Journal of Clinical Epidemiology 107 (2019) 89–100

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size

Felizitas A. Eichner[a,*,1], Rolf H.H. Groenwold[a,b], Diederick E. Grobbee[a], Katrien Oude Rengerink[a]

[a]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands*
[b]*Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands*

Accepted 14 November 2018; Published online 17 November 2018

Abstract

Objective: To determine how often stepped-wedge cluster randomized controlled trials reach their planned sample size, and what reasons are reported for choosing a stepped-wedge trial design.

Study Design and Setting: We conducted a PubMed literature search (period 2012 to 2017) and included articles describing the results of a stepped-wedge cluster randomized trial. We calculated the percentage of studies reaching their prespecified number of participants and clusters, and we summarized the reasons for choosing the stepped-wedge trial design as well as difficulties during enrollment.

Results: Forty-six individual stepped-wedge studies from a total of 53 articles were included in our review. Of the 35 studies, for which recruitment rate could be calculated, 69% recruited their planned number of participants, with 80% having recruited the planned number of clusters. Ethical reasons were the most common motivation for choosing the stepped-wedge trial design. Most important difficulties during study conduct were dropout of clusters and delayed implementation of the intervention.

Conclusion: About half of recently published stepped-wedge trials reached their planned sample size indicating that recruitment is also a major problem in these trials. Still, the stepped-wedge trial design can yield practical, ethical, and methodological advantages.

Keywords: Stepped-wedge cluster randomized trial; Recruitment; Sample size; Systematic review

1. Introduction

The stepped-wedge cluster randomized trial design is a subtype of the parallel cluster design and has first been introduced about 1987 [1]. In contrast to the parallel cluster randomized trial design, all clusters in a stepped-wedge trial will receive the new intervention, yet the moment at which they do so is determined by chance [2,3].

Over the years, the stepped-wedge cluster randomized trial design has become popular in various research settings but mainly for investigating health care interventions [3,4]. Compared to the parallel cluster randomized trial, this design is especially attractive when it appears to be impractical that all clusters receive the intervention simultaneously—because of logistical issues or because contamination between both arms may occur—or when it is unethical to withhold the intervention for some patients as it has been suggested to be potentially advantageous in previous studies [2,3,5].

However, it is a well known and common challenge for all randomized trials to recruit the required number of participants within the set time [6–8]. The reasons for this problem are diverse and include overestimation of the available number of eligible participants, delayed start of recruitment, and inadequate planning [6]. Not only for the study as a whole but also across centers, recruitment may vary [9]. Furthermore, many physicians underestimate the time and effort needed for recruitment of their patients into a trial. Moreover, physicians or patients may refuse to participate in clinical trials, for instance because of concerns about the randomization and possible side effects [10,11]. Difficulties in recruiting sufficient participants

Conflict of interest statement: None.

[1] Present address: Institute for Clinical Epidemiology and Biometry, University Würzburg, Germany.

* Corresponding author. Institute for Clinical Epidemiology and Biometry University of Würzburg, Josef-Schneider-Straße 2, 97078 Würzburg, Germany. Tel.: + 49 (0) 931 201 47311; fax: +49 (0) 931 201 647310.

E-mail address: felizitas.eichner@uni-wuerzburg.de (F.A. Eichner).

https://doi.org/10.1016/j.jclinepi.2018.11.013
0895-4356/

What is new?

- First systematic review to assess recruitment rates into stepped-wedge cluster randomized trials.
- 80% of the included studies recruited the planned number of clusters, yet only 69% recruited their planned number of participants.
- Ethical reasons were the most common reasons to use a stepped-wedge design.
- Researchers are advised to be conservative when estimating the available number of patients per cluster.
- To increase transparency, improved sample size reporting in stepped-wedge cluster trials is still needed.

can threaten the power and hence the value of the study, whereas those who were included in the study were still exposed to the burdens and potential risks of trial participation, which is deemed unethical [12].

The stepped-wedge design has the potential advantage that clusters are not randomized to *if* they receive the intervention but to *when* they receive the intervention. This could lead to a higher willingness of clusters to participate in the trial and thus improve recruitment rates [5,13]. In addition, especially in cohort stepped-wedge cluster trials, thus stepped-wedge trials in which all participants are exposed to both the control and the intervention, patients might also be more willing to participate [14].

It is unknown whether the potential advantages in recruitment of stepped-wedge trials are actually utilized and whether such studies successfully recruit their required sample size. Although several reviews exist that assess recruitment rates in parallel randomized trials, no such review exists for the stepped-wedge cluster design. We therefore performed a systematic review of stepped-wedge cluster randomized trials to assess to what extent studies using this design reached their planned sample size.

2. Materials and methods

2.1. Literature search

We searched PubMed on February 14, 2017 to identify articles published between January 1, 2012 and January 1, 2017. We restricted our search to articles published in these 5 years as we wanted to include recently performed trials reflecting recruitment rates that are likely to be comparable to trials performed nowadays. No language restrictions were applied. We also screened the references of the identified studies and previously published reviews to identify additional articles we might have missed in our search.

Articles were included in our review if they had a sequential rollout of the intervention to all clusters, used cluster randomization, were conducted in human subjects, and were original studies. We excluded articles that were called stepped-wedge trials, but randomized patients on an individual level.

Our search was based on common descriptions for stepped-wedge trial designs and randomization (Table 1).

2.2. Selection of articles

Title and abstract of all articles identified by the search were screened for eligibility by one reviewer, Felizitas A. Eichner (FAE); any uncertainty was resolved by discussion with three reviewers FAE, Katrien Oude Rengerink (KOR) and Rolf H.H. Groenwold (RHHG). Of potentially eligible articles, the full-text was retrieved and eligibility was assessed by one reviewer (FAE), and any uncertainty or disagreement was resolved by discussion with three reviewers (FAE, KOR, RHHG).

2.3. Data extraction

From the eligible articles, we extracted information on the type of intervention, the country where the trial was performed, the planned and final sample size and the arguments for choosing this sample size, the planned and final number of clusters, the planned and final number of steps, the planned and final duration of the study, difficulties during intervention rollout, and the reasons for choosing the stepped-wedge trial design. If information was not reported in the main article, we searched whether a study protocol was available which obtained the missing information.

