Reporting of The CONSORT extension for Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration

Hemming, Karla; Taljaard, Monica; McKenzie, Joanne E; Hooper, Richard; Copas, A; Thompson, JA ; Dixon-Woods, M; Aldcroft, A ; Doussau, A ; Grayling, M ; Kristunas, Caroline; Goldstein, CE; Campbell, MK; Girling, Alan; Eldridge, S; Campbell, MJ; Lilford, Richard; Weijer, C ; Forbes, A; Grimshaw, JM

Reporting of The CONSORT extension for Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration

Journal:	*BMJ*
Manuscript ID	BMJ.2017.042390.R1
Article Type:	Research methods and reporting
BMJ Journal:	BMJ
Date Submitted by the Author:	16-Mar-2018
Complete List of Authors:	Hemming, karla; birmingham Taljaard, Monica; Ottawa Hospital Research Institute, Clinical Epidemiology Mckenzie, Jo; Monash Hooper, Richard; Queen Mary University of London, Institute of Health Sciences Education Copas, Andrew; University College London, Department of Infection and Population Health Thompson, Jennifer; London School of Hygiene and Tropical Medicine Faculty of Epidemiology and Population Health Dixon-Woods, Mary; University of Leicester, Health Sciences Aldcroft, Adrian; BMJ, Doussau, Adelaide; McGill University Health Centre Grayling, Michael; University of Cambridge Department of Engineering Kristinus, Caroline; University of Leicester Medical School Goldstein, Cory; Western University, Philosophy Campbell, Marion; University of Aberdeen, Health Services Research Unit Girling, Alan; Institute of Applied Health Research Eldridge, Sandra; Queen Mary, University of London, Primary Care and Public Health Campbell, Michael; University of Sheffield, Health Services Research ScHARR Lilford, Richard; University of Warwick, Weijer, Charles; University of Western Ontario, Philosophy Forbes, Andrew; Monash University, School of Public Health and Preventive Medicine Grimshaw, JM; Ottawa Hospital Research Institute, Clinical Epidemiology Program
Keywords:	CONSORT, Stepped-wedge, cluster, reporting guideline

Reporting of Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration

K Hemming[1], M Taljaard[2], JE McKenzie[3], R Hooper[4], A Copas[5], JA Thompson[5 6], M Dixon-Woods[7], A Aldcroft[8], A Doussau[9], M Grayling[10], C Kristunas[11], CE Goldstein[12], MK Campbell[13], A Girling[14], S Eldridge[15], MJ Campbell[16], RJ Lilford[17], C Weijer[18], A Forbes[19], JM Grimshaw[2 20]

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. k.hemming@bham.ac.uk;

[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling Avenue, Ottawa, Ontario, Canada; and School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Canada. mtaljaard@ohri.ca;

[3] School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. Joanne.mckenzie@monsh.edu;

[4]Pragmatic Clinical Trials Unit, Centre for Primary Care & Public Health, Queen Mary University of London, London, UK. r.l.hooper@qmul.ac.uk;

[5]London Hub for Trials Methodology Research, MRC Clinical Trials Unit at University College London, London, UK. a.copas@ucl.ac.uk;

[6]Department for Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. jennifer.thompson@lshtm.ac.uk;

[7]THIS Institute, University of Cambridge, Cambridge Biomedical Campus, Bay 13 Clifford Allbutt Building, Cambridge CB2 OAH. md753@medschl.cam.ac.uk;

[8]BMJ Publishing Group, London, UK. aaldcroft@bmj.com

[9]Biomedical Ethics Unit, McGill University School of Medicine, Montreal, Canada. Adelaide.doussau@mail.mcgill.ca;

[10]MRC Biostatistics Unit, Cambridge, UK. michael.grayling@mrc-bsu.cam.ac.uk;

[11]Department of Health Sciences, University of Leicester, Leicester, UK. Cak21@le.ac.uk;

[12]Rotman Institute of Philosophy, Western University, London, Canada. cgoldst2@uwo.ca;

[13] Health Services Research Unit, University of Aberdeen, Aberdeen, UK. m.k.campbell@abdn.ac.uk;

[14]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. a.j.girling@bham.ac.uk;

[15]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK. s.eldridge@qmul.ac.uk;

[16]ScHARR, University of Sheffield, Sheffield, UK. m.j.campbell@sheffield.ac.uk;

[14]University of Warwick, Coventry, UK. R.J.Lilford@warwick.ac.uk;

[18]Rotman Institute of Philosophy, Western University, London, Canada. cweijer@uwo.ca;

[19]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. Andrew.Forbes@monash.edu;

[20] Department of Medicine University of Ottawa, Ottawa, Canada. jgrimshaw@ohri.ca.

