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

Ann M. Nguyen, *PhD, MPH*[1]

Charles M. Cleland, *PhD*[2]

L. Miriam Dickinson, *PhD*[3]

Michael P. Barry[4]

Samuel Cykert, *MD*[5]

F. Daniel Duffy, *MD, MACP*[6]

Anton J. Kuzel, *MD, MHPE*[7]

Stephan R. Lindner, *PhD*[8]

Michael L. Parchman, *MD, MPH*[9]

Donna R. Shelley, *MD, MPH*[10]

Theresa L. Walunas, *PhD*[11]

[1]Rutgers University, Center for State Health Policy, New Brunswick, New Jersey

[2]NYU Langone Health, New York, New York

[3]University of Colorado, Aurora, Colorado

[4]SUNY Downstate Health Sciences University College of Medicine, Brooklyn, New York

[5]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

[6]University of Oklahoma Health Sciences Center, Tulsa, Oklahoma

[7]Virginia Commonwealth University, Richmond, Virginia

[8]Oregon Health & Science University, Portland, Oregon

[9]Kaiser Permanente Washington Health Research Institute, Seattle, Washington

[10]New York University School of Global Public Health, New York, New York

[11]Northwestern University, Feinberg School of Medicine, Chicago, Illinois

CORRESPONDING AUTHOR

Ann M. Nguyen
Center for State Health Policy,
Rutgers University
112 Paterson Street, 5th Floor
New Brunswick, NJ 08901
anguyen@ifh.rutgers.edu

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.

👆 **VISUAL ABSTRACT**

INTRODUCTION

As 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.[1,2] 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.[2]

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.[3] Examining 2 systematic reviews of SW-CRT studies,[1,4] only 14 SW-CRTs have been conducted in primary care settings; those study teams selected the design on the basis of resource constraints,[5] methodologic preferences (eg, phased implementation,
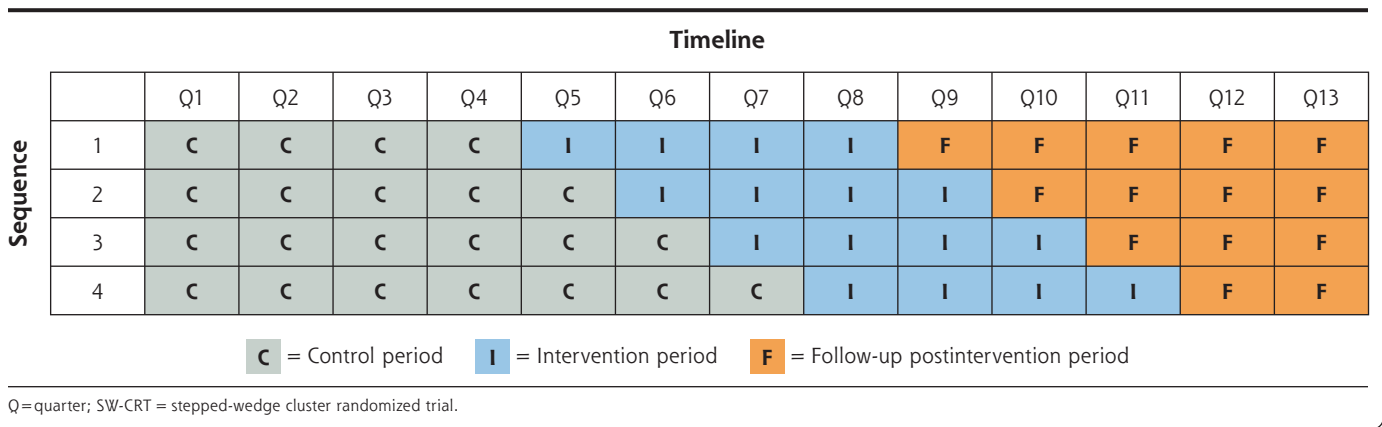
Figure 1. Sample SW-CRT scheme.