The information was extracted by one researcher (FAE); in case of uncertainty this was discussed with two other researchers (KOR, RHHG).

Table 1. Search strategy used to identify articles about recruitment into stepped-wedge cluster randomized controlled trials

Search	Query	Items
#1	random*[tiab]	888,882
#2	(stepped [tiab] AND wedge [tiab])	349
#3	(stepped [tiab] AND cluster [tiab])	286
#4	(step [tiab] AND wedge [tiab])	560
#5	(phased [tiab] AND implementation [tiab])	388
#6	#2 OR #3 OR #4 OR #5	1338
#7	#1 AND #6	394
#8	NOT review [pt] NOT review [tiab]	343
#9	Publication date from 2012/01/01 to 2017/01/01	277

2.4. Analysis and synthesis of the results

We defined a sample size being reached if at least 100% of the planned sample size was included in the study. To determine whether a study reached their predefined sample size, we calculated the ratio between the planned and final sample size. In addition, we categorized the reasons for choosing the stepped-wedge trial design into ethical reasons, practical reasons, methodological reasons, and other reasons. Ethical reasons applied when equipoise was considered not to be sufficiently present or when eventual withholding the intervention was deemed unethical (all clusters desired to implement the intervention). Reasons were categorized as practical when it was impossible or difficult to introduce the intervention in multiple centers simultaneously because of logistical or monetary constraints, when the number of available staff members to implement the intervention was limited, and when the stepped implementation facilitated the rollout for other reasons. Reasons categorized as methodological included, among others, adjustment for time effects, avoidance of contamination bias, the advantage of clusters functioning as their own controls, higher power, and fewer clusters needed compared to the parallel trial design. Reasons that did not fit the classification ethical, practical, or methodological were classified as other reasons.

3. Results

3.1. Included studies

The PubMed search and reference screening revealed 280 articles. After the selection process, we included 53 articles in this review describing 46 studies [15–45] [46–67] (Table 2, Appendix A. Figure 1).

Of the 46 studies, 18 were conducted in Europe [21,23,28,31,34,35,38,39,41,44,51,54,57,61–63,65,67], 11 in America [25,26,29,32,33,42,47,50,53,55,58], seven in Africa [15,19,22,27,37,48,49], five in Australia or New Zealand [17,36,45,56,60], two in Asia [46,64], and two in the Middle East [16,40] (Figure 1). Lashoher et al. conducted the only multicountry trial, in which the intervention was implemented in nine different countries [43].

The trials included two to 190 clusters, with two to 15 steps (Table 2). The units of randomization (clusters) were diverse, ranging from health facilities such as hospitals, intensive care units, and general practices ($N = 27$) [15–17,22,23,25,28,33–35,38,40,42,43,47–50,53–56,58,60,61,63,65], geographical units of the general population such as villages and communities ($N = 7$) [19,27,29,32,37,57,62], public institutions such as universities, children day care centers or prisons ($N = 4$) [21,24,45,46,64], care homes for the elderly ($N = 4$) [39,41,44,67] as well as primary care laboratories ($N = 1$) [26], rehabilitation units ($N = 1$) [36], palliative home care sites ($N = 1$) [31] and one study in which clusters were defined as workers in elderly care homes in the same district (Table 2) [51]. In line with this result, also the type of study participants was diverse (Appendix B. Table 1). The type of data collection was cross-sectional in 20 studies, whereas seven used an open and 16 a closed cohort design (Appendix B. Table 1). For four studies, the type of data collection was unclear.

The most often implemented intervention was a health care intervention ($N = 12$), implemented either in health facilities or in nursing homes (Appendix C. Table 2).

3.2. Reasons for choosing the stepped-wedge trial design

Sixteen studies stated at least one reason for choosing the stepped-wedge cluster randomized design and 23 stated two or more reasons, wheres seven studies did not report their reason. Twenty-two articles stated an ethical reason, 18 a practical one, 16 a methodological one, and 11 a reason categorized as "other" (Appendix C. Table 2).

For all drug interventions, an ethical reason was stated, whereas practical reasons were more often found among the

Figure 1. Countries of study conduct of 46 stepped-wedge cluster randomized trials included in a literature review. All shaded countries were represented in this review in at least one study. The following regions were defined: Europe, America, Africa, Australasia, Asia, and Middle East.

screening and prevention interventions. No other relevant differences between the types of interventions were observed.

3.3. Recruitment achieved in stepped-wedge cluster randomized trials

Information about recruitment rates was reported for 48 of the 53 studies included in this review (five pairs of publications reported on the same study, yet answered different research questions) (Table 3, Appendix B. Table 1).

There was a marked variability in the number of participants included in the studies ranging from 16 to 292,000 (median of 1,720 participants) [48,64].

The prespecified sample size was most often based on formal sample size calculations ($N = 21$), in which the parameters for the calculations were obtained from pilot studies, simulation studies, or studies conducted by other researchers [16–19,23,25,26,34,37,38,41,43,51,53,56–58,60,63,65,67] (Table 3). Eleven studies based their prespecified sample size on what was considered feasible given the study setting [15,28,29,32,35,40,42,47,48,50,55] (Table 3). Seven studies stated a combination of statistical calculations and restricted study settings [27,33,36,44,45,49,62]. In eight studies, the basis for the sample calculation was not reported [21,22,31,39,46,59,61,64]. In 16 of the 48 studies, the planned sample size was specified as estimated participants per cluster or per data collection time point without stating an overall sample size [16,18,19,21,29,33,38,41,47,49,53,54,56,61,63,67]. In four studies, the planned sample size could be calculated based on other parameters [25,57,58,62] (Appendix B. Table 1).

For 36 studies, we could calculate whether the planned sample size was achieved, from 9 articles the planned ($N = 7$) [22,26,39,43,46,59,64] or final sample size ($N = 2$) [26,50] could not be obtained or was not applicable (Table 3). In two studies, recruitment was performed on cluster level only [23,28]; in one study, the units of measurement were unequal between the planned (participants) and final (total tests) sample size [21]; in another study, the planned sample size was only stated as a range [32].