Acknowledgements

With acknowledgement to those who participated in the Delphi survey and Peter Chilton who provided administrative support.

Author contributions

KH led the development of the project, the Delphi survey, the consensus meeting, drafting of the items; and wrote the first draft of the paper. MT, JG, AF, CW and JM made a substantial contribution to all stages of the project. CW and MT gave insight into the ethical aspects of the project. KH, MT, JM, CW and AF contributed to the development of the items. SE and MJC gave critical insights into reporting guidelines. AF and JMG provided project leadership and guidance. JMG facilitated the consensus meeting. RL provided critical insight into the early stages of the project. All authors participated in the consensus meeting and commented on the draft paper.

Funding

This research was funded by the Australian National Health and Medical Research Council (NHMRC) project grant (1108283) and also partly funded by the UK NIHR Collaborations for Leadership in Applied Health Research and Care West Midlands initiative. Mary Dixon-Woods is funded by a Welcome Trust Senior Investigator award WT097899. Jennifer A Thompson is funded by the Medical Research Council Network of Hubs for Trials Methodology Research (MR/L004933/1-P27). Jeremy Grimshaw holds a Canada Research Chair in Health Knowledge Transfer and Uptake. Charles Weijer holds a Canada Research Chair. Joanne E McKenzie holds an NHMRC Australian Public Health Fellowship (1072366). Karla Hemming holds an NIHR Senior Research Fellowship (SRF-2017-002).

Competing Interests

We have read and understood the BMJ Group policy on declaration of interests and declare the following interests: none.

Summary

This document presents the Consolidated Standards Of Reporting Trials (CONSORT) extension for the stepped-wedge cluster randomised trial (SW-CRT). The SW-CRT involves randomisation of clusters to different sequences that dictate the order (or timing) at which each cluster will switch to the intervention condition. The development of this statement was motivated by the unique design characteristics of this study design, including the need to allow for time effects and because the design is increasingly being used. The guideline was developed using a Delphi survey and consensus meeting; and is informed by the CONSORT statements for individually and cluster randomised trials. Reporting items along with explanations and examples are provided. We include a glossary of terms, and explore the key properties of the SW-CRT which require special consideration in their reporting.

Introduction

The CONSORT (Consolidated Standards Of Reporting Trials) statement, initially published in 1996 and updated in 2001 and 2010, outlines essential items to be reported in a parallel arm individually randomised trial [Begg 1996; Rennie 2001; Schulz 2010]. The CONSORT extension for cluster randomised trials, initially published in 2004 and updated in 2012, extended this guidance for trials in which groups of individuals (clusters – for a full glossary of terms see Table 1) are randomised to different treatment conditions [Campbell 2004; Campbell 2012]. In recent years, a novel type of cluster randomized design - the stepped-wedge cluster randomised trial (SW-CRT) - has become increasingly popular [Brown 2006; Mdege 2011, Martin 2017]. The SW-CRT involves randomisation of clusters to different sequences. These sequences dictate the order (or timing) with which each cluster will switch to the intervention condition.

The basic components of the design, as well as illustrative examples of studies which have used this design, have been described previously [Hemming 2015]. The unit of randomisation in these trials is the cluster with clusters (or groups of clusters) allocated to different sequences (as opposed to different "arms" in a parallel trial). These sequences specify the number of time periods spent in the control condition and the number of time periods in the intervention condition. In Figure 1, for example, there are four groups of clusters allocated to four different sequences. Each cluster contributes data to the analysis from each measurement period. In the example in Figure 1 there are five measurement periods. The point at which a cluster switches to the intervention condition is called a "step". Sometimes a transition period is built into the design, during which the intervention is implemented in the cluster.

