Review Article

Planning and Analysis of Trials Using a Stepped Wedge Design

Part 26 of a Series on Evaluation of Scientific Publications

Stefan Wellek, Norbert Donner-Banzhoff, Jochem König, Philipp Mildenberger, Maria Blettner

Institute for Medical Biostatistics, Epidemiology and Informatics (IMBEI), Faculty of Medicine, Johannes Gutenberg University of Mainz; Institute for Medical Biostatistics, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany: Prof. Dr. rer. nat. Stefan Wellek

Department of General Practice/Family Medicine, University of Marburg: Prof. Dr. med. Norbert Donner-Banzhoff, MHSc

Institute for Medical Biostatistics, Epidemiology and Informatics (IMBEI), Faculty of Medicine, Johannes Gutenberg University of Mainz: Dr. sc. hum. Jochem König

Institute for Medical Biostatistics, Epidemiology and Informatics (IMBEI), Faculty of Medicine, Johannes Gutenberg University of Mainz: Philipp Mildenberger, MSc

Institute for Medical Biostatistics, Epidemiology and Informatics (IMBEI), Faculty of Medicine, Johannes Gutenberg University of Mainz: Prof. Dr. rer. nat. Maria Blettner

Summary

Background: The stepped-wedge design (SWD) of clinical trials has become very popular in recent years, particularly in health services research. Typically, study participants are randomly allotted in clusters to the different treatment options.

Methods: The basic principles of the stepped wedge design and related statistical techniques are described here on the basis of pertinent publications retrieved by a selective search in PubMed and in the CIS statistical literature database.

Results: In a typical SWD trial, the intervention is begun at a time point that varies from cluster to cluster. Until this time point is reached, all participants in the cluster belong to the control arm of the trial. Once the intervention is begun, it is continued without change until the end of the trial period. The starting time for the intervention in each cluster is determined by randomization. At the first time point of measurement, no intervention has yet begun in any cluster; at the last one, the intervention is in progress in all clusters. The treatment effect can be optimally assessed under the assumption of an identical correlation at all time points. A method is available to calculate the power and the number of clusters that would be necessary in order to achieve statistical significance by the appropriate type of significance test. All of the statistical techniques presented here are based on the assumptions of a normal distribution of cluster means and of a constant intervention effect across all time points of measurement.

Conclusion: The necessary statistical tools for the planning and evaluation of SWD trials now stand at our disposal. Such trials nevertheless are subject to major risks, as valid results can be obtained only if the far-reaching assumptions of the model are, in fact, justified.

Cite this as:
Wellek S, Donner-Banzhoff N, König J, Mildenberger P, Blettner M:
Planning and analysis of trials using a stepped wedge design—part 26 of a series on evaluation of scientific publications.
Dtsch Arztebl Int 2019; 116: 453–8. DOI: 10.3238/arztebl.2019.0453

The value of the principle of randomization to compare treatments and interventions remains undisputed in medical research, and randomized controlled trials (RCTs) are the acknowledged gold standard. Due to practical considerations, a number of variations have been developed in addition to the traditional RCT design, including cluster-randomized trials and the stepped-wedge design (SWD). In cluster-randomized, parallel-group trials—the prevailing type of cluster-randomized trial—groups of individuals (e.g. doctors' practices, school classes, regions), rather than individuals themselves, are randomized to receive the intervention. These groups are generally—as well as in the rest of this article—referred to as clusters.