In eight of the 36 studies, the sample size was not based on the number of participants but on the number of outcomes collected (such as survey responses or number of tests performed) [18,27,29,31,50,57,58,61]. For these studies, if the planned and final sample size referred to the same unit of measurement, sample size calculations for the review were based on the number of outcomes.

Thirty six studies recruited between 48% and 902% of the planned number of patients [35,51]. The planned number of patients was included in 69% [24/35] of the studies. Twenty-three studies recruited more participants than previously planned, with seven <50% more [17,18,31,33,44,45,49] and nine >50% more than previously planned [29,34,37,51,54,55,57,65,67]. For seven studies, the exact percentage of over-recruitment could not be calculated, as the planned sample size was rather an estimation than an exactly calculated number. All these studies however recruited more participants than the estimated sample size [15,19,25,42,48,53,58]. Of the 12 studies that did not reach the prespecified sample size, only one study recruited less than 50% of the planned participants [35]; others recruited between 50 and 99.6% of the patients [27,36,38,40,41,47,56,60–63]. The median recruitment rate in these 12 studies was 92%.

Recruitment rates on the cluster level could be calculated for 41 studies only. For seven studies (17%), the planned number of clusters was missing or unclear (Table 3).

Of the 41 studies, 31 met (76%) and five exceeded (12%) their planned number of clusters. The studies did not report a reason for recruiting more clusters than originally planned. Five studies (10%) recruited less clusters than planned, but still recruited at least 80% of the required sample size [23,29,43,45,60].

3.4. Study duration

The studies lasted 9 days to 4 years (median 20 months). For one study, duration was not reported [31], and for six studies, planned study duration was not reported [16,22,31,43,46,64].

Thirty-five studies (76%) completed their study within the planned study period (Table 2). Five studies reported a prolongation of the implementation or follow-up period with 3 months ($N = 1$) [28], 6 months ($N = 2$) [25,62], and 1 year ($N = 1$) [21]. In one study, the exact extension could not be calculated [55].

Although only a small percentage of the studies extended their planned study duration, difficulties during the study appeared more frequently. Twenty studies (43%) reported one or multiple difficulties with the study rollout, data collection, or data analysis (Appendix C. Table 2).

Six trials reported the dropout of clusters after randomization or after data collection had started [23,47,53,54,60,61]. In three of these studies, it was decided to exclude the data from these clusters from the final analysis [23,47,60]. Another full cluster was excluded from an analysis as the intervention was implemented before the randomization date [53]. Furthermore, the interventions were delayed in four studies [21,28,42,43]. Also, several clusters or cluster members did not receive the intervention at all in five studies [22,32,37,43,48]. In one study that implemented a safety checklist for surgeries, the checklist was only applied in about 75% of the cases [34]. Finally, one study reported a shortening of the lead-in period by 9 weeks because of a strike [22] and one study had to send more surveys than planned to reach a sufficient response rate [57].

The reasons leading to these difficulties were not always stated but included fluctuations in staffing, retirement of staff members, delayed or non-approval by the ethical

Table 2. Characteristics of 46 studies included in the review

First author	Year	Country	Setting	Study duration [months]			Final number of steps
				Planned	Final	% of planned	
Amanyire	2016	Uganda	Health facility	28	28	100%	4
Bashour	2013	Syria	Maternity hospital	NR	10	CBD	4
Bennett	2016	Australia	Dialysis clinic	11	11	100%	3
Chinbuah and Chinbuah[a]	2012/2013	Ghana	Community	48	48	100%	2
Cissé and NDiaye[a]	2016	Senegal	Health post	36	36	100%	3
Craine	2015	United Kingdom	Prison	~6	18	~300%	5
Dryden-Peterson	2015	Botswana	Antenatal clinic	NR	10	CBD	10
Due	2014	Denmark	General practice	16	16	100%	2
Duijster and Monse[a]	2013/2012	Philippines	Day care center	NR	~8	CBD	2
Durovni and Golub[a]	2013/2015	Brazil	HIV clinics	42	48	114%	14
Durovni and Trajman[a]	2014/2015	Brazil	Primary care laboratories with serviced clinics	8	8	100%	7
Fink	2013	Burkina Faso	Community	72	72	100%	3
Fuller	2012	United Kingdom	Hospital	36	39	108%	5
Golden	2015	United States	Local health jurisdiction	22	22	100%	4
Grande[b]	2015	United Kingdom	Palliative home care site	NR	NR	CBD	4[c]
Gruber	2013	Mexico	Community	17	17	100%	6
Happ	2015	United States	Intensive care unit	23	23	100%	6
Haugen	2015	Norway	Surgery specialty in hospital	11	11	100%	5
Hemming	2016	United Kingdom	General practice	30	30	100%	15
Hill	2015	Australia	Rehabilitation unit	12	12	100%	4
Homan	2016	Kenya	Household	35	35	100%	9
Hoogendijk	2015	Netherlands	Primary care practice	24	24	100%	4
Horner	2012	United Kingdom	Care home	28	28	100%	3
Jeddian	2016	Iran	General hospital ward	17	17	100%	6
Jordan	2015	United Kingdom	Care home	6	6	100%	5
Karkouti	2016	Canada	Hospital	7	7	100%	6
Lashoher	2016	Nine low-, middle- and high-income countries	Hospital	NR	24	CBD	NR
Leontjevas	2013	Netherlands	Nursing home	23	23	100%	5
Mhurchu	2013	New Zealand	School	12	12	100%	4
Morrison	2015	Canada	Hospital	36	36	100%	4
Ononge	2015	Uganda	Health facility	12	12	100%	3
Pickering	2015	United States	Intensive care unit	2	2	100%	~4
Rasmussen and Rasmussen[a]	2015/2016	Denmark	Clusters not based on institutions, but workers in elderly care in the same district	30	30	100%	4
Rodriguez	2015	Argentina	Intensive care unit	~9	9	~100%	4
Roy	2013	United Kingdom	Clinic	7	7	100%	6
Scales	2016	Canada	Hospital	~36	37	>100%	4
Schultz	2014	Australia	Hospital ward	17	17	100%	3[d]

(*Continued*)

