulted in different or additional insights.

CONCLUSION
The challenges experienced by EvidenceNOW grantees suggest that certain favorable real-world conditions increase the odds of successful use of the SW-CRT design for large-scale intervention implementations. First, SW-CRTs might be more feasible when there are many practices in the region, when there is existing infrastructure to support recruitment, and/or when the implementation period is shorter so that there is less waiting time for practices that are randomized to later sequences. Second, there needs to be a comprehensive recruitment and retention plan in place. Third, strategies are needed to minimize the burden of capturing data at multiple time points from all study sites associated with the design. The feasibility and cost of data collection should be determined at the outset, and if the outcomes are not automatically captured in routine practice, researchers and funders might need to reconfigure the data collection process. Before specifying SW-CRT as the study design—particularly for large-scale intervention implementations for which the stakes might be high—researchers and funders should consider whether the study conditions are conducive for SW-CRT design. It is then up to the study team to determine whether the advantages outweigh the challenges.

Read or post commentaries in response to this article.

Key words: stepped wedge cluster randomized trial; practice improvement; study design; implementation; qualitative

Submitted September 2, 2020; submitted, revised, September 1, 2021; accepted September 30, 2021.

Funding support: This project was funded under grant no. 1R18HS023922 from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services (HHS). The authors are solely responsible for this document’s contents, findings, and conclusions, which do not necessarily represent the views of the AHRQ. Readers should not interpret any statement in this report as an official position of the AHRQ or of the HHS.

Acknowledgments: We thank program officer Robert McNellis, MPH, PA, for his insights and support in this article and throughout the initiative.

Supplemental materials

References


22. Tongco MDC. Purposive sampling as a tool for informant selection. Ethnobot Res Appl. 2007;5:147-158. 10.17348/era.5.0.147-158


261