Basic principle, model assumptions, estimator of treatment effect

In SWD trials, all individuals or clusters are observed first for a certain period of time under control conditions and then under intervention conditions until the end of the trial. Randomization is used to decide when the transition to the intervention is made. The number of consecutive points in time at which the outcome variable is observed is identical for all clusters, except for cases with missing values. Individuals may either be treated only once (cross-sectional SWD) or switch from control treatment to the intervention during the trial (open- versus closed-cohort SWD). In principle, the unit of observation in an SWD may be either an

TABLE 1

Stepped-wedge design with 5 possible intervention start times (T = 5) and one cluster per start time (n = 1)

Intervention start time	Cluster no.	Time of measurement					
		0	1	2	3	4	5
1	1	C	I	I	I	I	I
2	2	C	C	I	I	I	I
3	3	C	C	C	I	I	I
4	4	C	C	C	C	I	I
5	5	C	C	C	C	C	I

I: Intervention; C: Control

TABLE 2

Optimum weights of cluster means for n = 1, T = 5

a) $\rho = 2/3$ (values in table: numerators of fractions with denominator 160)

Cluster no.	Time of measurement					
	0	1	2	3	4	5
1	–20	32	19	6	–7	–20
2	–10	–23	29	16	3	–10
3	0	–13	–26	26	13	0
4	10	–3	–16	–29	23	10
5	20	7	–6	–19	–32	20

b) $\rho = 0$ (values in table: numerators of fractions with denominator 20)

Cluster no.	Time of measurement					
	0	1	2	3	4	5
1	0	4	3	2	1	0
2	0	–1	3	2	1	0
3	0	–1	–2	2	1	0
4	0	–1	–2	–3	1	0
5	0	–1	–2	–3	–4	0

individual or a cluster. In practice, however, SWD trials are usually conducted as an alternative to cluster-randomized trials.

In recent years, SWD trials have gained considerable popularity for planning scientific studies in medicine and health care research. This is reflected in the volume of medical literature on SWD trials: for example, a PubMed search using the keywords "stepped wedge" for a systematic review of the literature, covering publications from 2010 through 2014, yielded a total of 491 hits (1) (as of June 8, 2018). Among the health care research projects funded by the Innovation Fund of Germany's Federal Joint Committee (G-BA, *Gemeinsamer Bundesausschuss*) since 2015, there are several trials following an SWD.

SWD trials were described in the literature on experimental design as early as the late 1970s (2). The first large-scale study conducted and termed an SWD trial dates from 1987 (3). In the course of that project, a large-scale vaccination program was implemented in Gambia, for which 17 teams were formed. All the teams initially started a standard vaccination program. Hepatitis vaccination was adopted gradually, by one team at a time. The aim was to have vaccinated all children against hepatitis B viruses (HBVs) after approximately 4 years. The main reason given for proceeding in this way was logistics, including vaccine availability. The outcome was evaluated in terms of the incidence of liver tumors. Indirect evidence that vaccination effectively reduced HBV infection had already been confirmed before in a number of studies in high-risk groups. It was also known that HBV infection was a risk factor for liver cancer. According to the authors of the trial, it would be valuable to obtain direct evidence that vaccination reduced the incidence of liver tumors. With respect to this trial, there was also debate as to whether a 4-year traditional parallel-group trial should be conducted instead of the SWD. However, there were many organizational arguments against this, so the SWD design was chosen.

SWD trials are often also referred to as unidirectional crossover trials (4). This can be explained by the schedule shown in *Table 1* for clusters' transition from the control arm to the intervention arm of the trial for the standard case of a 2-armed SWD trial: each cluster begins in the control arm (C). The transition to the intervention treatment (I) occurs at the latest at the last follow-up time. This means that the only possible combinations for 2 consecutive points in time are C-C, C-I, and I-I, whereas I-C is impossible. This means that unlike in true, bidirectional crossover trials (5) there are no observational units for which the outcome variable is are measured without intervention after the end of the trial's intervention phase. Which cluster is allocated to which row of the scheme is determined by randomization. *Table 1* shows a specific example, and one recognizes the stepped-wedge shape between control and intervention periods that gives SWD trials their name. The number of clusters per start time need not be restricted to one, but it should remain constant over time where possible.

SWD trials are preferred over parallel-group or (true) crossover trials if it is assumed that the intervention will be considered worthwhile and beneficial, and those planning the trial cannot (or do not want to) justify interrupting the intervention once it has been started. The SWD also has the advantage that the intervention only needs to be started in a few clusters at once, which from an organizational perspective is often a very important factor. For example, in the trial conducted in Gambia described above it was not possible, for organizational reasons, to begin HBV vaccination in all 60 000 children (50% of the study population) at the same time.