Table 2. Continued

First author	Year	Country	Setting	Study duration [months]			Final number of steps
				Planned	Final	% of planned	
Solomon	2014	United Kingdom	Rural village	20	20	100%	4
Stern	2014	Canada	Long-term care facility	17	17	100%	11
Turner	2016	Australia	Cancer treatment center	24	24	100%	4[d]
van de Steeg	2014	Netherlands	Hospital	11	11	100%	10[d]
van den Broek	2012	Netherlands	Geographical and administrative units	26	32	123%	3
van Leeuwen	2015	Netherlands	Primary care practice	24	24	100%	4
Wu	2015	Taiwan	University building	NR	0.3[e]	CBD	2
Zhan	2014	Netherlands	Hospital	24	24	100%	5
Zwijsen and Zwijsen[a]	2014/2015	Netherlands	Dementia special care unit	20	20	100%	5

Abbreviations: NR, not reported; CBD, cannot be determined.
[a] Same studies but different research questions.
[b] Because of pragmatic constraints, the sequence of sites beginning the intervention was not fully randomized.
[c] Article states five steps, but one center received intervention not through randomization but because it was estimated to be the most ready center.
[d] Only articles which performed one step less than originally planned.
[e] Experimental period only, full study duration not stated.

committee, operational difficulties, and other competing interventions.

4. Discussion

Our systematic literature review of recruitment rates in stepped-wedge cluster randomized trials showed that about 69% of the 46 included studies reached or even exceeded their prespecified sample size, with 80% reaching the prespecified cluster size. The reason for choosing the stepped-wedge cluster randomized trial design was most often an ethical one ($N = 22$). Only one study reported explicitly that this trial design was applied to facilitate the recruitment of participants [65]. However, description of the planned and final sample size was lacking or insufficient in many articles, which shows that improved reporting is needed.

In this review, the recruitment failure rate in stepped-wedge cluster randomized trials was estimated to be 31%. Yet, we based our calculation of planned sample size reached only on the 35 studies in which both planned and final sample size were available. If including the missing and using all 48 studies as the denominator, achieved recruitment rate would still be at least 50%. We did not identify other articles that have addressed recruitment rates in stepped-wedge cluster trials. When compared to parallel cluster studies, Eldridge et al. (2004) reported low recruitment in 1% of 152 published cluster trials; however, a definition of low recruitment was lacking and this aspect was not part of the main research question [68]. Another study that explored barriers and facilitators for recruitment into cluster trials reported that only 30% of the participating general practitioners reached the planned sample size [69]. Furthermore, it has previously been suggested that recruitment into stepped-wedge cluster trials might be more successful compared to traditional randomized trials [2]. Yet different reviews reported that the planned recruitment goal was achieved in between 30 and 48% of these studies [7,70], which is comparable to the findings of our review.

Our results could not show recruitment benefit of stepped-wedge cluster randomized trials over parallel cluster trials or randomized controlled trials. However, it is difficult to compare stepped-wedge cluster and randomized trials directly because there are often large differences in interventions and research settings. The parallel group randomized controlled trial design is most often used for efficacy trials, whereas stepped-wedge cluster trials normally take place in a real-world setting, which better reflects effectiveness trials [71,72]. It could be that if the reported studies had been designed as parallel group trials, recruitment would have been lower.

On the other hand, a more extensive indirect comparison of recruitment rates between parallel cluster randomized trials and stepped-wedge cluster trials can only be performed when average recruitment rates of recently published parallel cluster randomized trials are known. However, although several reviews about parallel cluster trials exist, none of them addressed planned and final recruitment rates [68,73,74].

Regarding our findings about recruitment rates, a few other aspects need to be discussed in more detail. Because more studies (~80%) reached their prespecified number of clusters than their number of participants in our review, one must assume that the researchers were overall too

Table 3. Sample size considerations for 48 articles included in a literature review about stepped-wedge cluster randomized trials

	Sample size			Clusters			Rationale of sample size
First author	Planned	Achieved	% of planned	Planned	Final	% of planned	
Amanyire	~12,000	12,024	>100%[a]	20	20	100%	Practical limitations
Bashour	2,000	2,000	100%	4	4	100%	Statistical
Bennett	180	228	127%	15	15	100%	Statistical
Chinbuah 2012[b]	(1) 10,000[c] (2) 6,000	12,333	>100%[a]	60 or 100	114	>100%	Statistical
Chinbuah 2013[b]	5,600 survey responses	7,794 survey responses	139%	NR	114	CBD	Statistical
Cissé and Ndiaye[d]	~75,600 children	~292,000 children treated	>100%[a]	NR	54	CBD	Practical limitations
Craine	750 participants	2,237 total tests	CBD	5	5	100%	Unclear
Dryden-Peterson	NR	366	CBD	NR	20	CBD	NR
Due	NA (assessment on cluster level only)	NA	CBD	189	183	96.8%	Statistical
Duijster and Monse[d]	NR	202	CBD	13	13	100%	NR
Durovni and Golub[d]	~8,700	12,816	>100%[a]	29	29	100%	Statistical
Durovni 2014[b]	unclear	Probably smaller than planned	CBD	14	14	100%	Statistical
Trajman 2015[b]	NR	4,088 enrolled	CBD	14	14	100%	NR
Fink	990 households	983 households	99.3%	33	33	100%	Statistical and practical
Fuller	NA	NA	CBD	16	16	100%	Probably practical limitations
Golden	19,440 tests	91,971 tests	473,1%	24	23	95.8%	Practical limitations
Grande	540 survey responses	681 survey responses	126.1%	6	6	100%	Unclear
Gruber	400-500 households *Note: in the sample size calculation, they assume 4.5 individuals per household*	444 households *Note: based on 444 households, they state a max. of 3,108 observations, which is equal to 1 individual per household*	CBD	NR	24	CBD	Practical limitations
Happ	(1) 300 nurses (2) 1,440 patients	(1) 323 nurses (2) 1,440 patients	(1) 107.7% (2) 100%	~6	6	~100%	Statistical and practical
Haugen	2,220	3,811	171.7%	5	5	100%	Statistical
Hemming	6,240	2,969	47.6%	26	26	100%	Practical limitations
Hill	(1) 2,409 (2) 3,744	(2) 3,606	96.3%	8	8	100%	(1) Statistical (2) Practical limitations
Homan	7,914 persons	34,041 participants	430.1%	81	81	100%	Statistical

(*Continued*)