		Timeline												
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Sequence	1	C	C	C	C	I	I	I	I	F	F	F	F	F
	2	C	C	C	C	C	I	I	I	I	F	F	F	F
	3	C	C	C	C	C	C	I	I	I	I	F	F	F
	4	C	C	C	C	C	C	C	I	I	I	I	F	F

C = Control period   I = Intervention period   F = Follow-up postintervention period

Q = quarter; SW-CRT = stepped-wedge cluster randomized trial.

all participants receive the intervention),[6-12] or had no reason indicated.[13] The largest of the studies included 72 practices,[13] and only 1 was conducted in the United States.[5]

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.[14-17] We examined the considerations for selecting SW-CRT for large-scale implementations, specifically in primary care settings.

Table 1. Overview of Cooperatives

Cooperative	Region Served	No. of Participating Practices	Study Design
Midwest	Wisconsin, Illinois, Indiana	226	Parallel cluster randomized trial[a]
New York City	New York City	276	SW-CRT
North Carolina	North Carolina	228	SW-CRT
Northwest	Washington, Oregon, Idaho	209	2×2 parallel randomized factorial trial[b]
Oklahoma	Oklahoma	226	SW-CRT
Southwest	Colorado, New Mexico	202	Parallel cluster randomized trial[c]
Virginia	Virginia	208	SW-CRT

SW-CRT = stepped-wedge cluster randomized trial.

[a] This parallel cluster randomized trial had 4 waves with 2 intervention arms in each wave (intervention and intervention plus), which included additional support. Practices were recruited in 4 separate waves, and waves began at staggered times.

[b] The 2×2 randomized factorial trial had 2 factors (factor A and factor B), each with an on and off level potentially added to the standard level, leading to 4 study conditions. Practices were recruited up front, and all arms followed the same timeline.

[c] This parallel cluster randomized trial had 2 intervention arms (1 standard and 1 enhanced), which included additional support. Practices were recruited up front, and both arms followed the same timeline.

METHODS

Study Setting

EvidenceNOW was designed to improve cardiovascular health care delivery in the United States,[18] 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.[19] 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[20]; 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."[21] 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[22]);

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

---

Table 2. Advantages and Challenges of Using SW-CRTs

	Description	Representative Quote
Advantage		
Incentivized recruitment	All sites receive the intervention and serve as their own control.	"To ask a site to engage in research for a year without receiving resources doesn't seem possible." (New York City)
Staggered resource allocation	Resources (eg, personnel) can be allocated over a longer period.	"You can shift resources from one sequence to another, which eases workforce logistical concerns." (ESCALATES)
Statistical power	Power can be higher than parallel cluster randomized controlled trials under certain conditions.	"Stepped-wedge designs potentially have the power advantage over alternative designs, though the trade-off is the time issue." (New York City)
Challenge		
Time-sensitive recruitment	Recruitment must occur at all sites up front.	"Recruitment challenges were the main reason we chose the parallel cluster randomized trial and not a stepped-wedge design." (Midwest)
Retention	Sites might drop out owing to a long lag time between recruitment and the start of the intervention.	"By the time all partnership agreements were signed and sites were randomized, 47 had dropped out of the study." (Northwest)
Randomization requirements and practice preferences	It might be difficult to randomize sites into sequences, given that real-world practice priorities are often changing.	"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)
Achieving treatment schedule fidelity	It might be difficult to deliver the intervention as prescribed (eg, sites might cross-talk across sequences).	"We held weekly meetings with all facilitators to deliberately talk about cross-contamination and staying with the timeline." (Oklahoma)
Intensive data collection	Sites might have difficulty contributing data for specified outcome measures for every time block of the implementation timeline.	"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)
Hawthorne effect	Sites might modify their behavior before the intervention begins.	"We anticipated that the sites would start preparing (before the intervention started)." (Midwest)
Temporal trends	Effect of intervention might be confounded by underlying temporal trends.	"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)

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 recommended 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) incentivized 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 intervention 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 implementations 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 comparison, 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 electronic 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 Hawthorne 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 randomization. The SW-CRT design does not allow for staggered recruitment; staggered recruitment prevents full randomization. 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, particularly 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 to 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 Northwest cooperative reported that 47 sites dropped out by the time all partnership agreements were signed and sites randomized. Others lost sites randomized to later sequences, which involved waiting more than a year before starting the intervention. 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 randomization 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' preferences 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 Virginia cooperative, which used SW-CRT design, opted to randomize groups of practices as a block to contain any crosstalk within sequences. There was also the risk that facilitators working across multiple sequences were delivering the intervention 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 process[28]; 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 care[1,4,15] 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.[31] 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 have included mixed-effects regression models,[32-34] generalized estimating equations,[35] and robust nonparametric methods.[36-38] 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.**