Table 2 shows optimum weighting for the trial design shown in *Table 1*, as an example. The results hold under the following simplifying conditions (4, 6):

- **Condition 1:** Analysis is performed in 2 steps; the first one consists of calculating averages for each cluster and point in time. Subsequent analysis relates to these aggregate values alone, and for these the basic distributional assumptions are required to hold.
- **Condition 2:** Cluster means are normally distributed (at least approximately) with a variance, being independent of both point in time and treatment.
- **Condition 3:** Cluster means are correlated between times at which parameters are measured. However, the magnitude of this correlation depends on neither temporal distance nor the type of treatment (control or intervention). In principle, correlations depend on whether and how often repeated measurements are taken from the same individual.
- **Condition 4:** As an average over the population of all clusters, the clusters' arithmetical means are the sum of a period effect specific to the time at which parameters are measured and the time-independent effect (hereafter referred to as θ) of the treatment being investigated (the intervention).

Using these conditions, the standard error (*stderr*) of the optimum estimator of the treatment effect can be calculated exactly. A relatively simple formula can be used *(Box 1)* to do so for arbitrary numbers of intervention start times (T) and clusters (n) that transition from the control phase to the intervention phase at the same time. This formula can be used to calculate a confidence interval for the estimated treatment effect obtained by analyzing an SWD trial. The entries in the table in *Box 1* show how the width of this confidence interval, and therefore the statistical precision of the estimator, is affected by the parameters underlying the trial design.

Significance testing, power, and sample size planning

Just as simple as calculating the limits of confidence intervals is statistical testing of the null hypothesis that the "true" treatment effect θ (i.e. the effect without superposition of chance deviations) is 0.

When planning an SWD, it is important to realize that the procedure to be used for calculating power cannot be converted into a simple formula for the number n of clusters that start the intervention at the same time. As shown in the formula given in *Box 1*, the standard error of θ_{est}, as well as the power, depend not only on the variance (σ^2) of the cluster means and the number of clusters (n), but also on the number of intervention start times (T) and the correlation between repeated measurements in a single cluster. The conclusions to be drawn from comparative investigations into the efficiency of various SWD trials, cluster-randomized parallel-group trials, and individually randomized trials therefore depend on the number of participating individuals, the number of times measurements are repeated per individual, the number of clusters starting intervention at the same time, and the number of possible starting times.

The *eBox* compares cluster-randomized SWD trials and parallel-group trials in various scenarios in which both the variance σ^2 of the cluster means and their correlation ρ between time points are functions of the

BOX 1

Error variance (SE^2) of the optimum estimator of the treatment effect

Symbols:

T = no. of times outcome measured or no. of intervention start times

n = no. of clusters beginning the intervention at the same time

σ^2 = variance of cluster means

ρ = correlation coefficient between measurements for a single cluster at 2 different points in time

stderr = standard error

$$stderr = \sqrt{\frac{\sigma^2}{n}\frac{(1-\varrho)(\varrho T+1)}{(T-1)(T+1)(\varrho T+2)/12}}$$

(Source: Rhoda et al. [7]; Hughes et al. [8])

Width of 95% confidence interval (CI) for θ in for different combinations of the design parameters T, n, and ρ when $\sigma^2 = 1$

T	ρ	n	Width
2	0.10	1	5.49
		5	2.46
		50	0.78
2	0.50	1	4.53
		5	2.02
		50	0.64
2	0.90	1	2.13
		5	0.95
		50	0.30
5	0.10	1	2.04
		5	0.91
		50	0.29
5	0.50	1	1.73
		5	0.77
		50	0.24
5	0.90	1	0.81
		5	0.36
		50	0.11

BOX 2

Example of the planning and statistical analysis of an SWD trial (according to [10])

Aim

- To obtain evidence that the quality of life of frail senior citizens can be improved by geriatric training for nursing staff