Table 3. Continued

First author	Sample size			Clusters			Rationale of sample size
	Planned	Achieved	% of planned	Planned	Final	% of planned	
Hoogendijk	1,244	1,147	92.2%	35	35	100%	Statistical
Horner	NR	2,492	CBD	NR	68	CBD	NR
Jeddian	23,000 admissions	22,919 admissions	99.6%	13	13	100%	Practical limitations
Jordan	50	43	86.0%	5	5	100%	Statistical
Karkouti	~7,000	7,402	>100%[a]	12	12	100%	Practical limitations
Lashoher	unclear	3,422 patients enrolled	CBD	13	11	84.6%	Statistical
Leontjevas	720	793	110.1%	32 (16 somatic, 16 dementia units)	33 (17 somatic, 16 dementia units)	103.1%	Statistical and practical
Mhurchu	400	424	106.0%	16	14	87.5%	Statistical and practical
Morrison	1,020	934	91.6%	32	32	100%	Practical limitations
Ononge	2,400	2,466	102.8%	6	6	100%	Statistical and practical
Pickering	400 survey responses[e]	unclear	CBD	4	4	100%	Practical limitations
Rasmussen and Rasmussen[d]	65 participants for analysis	586 participants included in analysis	901.5%	21	21	100%	Statistical
Rodriguez	300-330	705	>100%[a]	10-11	10	~100%	Statistical
Roy	800	1,315	164.4%	8	24	300.0%	Statistical
Scales	586 in 2 years	905 in 3 years (whole study duration)	154.4%	18	18	100%	Practical limitations
Schultz	400	375	93.8%	25	25	100%	Statistical
Solomon	6,400 survey responses	10,412 survey responses	162.7%	128	128	100%	Statistical
Stern	~126 ulcers	259 ulcers (137 residents)	>100%[a]	10	12	120.0%	Statistical
Turner	600	469	78.2%	5	4	80.0%	Statistical
van de Steeg	3,960 patient records	3,273 patient records	82.7%	18	18	100%	Unclear
van den Broek	95,191	79,173	83.2%	190	190	100%	Statistical and practical
van Leeuwen	1,155	1,147	99.3%	35	35	100%	Statistical
Wu	NR	16	CBD	2	2	100%	NR
Zhan	1,600[f]	3,223	201.4%	11	11	100%	Statistical
Zwijsen and Zwijsen[d]	(1) 280 residents (2) 1,680 measurements	(1) 659 residents (2) 2,292 measurements	(1) 235.4% (2) 136.4%	at least 14	17	121.4%	Statistical

Abbreviations: NR, not reported; CBD, cannot be determined; NA, not applicable.

[a] Planned sample size was exceeded, but exact percentage could not be calculated.

[b] Chinbuah/Chinbuah and Durovni/Trajman are the same studies, but they report different outcomes, hence each specified an individual sample size to answer their research question.

[c] Two sample sizes were calculated based on two different effect sizes.

[d] Same studies but different research questions.

[e] Sample size calculation states 400 provided responses, but outcome is divided in three survey rounds.

[f] Sample size was recalculated retrospectively taking into account the design effect; this revealed an actual sample size of 10,285 participants needed.

optimistic about the recruitment rates of patients per cluster. Some studies like Roy et al. (2013) have tried to compensate this by increasing the number of clusters [54]. However, recruiting more clusters than planned is not always feasible. Moreover especially in stepped-wedge cluster randomized trials, it is methodologically difficult to do so, because additional clusters could also increase the number of required steps and hence prolong the planned study duration substantially [5].

Another possibility for unsuccessful recruitment of clusters could be that not all planned steps of a trial were executed. However, unperformed steps rarely occurred in the studies of this review and were hence no cause for unsuccessful recruitment of clusters.

Yet, we noticed that some articles also counted the prerollout phase as a step, which is methodologically not correct. A good summary of the terminology of stepped-wedge trials is given by Copas et al. [14]. To facilitate the comparison of different articles in the field of stepped-wedge cluster randomized trials, usage of consistent terms is advised.

Besides, many studies of our review did not clearly report their sample size estimation. This shows that there is still a need for improved reporting of sample size methodology in stepped-wedge cluster randomized trials. Also, it is possible that sample size reporting was especially transparent in the studies where recruitment of participants was successful. This has also been reported for other study designs [75]. On the other hand, one should also mention that we specifically looked into planned and achieved recruitment rates of stepped-wedge cluster trials without incorporating the actual results of the studies. We therefore suggest as future work to investigate what proportion of studies that did not reach their sample size did eventually yield clear (significant) results. Vice versa, one must also note that even studies that achieve their recruitment goal may present a quite imprecise effect estimation because of unrealistic assumptions in the sample size calculation.

Regarding the calculation of recruitment rates, it should be mentioned that our definition of reaching the prespecified sample size was quite conservative and had to be at least 100%.

When looking at the reasons for choosing the stepped-wedge design, most studies were conducted in developed countries and were aimed at health care improvements. Beard et al. (2015) have reported a similar finding [4]. We identified ethical reasons as the most common reasons to use a stepped-wedge design, whereas Beard et al. (2015) reported most often logistical reasons. Both of our findings are likely to reflect that stepped-wedge cluster randomized studies are increasingly applied when testing a health care or quality improvement intervention in an entire facility, which is to some extent proven to be beneficial. This was also found by Mdege et al. (2011) who concluded that the design is "mainly used for evaluating interventions during routine implementation" [76].

Besides, only one study chose the stepped-wedge design to facilitate recruitment of participants. However, it is possible that the expectation of facilitated recruitment was a minor reason in some studies and therefore not specifically stated in the article.

Last, many studies included in this review reported difficulties during the intervention's rollout or data collection. This might firstly be explained by the fact that stepped-wedge cluster trials are often conducted on a large scale and are therefore sensitive to external challenges that cannot directly be influenced by the researchers such as change of hospital staff or withdrawal of full clusters. Similar findings have been reported by Kotz et al. (2012) [77]. On the other hand, the stepped-wedge cluster design only exists since the late 1980s [78]. Since then, about 60 stepped-wedge cluster trials with results have been published [3,4]. As experience with the conduct of these studies will grow, more strategies to prevent difficulties during trial conduct will likely be developed. Complete and consistent reporting of study results will be crucial for this process and useful reporting guidelines have already been published [78].