☜ **Supplemental materials**

References

1. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol*. 2011;64(9):936-948. 10.1016/j.jclinepi.2010.12.003

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. 10.1136/bmj.h391

3. Hemming K, Taljaard M, Grimshaw J. Introducing the new CONSORT extension for stepped-wedge cluster randomised trials. *Trials*. 2019;20(1):68. 10.1186/s13063-018-3116-3

4. Beard E, Lewis JJ, Copas A, et al. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials*. 2015;16:353. 10.1186/s13063-015-0839-2

5. Weiner M, El Hoyek G, Wang L, et al. A web-based generalist-specialist system to improve scheduling of outpatient specialty consultations in an academic center. *J Gen Intern Med*. 2009;24(6):710-715. 10.1007/s11606-009-0971-3

6. Liddy C, Hogg W, Singh J, et al. A real-world stepped wedge cluster randomized trial of practice facilitation to improve cardiovascular care. *Implement Sci*. 2015;10:150. 10.1186/s13012-015-0341-y

7. Dreischulte T, Grant A, Donnan P, et al. A cluster randomised stepped wedge trial to evaluate the effectiveness of a multifaceted information technology-based intervention in reducing high-risk prescribing of non-steroidal anti-inflammatory drugs and antiplatelets in primary medical care: the DQIP study protocol. *Implement Sci*. 2012;7:24. 10.1186/1748-5908-7-24

8. Gucciardi E, Fortugno M, Horodezny S, et al. Will Mobile Diabetes Education Teams (MDETs) in primary care improve patient care processes and health outcomes? Study protocol for a randomized controlled trial. *Trials*. 2012;13:165. 10.1186/1745-6215-13-165

9. Keriel-Gascou M, Buchet-Poyau K, Duclos A, et al. Evaluation of an interactive program for preventing adverse drug events in primary care: study protocol of the InPAct cluster randomised stepped wedge trial. *Implement Sci*. 2013;8:69. 10.1186/1748-5908-8-69

10. Marshall T, Caley M, Hemming K, Gill P, Gale N, Jolly K. Mixed methods evaluation of targeted case finding for cardiovascular disease prevention using a stepped wedged cluster RCT. *BMC Public Health*. 2012;12:908. 10.1186/1471-2458-12-908

11. Praveen D, Patel A, McMahon S, et al. A multifaceted strategy using mobile technology to assist rural primary healthcare doctors and frontline health workers in cardiovascular disease risk management: protocol for the SMART-Health India cluster randomised controlled trial. *Implement Sci*. 2013;8:137. 10.1186/1748-5908-8-137

12. Stringer JS, Chisembele-Taylor A, Chibwesha CJ, et al. Protocol-driven primary care and community linkages to improve population health in rural Zambia: the Better Health Outcomes through Mentoring and Assessment (BHOMA) project. *BMC Health Serv Res*. 2013;13(Suppl 2):S7. 10.1186/1472-6963-13-S2-S7

13. Fearon P, Quinn T, Wright F, Fraser P, McAlpine CSD. Evaluation of a telephone hotline to enhance rapid outpatient assessment of minor stroke or TIA: a stepped wedge cluster randomised control trial. *Int J Stroke*. 2013;8(3 Suppl):6-7. 10.1111/ijs.12213

14. 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(3):1043-1052. 10.1093/ije/dyaa077

15. Prost A, Binik A, Abubakar I, et al. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. *Trials*. 2015;16:351. 10.1186/s13063-015-0837-4

16. Joag K, Ambrosio G, Kestler E, Weijer C, Hemming K, Van der Graaf R. Ethical issues in the design and conduct of stepped-wedge cluster randomized trials in low-resource settings. *Trials*. 2019;20(Suppl 2):703. 10.1186/s13063-019-3842-1