Study procedure

- Start of intervention (training for nursing staff according to the Chronic Care Model [CCM], [11]) 6, 12, 18, or 24 months after the beginning of the project (<–> $T = 4$); each cluster is a practice caring for 20 patients each; intervention started in 8 practices at each of the 4 points in time

Outcome parameter

- Physical Composite Score (PCS) of the Short Form Quality-of-Life Questionnaire (SF-12) (12); high score to be rated as favorable

Assumptions for power calculation

- Cluster means are normally distributed with variance $\sigma^2 = 3.48$ and correlation $\rho = 0.66$ between repeated measurements.
- The mean improvement in score achieved by the intervention for all practices is $\theta = 1.00$.
- The significance level is set at $\alpha = 5\%$ (2-tailed).

Power when intervention is started in 8 practices every 6 months

- The formula shown in *Box 1* is used to calculate the standard error of the estimated intervention effect at 0.3046. Hence, the probability that the corresponding test yields a significant finding (power) is 90.71%.

Analysis of dataset shown in *Table 3*

- For the clustered PCS scores listed in *Table 3*, complete statistical analysis, including variance and correlation between points in time, yields the following values:
  - Estimated effect of intervention (± standard error): $\theta_{est} = 0.1717 \pm 0.2901$
  - 95% confidence interval: [-0.3969; 0.7403]
  - p-value (2-tailed) to test the null hypothesis that $\theta = 0$: $p = 0.5539$

In view of these results, the outcome of the trial is negative. In other words, one cannot conclude from the trial data that the intervention has a positive effect on patients' physical quality of life.

so-called intraclass correlation coefficients (ICCs) within clusters. If one measures the efficiency of a design in terms of the total number of clusters required to detect an effect of $\theta = 0.25$ with a probability (power) of 90% in a test at the usual significance level $\alpha = 0.05$ (2-tailed), the findings are as follows: in these settings, SWD trials are more efficient than parallel-group trials unless ICC values are very low *(eFigure)*. However, it should be noted that there is a fundamental qualitative change in this picture if, unlike in the scenarios investigated in the *eBox*, the number of measurements to be performed at each point in time in individual clusters is the same for all designs. Then, parallel-group trials are substantially more efficient than SWD trials unless ρ is very large.

How to proceed when outcome parameter variance and time-dependent correlation are unknown

The facts and conclusions on statistical planning and analysis of SWD trials that are summarized here hold under the assumption that both the variance σ^2 between clusters and the correlation coefficient ρ between measurements for one cluster at different times are known quantities. Whenever an SWD trial needs to be analyzed without this prior knowledge, a much more complicated statistical procedure must be used allowing to estimate in addition to the treatment effect θ, the parameter of primary interest, also σ^2 and ρ from the data obtained in the study.

This extended procedure was used to obtain the findings shown in *Box 2* by analyzing the sample SWD trial described in *Table 3*. Full details of the procedure can be found in the documentation relating to software programs for analyzing mixed linear models such as the SAS PROC MIXED Procedure (9). Such complex statistical models should also be used to analyze trials in which correlations between repeated measurements are assumed to be due to intraindividual effects. Among others, this this typically implies that the variation between clusters cannot be described any longer in accordance with Condition 2 by a single dispersion parameter. Even when σ^2 and ρ have to be estimated as part of the analysis of an SWD trial, the trial is usually (4) planned as described above for settings in which σ^2 and ρ are known.

Discussion

Like true crossover trials, SWD trials yield longitudinal data, since the outcome variable is measured repeatedly in each observational unit (cluster). Another common

feature of these two trial designs is that they entail a high risk of producing misleading results: if the very restrictive underlying assumption that there is no interaction between intervention effect and measurement time is incorrect, the treatment effect cannot be estimated without bias. It is very important to bear this caveat in mind when planning and interpreting trials.

Alternatively, an SWD trial can also be regarded as a sequence of $T + 1$ parallel-group trials, with a constant sample size (n) but a proportion of observational units allocated to the invention arm that varies over time (increasing from 0 to 100%).