A possible limitation of our study is that our search was limited to PubMed and reference screening of previous reviews. Our review might therefore not include all articles about stepped-wedge cluster randomized trials that were published between 2012 and 2017. Still, we assume that the great majority of articles are listed in PubMed because most interventions were conducted in a medical setting. In addition, data extraction was performed by a single reviewer, which might have caused some errors in interpretation. However, uncertainties during data extraction were always resolved together with two other researchers. Last, we must consider a possible publication bias in the reporting of the results of stepped-wedge trials, as trials that failed to reach their planned sample size might be more likely to remain unreported. This is unfortunately a common problem in systematic reviews and our review could nevertheless provide a first—probably optimistic—estimation of overall recruitment rates into stepped-wedge cluster trials.

5. Conclusion

In conclusion, the results of our analyses show that about 31% of recently published stepped-wedge cluster trials do not reach their planned recruitment, although several design features could appear appealing for potential participants. Still, the stepped-wedge trial design may yield practical, ethical, and methodological advantages. Improved sample size reporting will be needed for more concrete conclusions.

Acknowledgments

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

Authors' contributions: F.A.E. was involved in the conception of the study, data collection, data analysis and interpretation, drafting of the article, and gave her final approval of the version to be published. R.H.H.G. was involved in the conception of the study, data analysis and interpretation, critical revision of the article, and gave his final approval of the version to be published. D.E.G. was involved in the interpretation of the data, the critical revision of the article, and gave his final approval of the version to be published. K.O.R. was involved in the conception of the study, data analysis and interpretation, critical revision of the article, and gave her final approval of the version to be published.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2018.11.013.

References

[1] The Gambia Hepatitis Study Group. The Gambia hepatitis intervention study. Cancer Res 1987;47:5782−7.
[2] Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. BMC Med Res Methodol 2016;16:69.
[3] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006;6:54.
[4] Beard E, Lewis JJ, Copas A, Davey C, Osrin D, Baio G, et al. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. Trials 2015;16:353.
[5] de Hoop E, van der Tweel I, van der Graaf R, Moons KG, van Delden JJ, Reitsma JB, et al. The need to balance merits and limitations from different disciplines when considering the stepped wedge cluster randomized trial design. BMC Med Res Methodol 2015;15:93.
[6] Friedman LM, Furberg CD, DeMets DL. Recruitment of study participants. In: Fundamentals of clinical trials. 5th ed. Switzerland: Springer; 2015.
[7] McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 2006;7:9.
[8] Lovato LC, Hill K, Hertert S, Hunninghake DB, Probstfield JL. Recruitment for controlled clinical trials: literature summary and annotated bibliography. Control Clin Trials 1997;18:328−52.
[9] van den Bor RM, Grobbee DE, Oosterman BJ, Vaessen PWJ, Roes KCB. Predicting enrollment performance of investigational centers in phase III multi-center clinical trials. Contemp Clin Trials Commun 2017;7:208−16.
[10] Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. Lancet Oncol 2006;7:141−8.
[11] Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs 2010;19:227−33.
[12] Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002;288:358−62.
[13] 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.
[14] Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials 2015;16:352.
[15] Amanyire G, Semitala FC, Namusobya J, Katuramu R, Kampiire L, Wallenta J, et al. Effects of a multicomponent intervention to streamline initiation of antiretroviral therapy in Africa: a stepped-wedge cluster-randomised trial. Lancet HIV 2016;3:e539−48.
[16] Bashour HN, Kanaan M, Kharouf MH, Abdulsalam AA, Tabbaa MA, Cheikha SA. The effect of training doctors in communication skills on women's satisfaction with doctor-woman relationship during labour and delivery: a stepped wedge cluster randomised trial in Damascus. BMJ Open 2013;3:e002674.
[17] Bennett PN, Fraser S, Barnard R, Haines T, Ockerby C. Effects of an intradialytic resistance training programme on physical function: a prospective stepped-wedge randomized controlled trial. Nephrol Dial Transplant 2016;31:1302−9.
[18] Chinbuah MA, Adjuik M, Cobelens F, Koram KA, Abbey M. Impact of treating young children with antimalarials with or without antibiotics on morbidity: a cluster-randomized controlled trial in Ghana. Int Health 2013;5:228−35.
[19] Chinbuah MA, Kager PA, Abbey M, Gyapong M, Awini E. Impact of community management of fever (using antimalarials with or without antibiotics) on childhood mortality: a cluster-randomized controlled trial in Ghana. Am J Trop Med Hyg 2012;87:11−20.
[20] Cisse B, Ba EH, Sokhna C, NDiaye JL, Gomis JF. Effectiveness of seasonal malaria chemoprevention in children under ten years of age in senegal: a stepped-wedge cluster-randomised trial. PLoS Med 2016;13:e1002175.
[21] Craine N, Whitaker R, Perrett S, Zou L. A stepped wedge cluster randomized control trial of dried blood spot testing to improve the uptake of hepatitis C antibody testing within UK prisons. Eur J Public Health 2015;25:351−7.
[22] Dryden-Peterson S, Bennett K, Hughes MD, Veres A, John O. An augmented SMS intervention to improve access to antenatal CD4 testing and ART initiation in HIV-infected pregnant women: a cluster randomized trial. PLoS One 2015;10:e0117181.
[23] Due TD, Thorsen T, Kousgaard MB, Siersma VD, Waldorff FB. The effectiveness of a semi-tailored facilitator-based intervention to optimise chronic care management in general practice: a stepped-wedge randomised controlled trial. BMC Fam Pract 2014;15:65.
[24] Duijster D, Sheiham A, Hobdell MH, Itchon G, Monse B. Associations between oral health-related impacts and rate of weight gain after extraction of pulpally involved teeth in underweight preschool Filipino children. BMC Public Health 2013;13:533.
[25] Durovni B, Saraceni V, Moulton LH, Pacheco AG, Cavalcante SC. Effect of improved tuberculosis screening and isoniazid preventive therapy on incidence of tuberculosis and death in patients with HIV in clinics in Rio de Janeiro, Brazil: a stepped wedge, cluster-randomised trial. Lancet Infect Dis 2013;13:852−8.
[26] Durovni B, Saraceni V, van den Hof S, Trajman A. Impact of replacing smear microscopy with Xpert MTB/RIF for diagnosing tuberculosis in Brazil: a stepped-wedge cluster-randomized trial. PLoS Med 2014;11:e1001766.
[27] Fink G, Robyn PJ, Sie A, Sauerborn R. Does health insurance improve health?: evidence from a randomized community-based insurance rollout in rural Burkina Faso. J Health Econ 2013;32:1043−56.
[28] Fuller C, Michie S, Savage J, McAteer J. The Feedback Intervention Trial (FIT)-improving hand-hygiene compliance in UK healthcare workers: a stepped wedge cluster randomised controlled trial. PLoS One 2012;7:e41617.
[29] Golden MR, Kerani RP, Stenger M. Uptake and population-level impact of expedited partner therapy (EPT) on Chlamydia trachomatis and Neisseria gonorrhoeae: the Washington State community-level randomized trial of EPT. PLoS Med 2015;12:e1001777.