17. Hargreaves JR, Copas AJ, Beard E, et al. Five questions to consider before conducting a stepped wedge trial. *Trials*. 2015;16:350. 10.1186/s13063-015-0841-8

18. Agency for Healthcare Research and Quality. EvidenceNOW: advancing heart health in primary care. Accessed Apr 8, 2022. https://www.ahrq.gov/evidencenow/projects/heart-health/index.html

19. Agency for Healthcare Research and Quality. EvidenceNOW: cooperatives. Published Sep 2018. Accessed Jul 6, 2020. https://www.ahrq.gov/evidence-now/about/cooperatives/index.html

20. Meyers D, Miller T, Genevro J, et al. EvidenceNOW: balancing primary care implementation and implementation research. *Ann Fam Med*. 2018;16(Suppl 1):S5-S11. 10.1370/afm.2196

21. Beebe J. *Rapid Assessment Process: An Introduction*. AltaMira Press; 2001.

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

23. Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychol Health*. 2010;25(10):1229-1245. 10.1080/08870440903194015

24. Cuthel A, Rogers E, Daniel F, Carroll E, Pham-Singer H, Shelley D. Barriers and facilitators in the recruitment and retention of more than 250 small independent primary care practices for EvidenceNOW. *Am J Med Qual*. 2020; 35(5):388-396. 10.1177/1062860619893422

25. Sweeney SM, Hall JD, Ono SS, et al. Recruiting practices for change initiatives is hard: findings from EvidenceNOW. *Am J Med Qual*. 2018;33(3):246-252. 10.1177/1062860617728791

26. Fagnan LJ, Walunas TL, Parchman ML, et al. Engaging primary care practices in studies of improvement: did you budget enough for practice recruitment? *Ann Fam Med*. 2018;16(Suppl 1):S72-S79. 10.1370/afm.2199

27. Fernald DH, Jortberg BT, Hessler DM, et al. Recruiting primary care practices for research: reflections and reminders. *J Am Board Fam Med*. 2018; 31(6):947-951. 10.3122/jabfm.2018.06.180025

28. Cohen DJ, Dorr DA, Knierim K, et al. Primary care practices' abilities and challenges in using electronic health record data for quality improvement. *Health Aff (Millwood)*. 2018;37(4):635-643. 10.1377/hlthaff.2017.1254

29. Lenguerrand E, Winter C, Siassakos D, 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(2): 122-134. 10.1136/bmjqs-2018-008625

30. Hastings SN, Stechuchak KM, Choate A, et al. Implementation of a stepped wedge cluster randomized trial to evaluate a hospital mobility program. *Trials*. 2020;21(1):863. 10.1186/s13063-020-04764-7

31. Handley MA, Schillinger D, Shiboski S. Quasi-experimental designs in practice-based research settings: design and implementation considerations. *J Am Board Fam Med*. 2011;24(5):589-596. 10.3122/jabfm.2011.05.110067

32. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007;28(2):182-191. 10.1016/j.cct.2006.05.007

33. Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: an overview. *Stat Methods Med Res*. 2021;30(2):612-639. 10.1177/0962280220932962

34. Thompson JA, Fielding KL, Davey C, Aiken AM, Hargreaves JR, Hayes RJ. Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis. *Stat Med*. 2017;36(23):3670-3682. 10.1002/sim.7348

35. Scott JM, deCamp A, Juraska M, Fay MP, Gilbert PB. Finite-sample corrected generalized estimating equation of population average treatment effects in stepped wedge cluster randomized trials. *Stat Methods Med Res*. 2017;26(2): 583-597. 10.1177/0962280214552092

36. Hughes JP, Heagerty PJ, Xia F, Ren Y. Robust inference for the stepped wedge design. *Biometrics*. 2020;76(1):119-130. 10.1111/biom.13106

37. Thompson JA, Davey C, Fielding K, Hargreaves JR, Hayes RJ. Robust analysis of stepped wedge trials using cluster-level summaries within periods. *Stat Med*. 2018;37(16):2487-2500. 10.1002/sim.7668

38. Wang R, De Gruttola V. The use of permutation tests for the analysis of parallel and stepped-wedge cluster-randomized trials. *Stat Med*. 2017;36(18): 2831-2843. 10.1002/sim.7329