Even if the cluster means obtained in an SWD trial are normally distributed, approximations are usually needed to test hypotheses concerning the treatment effect. Various approaches are available for this purpose. They yield different results, and none can be said to be generally preferable to the others. Furthermore, as is often the case when analyzing longitudinal data, SWD trials are usually analyzed on the basis of assumptions about the correlation structure that greatly simplify the true situation (equicorrelation model).

The main practical incentive stated for conducting an SWD trial is usually a wish to give all patients access to the intervention being investigated, at least in the last period of the trial. This is particularly desirable if there is information available suggesting that the intervention is effective. This was the pivotal argument for the trial conducted in Gambia: those conducting the trial were sure that vaccination was essentially effective.

SWD trials are thus an alternative to conventional trials when there are practical limitations that preclude carrying out cluster-randomized trials. Conducting a cluster-randomized trial would require the nursing staff training, etc. associated with the trial intervention to be performed swiftly enough for the intervention to be started in all trial patients simultaneously.

If statistical analysis of the trial is performed correctly (and at an appropriate level of complexity), basic methodological requirements can be met. Although the conditions required for a valid statistical evaluation of the treatment effect can be specified clearly in theory, in practice they are difficult to test.

TABLE 3

Raw data for example described in Box 2

Practice no.	Intervention start time	Time measured (months)				
		0	6	12	18	24
1	6 months	48.05	48.75	49.60	50.75	50.40
2		51.30	53.00	51.10	51.45	51.95
3		48.30	47.05	47.70	47.40	47.00
4		52.00	51.75	51.65	52.15	49.25
5		51.45	52.25	52.55	52.25	51.50
6		51.20	53.90	53.25	53.30	52.65
7		50.35	50.60	52.25	50.50	52.20
8		52.10	48.90	50.20	53.00	48.85
9	12 months	50.90	52.05	52.25	51.05	52.85
10		50.70	50.70	49.40	49.45	51.45
11		49.05	49.05	49.10	48.45	48.70
12		49.90	48.95	49.55	49.80	49.20
13		48.85	51.75	52.25	50.45	50.60
14		48.65	49.55	49.80	49.90	51.10
15		47.40	48.25	48.40	48.85	49.35
16		48.90	49.05	48.90	48.05	47.35
17	18 months	53.60	52.75	50.50	53.60	51.90
18		48.25	50.30	48.15	47.10	50.10
19		50.50	51.10	49.00	49.45	51.75
20		47.55	47.55	47.60	48.25	47.45
21		49.10	49.00	49.70	49.45	48.45
22		51.50	51.65	50.90	49.70	51.65
23		48.60	48.40	48.00	47.10	50.75
24		49.50	49.50	50.35	51.45	50.95
25	24 months	51.65	50.30	48.10	49.75	51.05
26		48.45	52.05	49.15	50.65	49.20
27		50.35	51.00	50.00	51.20	50.20
28		51.20	50.70	51.45	52.00	50.15
29		47.20	49.20	49.70	48.55	49.10
30		46.10	47.20	47.60	45.75	45.70
31		51.35	48.05	51.25	49.30	49.70
32		49.25	48.00	48.45	51.75	49.70

Conflict of interest statement
The authors declare that no conflict of interest exists.

Manuscript received on 12 October 2018, revised version accepted on 15 April 2019.

Translated from the original German by Caroline Shimakawa-Devitt, M.A.

References

1. 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.
2. Cook TD, Campbell DT: Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin 1979.
3. Gambia Hepatitis Study Group: The Gambia Hepatitis Intervention Study. Cancer Res 1987; 47: 5782–7.
4. Hussey MA, Hughes JP: Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007; 28: 182–91.
5. Wellek S, Blettner M: On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2012; 109: 276–81.
6. Hemming K, Lilford R, Girling AJ: Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med 2015; 34: 181–96.
7. Rhoda DA, Murray DM, Andridge RR, Pennell ML, Hade EM: Studies with staggered starts: multiple baseline designs and group-randomized trials. Am J Public Health 2011; 101: 2164–9.
8. Hughes JP, Granston TS, Heagerty PJ: Current issues in the design and analysis of stepped wedge trials. Contemp Clin Trials 2015; 45 (Pt. A): 55–60.