[30] Golub JE, Cohn S, Saraceni V, Cavalcante SC, Pacheco AG, Moulton LH, et al. Long-term protection from isoniazid preventive therapy for tuberculosis in HIV-infected patients in a medium-burden tuberculosis setting: the TB/HIV in Rio (THRio) study. Clin Infect Dis 2015;60:639−45.
[31] Grande GE, Austin L, Ewing G, O'Leary N. Assessing the impact of a carer support needs assessment tool (CSNAT) intervention in palliative home care: a stepped wedge cluster trial. BMJ Support Palliat Care 2017;7:326−34.
[32] Gruber JS, Reygadas F, Arnold BF, Ray I. A stepped wedge, cluster-randomized trial of a household UV-disinfection and safe storage drinking water intervention in rural Baja California Sur, Mexico. Am J Trop Med Hyg 2013;89:238−45.
[33] Happ MB, Sereika SM, Houze MP, Seaman JB, Tate JA. Quality of care and resource use among mechanically ventilated patients before and after an intervention to assist nurse-nonvocal patient communication. Heart Lung 2015;44:408−415.e2.
[34] Haugen AS, Softeland E, Almeland SK, Sevdalis N. Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. Ann Surg 2015;261:821−8.
[35] Hemming K, Ryan R, Gill P, Westerby P. Targeted case finding in the prevention of cardiovascular disease: a stepped wedge cluster randomised controlled trial. Br J Gen Pract 2016;66:e758−67.
[36] Hill AM, McPhail SM, Waldron N, Etherton-Beer C. Fall rates in hospital rehabilitation units after individualised patient and staff education programmes: a pragmatic, stepped-wedge, cluster-randomised controlled trial. Lancet 2015;385:2592−9.
[37] Homan T, Hiscox A, Mweresa CK, Masiga D, Mukabana WR. The effect of mass mosquito trapping on malaria transmission and disease burden (SolarMal): a stepped-wedge cluster-randomised trial. Lancet 2016;388:1193−201.
[38] Hoogendijk EO, van der Horst HE, van de Ven PM, Twisk JW, Deeg DJ. Effectiveness of a Geriatric Care Model for frail older adults in primary care: results from a stepped wedge cluster randomized trial. Eur J Intern Med 2016;28:43−51.
[39] Horner C, Wilcox M, Barr B, Hall D, Hodgson G, Parnell P, et al. The longitudinal prevalence of MRSA in care home residents and the effectiveness of improving infection prevention knowledge and practice on colonisation using a stepped wedge study design. BMJ Open 2012;2:e000423.
[40] Jeddian A, Hemming K, Lindenmeyer A, Rashidian A, Sayadi L, Jafari N. Evaluation of a critical care outreach service in a middle-income country: a stepped wedge cluster randomized trial and nested qualitative study. J Crit Care 2016;36:212−7.
[41] Jordan S, Gabe-Walters ME, Watkins A, Humphreys I. Nurse-led medicines' monitoring for patients with dementia in care homes: a pragmatic cohort stepped wedge cluster randomised trial. PLoS One 2015;10:e0140203.
[42] Karkouti K, Callum J, Wijeysundera DN, Rao V. Point-of-care hemostatic testing in cardiac surgery: a stepped-wedge clustered randomized controlled trial. Circulation 2016;134:1152−62.
[43] Lashoher A, Schneider EB, Juillard C, Stevens K, Colantuoni E. Implementation of the world health organization trauma care checklist program in 11 centers across multiple economic strata: effect on care process measures. World J Surg 2016;41:954−62.
[44] Leontjevas R, Gerritsen DL, Smalbrugge M, Teerenstra S. A structural multidisciplinary approach to depression management in nursing-home residents: a multicentre, stepped-wedge cluster-randomised trial. Lancet 2013;381:2255−64.
[45] Mhurchu CN, Gorton D, Turley M, Jiang Y. Effects of a free school breakfast programme on children's attendance, academic achievement and short-term hunger: results from a stepped-wedge, cluster randomised controlled trial. J Epidemiol Community Health 2013;67:257−64.
[46] Monse B, Duijster D, Sheiham A, Grijalva-Eternod CS. The effects of extraction of pulpally involved primary teeth on weight, height and BMI in underweight Filipino children. A cluster randomized clinical trial. BMC Public Health 2012;12:725.
[47] Morrison LJ, Brooks SC, Dainty KN, Dorian P, Needham DM, Ferguson ND. Improving use of targeted temperature management after out-of-hospital cardiac arrest: a stepped wedge cluster randomized controlled trial. Crit Care Med 2015;43:954−64.
[48] NDiaye JL, Cisse B, Ba EH, Gomis JF. Safety of seasonal malaria chemoprevention (SMC) with sulfadoxine-pyrimethamine plus amodiaquine when delivered to children under 10 years of age by district health services in senegal: results from a stepped-wedge cluster randomized trial. PLoS One 2016;11:e0162563.
[49] Ononge S, Campbell OM, Kaharuza F, Lewis JJ, Fielding K. Effectiveness and safety of misoprostol distributed to antenatal women to prevent postpartum haemorrhage after child-births: a stepped-wedge cluster-randomized trial. BMC Pregnancy Childbirth 2015;15:315.
[50] Pickering BW, Dong Y, Ahmed A, Giri J. The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial. Int J Med Inform 2015;84:299−307.
[51] Rasmussen CD, Holtermann A, Bay H, Sogaard K, Birk Jorgensen M. A multifaceted workplace intervention for low back pain in nurses' aides: a pragmatic stepped wedge cluster randomised controlled trial. Pain 2015;156:1786−94.
[52] Rasmussen CD, Holtermann A, Jorgensen MB, Orberg A, Mortensen OS, Sogaard K. A multi-faceted workplace intervention targeting low back pain was effective for physical work demands and maladaptive pain behaviours, but not for work ability and sickness absence: stepped wedge cluster randomised trial. Scand J Public Health 2016;44:560−70.
[53] Rodriguez V, Giuffre C, Villa S, Almada G, Prasopa-Plaizier N. A multimodal intervention to improve hand hygiene in ICUs in Buenos Aires, Argentina: a stepped wedge trial. Int J Qual Health Care 2015;27:405−11.
[54] Roy A, Anaraki S, Hardelid P, Catchpole M, Rodrigues LC, Lipman M, et al. Universal HIV testing in London tuberculosis clinics: a cluster randomised controlled trial. Eur Respir J 2013;41:627−34.
[55] Scales DC, Golan E, Pinto R, Brooks SC, Chapman M. Improving appropriate neurologic prognostication after cardiac arrest. A stepped wedge cluster randomized controlled trial. Am J Respir Crit Care Med 2016;194:1083−91.
[56] Schultz TJ, Kitson AL, Soenen S, Long L, Shanks A, Wiechula R, et al. Does a multidisciplinary nutritional intervention prevent nutritional decline in hospital patients? A stepped wedge randomised cluster trial. e-SPEN J 2014;9:e84−90.
[57] Solomon E, Rees T, Ukoumunne OC, Metcalf B, Hillsdon M. The Devon Active Villages Evaluation (DAVE) trial of a community-level physical activity intervention in rural south-west England: a stepped wedge cluster randomised controlled trial. Int J Behav Nutr Phys Act 2014;11:94.
[58] Stern A, Mitsakakis N, Paulden M, Alibhai S. Pressure ulcer multidisciplinary teams via telemedicine: a pragmatic cluster randomized stepped wedge trial in long term care. BMC Health Serv Res 2014;14:83.
[59] Trajman A, Durovni B, Saraceni V, Menezes A, Cordeiro-Santos M. Impact on patients' treatment outcomes of XpertMTB/RIF implementation for the diagnosis of tuberculosis: follow-up of a stepped-wedge randomized clinical trial. PLoS One 2015;10:e0123252.
[60] Turner J, Kelly B, Clarke D, Yates P, Aranda S. A tiered multidisciplinary approach to the psychosocial care of adult cancer patients integrated into routine care: the PROMPT study (a cluster-randomised controlled trial). Support Care Cancer 2017;25:17−26.
[61] van de Steeg L, IJkema R, Langelaan M, Wagner C. Can an e-learning course improve nursing care for older people at risk of delirium: a stepped wedge cluster randomised trial. BMC Geriatr 2014;14:69.