Key messages

- The main potential benefit of the stepped-wedge design (SWD) is that it makes it possible to compare an intervention that cannot be administered to all trial patients at the same time in a randomized controlled trial.
- The SWD bears only a slight similarity to the conventional crossover trial design, as the outcome variable is not measured for each observational unit under both control and intervention conditions.
- The major drawback of the SWD is that correct statistical analysis is only possible if the effect of the intervention investigated is guaranteed to depend on neither the duration of the intervention nor when it is started within the trial.
- Some authors prefer SWD trials to parallel-group trials regardless of feasibility issues, as it makes it possible to give all trial patients access to the intervention in at least one period of the trial.
- Like conventional, true crossover trials, easily interpretable models and procedures for the statistical analysis of SWD trials exist only if data is approximately normally distributed. There are several competing approaches for trials with binary or categorical data.

9. SAS: SAS/STAT(R) 14.1 User's guide. The MIXED procedure. support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_mixed_details.htm (last accessed on 20 May 2019).
10. Hoogendijk EO, van der Horst HE, van de Ven PM, et al.: 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.
11. Coleman K, Austin BT, Brach C, Wagner EH: Evidence on the Chronic Care Model in the new millennium. Health Affairs 2009; 28: 75–85.
12. Brook RH, Ware JEJ, Davies-Avery A , et al.: . Overview of adult health measures fielded in Rand's health insurance study. Med Care 1979 ; 17: 1–131.

Corresponding author:
Prof. Dr. rer. nat. Maria Blettner
Institute for Medical Biostatics, Epidemiology and Informatics
Johannes Gutenberg University Mainz
Obere Zahlbacher Str. 69
55131 Mainz, Germany
blettner@uni-mainz.de

Cite this as:
Wellek S, Donner-Banzhoff N, König J, Mildenberger P, Blettner M: Planning and analysis of trials using a stepped wedge design—part 26 of a serieson evaluation of scientific publications. Dtsch Arztebl Int 2019; 116: 453–8. DOI: 10.3238/arztebl.2019.0453

► **Supplementary material:**
eBox, eFigure, eTable:
www.aerzteblatt-international.de/19m0453

Supplementary material to:

Planning and Analysis of Trials Using a Stepped Wedge Design

Part 26 of a Series on Evaluation of Scientific Publications

by Stefan Wellek, Norbert Donner-Banzhoff, Jochem König, Philipp Mildenberger, and Maria Blettner

Dtsch Arztebl Int 2019; 116: 453–8. DOI: 10.3238/arztebl.2019.0453

eBOX

Planning and analyzing stepped-wedge design (SWD) trials

This *eBox* compares sample cluster-randomized SWD and parallel-group trials in various scenarios in which both the variance σ^2 of cluster means and their correlation ρ between times at which the outcomes are measured depend on the intraclass correlation coefficient (ICC) within clusters.

The scenarios result from the following assumptions (for full specifications see the *eTable*):