[62] van den Broek IV, van Bergen JE, Brouwers EE, Fennema JS. Effectiveness of yearly, register based screening for chlamydia in The Netherlands: controlled trial with randomised stepped wedge implementation. BMJ 2012;345:e4316.
[63] van Leeuwen KM, Bosmans JE, Jansen AP, Hoogendijk EO. Cost-effectiveness of a chronic care model for frail older adults in primary care: economic evaluation alongside a stepped-wedge cluster-randomized trial. J Am Geriatr Soc 2015;63:2494−504.
[64] Wu MT, Wu CF, Chen BH. Behavioral intervention and decreased daily melamine exposure from melamine tableware. Environ Sci Technol 2015;49:9964−70.
[65] Zhan Z, van den Heuvel ER, Doornbos PM, Burger H, Verberne CJ. Strengths and weaknesses of a stepped wedge cluster randomized design: its application in a colorectal cancer follow-up study. J Clin Epidemiol 2014;67:454−61.
[66] Zwijsen SA, Gerritsen DL, Eefsting JA, Smalbrugge M, Hertogh CM, Pot AM. Coming to grips with challenging behaviour: a cluster randomised controlled trial on the effects of a new care programme for challenging behaviour on burnout, job satisfaction and job demands of care staff on dementia special care units. Int J Nurs Stud 2015;52:68−74.
[67] Zwijsen SA, Smalbrugge M, Eefsting JA, Twisk JW, Gerritsen DL, Pot AM. Coming to grips with challenging behavior: a cluster randomized controlled trial on the effects of a multidisciplinary care program for challenging behavior in dementia. J Am Med Dir Assoc 2014;15:531.e1−531.e10.
[68] Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. Clin Trials 2004;1:80−90.
[69] Foster JM, Sawyer SM, Smith L, Reddel HK, Usherwood T. Barriers and facilitators to patient recruitment to a cluster randomized controlled trial in primary care: lessons for future trials. BMC Med Res Methodol 2015;15:18.
[70] Haidich A-B, Ioannidis JPA. Patterns of patient enrollment in randomized controlled trials. J Clin Epidemiol 2001;54:877−83.
[71] Roy-Byrne PP, Sherbourne CD, Craske MG, Stein MB, Katon W, Sullivan G, et al. Moving treatment research from clinical trials to the real world. Psychiatr Serv 2003;54:327−32.
[72] Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol 2014;5:e45.
[73] Brierley G, Brabyn S, Torgerson D, Watson J. Bias in recruitment to cluster randomized trials: a review of recent publications. J Eval Clin Pract 2012;18:878−86.
[74] Rutterford C, Taljaard M, Dixon S, Copas A, Eldridge S. Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review. J Clin Epidemiol 2015;68:716−23.
[75] Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: a systematic review. Br J Anaesth 2015;115:699−707.
[76] 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;64:936−48.
[77] Kotz D, Spigt M, Arts ICW, Crutzen R, Viechtbauer W. Use of the stepped wedge design cannot be recommended: a critical appraisal and comparison with the classic cluster randomized controlled trial design. J Clin Epidemiol 2012;65:1249−52.
[78] 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.