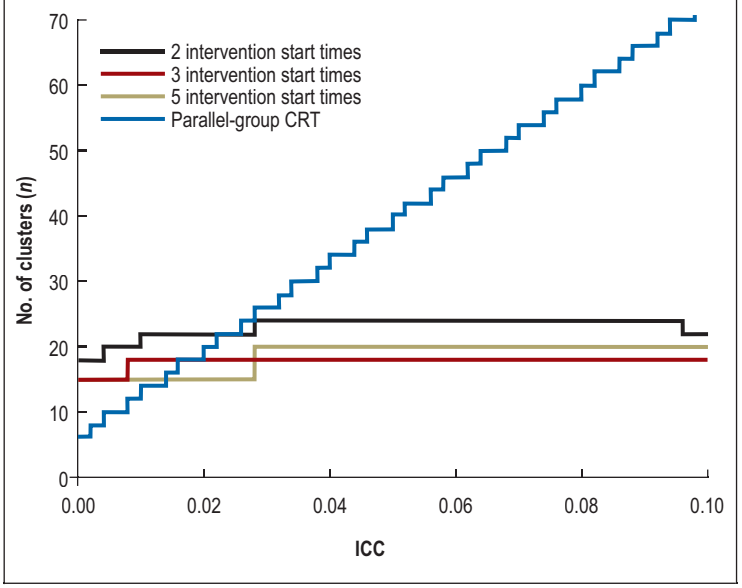
- Each individual is observed only once.
- The number of individuals participating in a cluster per month is fixed at 10.
- Study duration is set at 12 months.
- SWD trials with 2, 3, or 5 intervention start times and a parallel-group CRT (cluster-randomized trial) with recruitment lasting 12 months are compared with each other.
- The number of study periods is therefore 3, 4, or 6, and 1 for the parallel-group CRT.
- Study periods last 4, 3, or 2 months, and 12 months for the parallel-group CRT.
- The number of patients observed per cluster and study period is 40, 30, 20, or 120.
- Random cluster effects are constant over all periods.
- Both the variance of the cluster means and their correlation between times at which the outcome is measured depend on the ICC as follows: $\sigma^2 = ICC + (1 - ICC)/m$, $\rho = ICC/\sigma^2$, where m is the number of individuals per time of measurement.

The *eTable* shows the total number of clusters needed as a function of ICC, to reveal an effect of $\theta = 0.25$ with a probability (power) of 0.90 when the test described above is performed as a 2-tailed test with significance level $\alpha = 0.05$.

Because all clusters recruit over an equal period, 12 months, the number of individuals observed without intervention and under and intervention conditions is a constant multiple—120-fold—of the number of clusters. The *eFigure* shows that the following points are true for the design options selected here:

- Design efficiency depends on ICC.
- SWD trials with many intervention start times are more efficient than those with few. However, in some cases it will not be possible for practical reasons to choose the maximum number of periods per start time in one cluster, as this would make the periods too short.
- Costs/expenditure will usually be fixed per cluster. In such cases the scenarios described here will not be decisive but may be significant if there are various options available regarding total length of recruitment or recruitment rate for each design.
- Finally, the assumption that period effect is additive and a number of requirements concerning correlation structure in parallel-group CRTs (cluster-randomized trials) need not be met, so parallel-group CRTs have a lower risk of bias and higher level of evidence, other things being equal.

eFIGURE



Number of clusters required for various SWDs and a parallel-group CRT, by ICC

CRT: Cluster-randomized trial; ICC: Intraclass correlation coefficient; SWD: Stepped-wedge design

eTABLE

Specifications of compared trial designs

	Design			
	SWD	SWD	SWD	Parallel-group CRT
No. of intervention start times	2	3	5	1
No. of individuals per cluster per month	10	10	10	10
Study duration (months)	12	12	12	12
No. of study periods	3	4	6	1
Duration of study periods (months)	4	3	2	12
No. of individuals per period per cluster (m)	40	30	20	120
No. of clusters per start time	Depends on correlation parameters			
Marginal interindividual variance	1	1	1	1
Variance σ^2 of cluster means				
For ICC = 0.01	0.035	0.043	0.060	0.018
For ICC = 0.05	0.074	0.082	0.098	0.058
Correlation ρ between cluster means				
For ICC = 0.01	0.29	0.23	0.17	Irrel.
For ICC = 0.05	0.68	0.61	0.51	Irrel.

$\sigma^2 = \mathrm{ICC} + \frac{1}{m}(1 - \mathrm{ICC}), \rho = \frac{\mathrm{ICC}}{\sigma^2}$

CRT: Cluster-randomized trial; ICC: Intraclass correlation coefficient; Irrel.: Irrelevant; SWD: Stepped-wedge design