This design has numerous methodological complexities, including potential confounding with time [Hemming 2017]; changes in correlation structures over time [Girling 2016; Hooper 2016; Kasza 2017]; the possibility of within cluster contamination over time [Copas 2015]; the possibility of time varying treatment effects [Davey 2015, Hemming 2017]; and different design variations [Prost 2015; Hargreaves 2015], all of which increase the complexity of reporting [Hemming 2015]. Perhaps unsurprisingly, systematic reviews examining the adequacy of reporting of SW-CRTs have revealed numerous inadequacies, including absence of essential details of the design, inconsistent use of terminology [Brown 2006; Mdege 2011; Martin 2016; Grayling 2017; Taljaard 2017]; frequent lack of clarity in reporting of adjustment for time effects [Hemming 2017; Martin 2017]; as well as frequent failure to report ethical review and trial registration [Taljaard 2017]. These findings suggest there is a need for a specific reporting guideline for this trial design. Here we report the results of a consensus process to develop an extension to the CONSORT statement for use with SW-CRTs. The ultimate goal of this extension is to improve the standards of reporting of this important and increasingly used research design.

Scope of this statement

This reporting statement should be followed when reporting results from any SW-CRT. In line with other CONSORT statements this guideline includes the minimum set of items that should be reported; it is not intended to be a comprehensive list of all possible items that could be reported.

A wide variety of terms have been used to describe aspects of the SW-CRT design. For the purpose of this reporting statement, the key components of the design are defined in Figure 1 and a glossary of terms is provided in Table 1. Generally, SW-CRTs have a minimum of three sequences. Trials with two sequences and three periods, for example, a two-arm cluster randomised trial in which both arms are initially observed under the control condition and in addition, the control arm adopts the intervention during a third measurement period might also technically be considered a SW-CRT. The statement was developed for comparisons of two treatment conditions. So as to take a broader perspective on the range of designs that can be included, we are not restricting our definition to designs with all clusters initiating in the control condition and ending up in the intervention condition [Hooper 2016].

Extending the CONSORT statement to SW-CRTs

We developed this extension using methods recommended for developing reporting guidelines [Moher 2010]. We registered our protocol on the EQUATOR website in July 2015 [Hemming 2015c] and identified relevant and related reporting guidelines. We conducted several systematic reviews of published SW-CRTs examining aspects of reporting and methodological conduct and undertook a consensus process.

Results from systematic reviews examining SW-CRT methods and reporting

We conducted several systematic reviews in advance of the consensus process [Martin 2016; Taljaard 2017; Grayling 2017; Martin 2017]. Martin et al. (2016) found that the SW-CRT is increasingly being used and that the majority of trials are conducted in advanced economies and in healthcare settings; although a significant minority are conducted in lower middle income settings; with most trials having less than 20 clusters and a smaller number of time periods [Martin 2016].

Reviews of the quality of reporting of sample size and analysis methods revealed incomplete or inadequate reporting overall, and specifically, lack of reporting of how time effects and extended correlation structures were incorporated both at the design and analysis stages [Davey 2015; Martin 2016; Grayling 2017; Martin 2017]. Reviews of the ethical conduct and reporting revealed that many SW-CRTs do not report research ethics review; do not clearly identify from whom and for what consent was obtained; and a significant number do not pre-register with a trial registration database [Taljaard 2017]. Reviews of the methodological literature have identified several key aspects of the SW-CRT which are associated with bias [Barker 2016; Martin 2017]. Clear reporting of these aspects is essential to facilitate interpretation of trial results in published reports.

Firstly, time is a potential confounder in a SW-CRT and requires special consideration both at the design and analysis stage [Hughes 2007; Hemming 2017]. Secondly, as the SW-CRT is a longitudinal and clustered study, correlation structures are more complex than those of a parallel CRT carried out at a single cross-section in time [Hooper 2016]. Thirdly, some SW-CRTs are at risk of within-cluster contamination. Within-cluster contamination can arise either when outcomes in the intervention condition are obtained from participants who are yet to be exposed to the intervention, or alternatively, when outcome assessments in the control condition are from participants already exposed to the intervention [Copas 2015]. Contamination arising from observations yet to be fully exposed to the intervention condition can be allowed for by building in transition periods into the design; or by modelling these effects (referred to as lag effects) [Hughes 2015]. Interactions between time and treatment can also arise. These time varying effects are more likely to arise when the intervention is not continuously delivered, does not create a permanent change, or where its impact might wane or grow over time [Davey 2015].

These complexities differ according to the many different ways that a SW-CRT can be conducted, including whether the same or different participants are repeatedly assessed, whether participants are continuously recruited and the duration of their exposure, and whether a complete enumeration of the cluster is taken [Hemming 2015; Copas 2015]. With practical and ethical considerations also in play, the adoption of this design requires careful justification [Prost 2015; Doussau 2016]. A summary of key methodological issues which need extra consideration when reporting a SW-CRT is presented in Table 2.

Consensus process

Members of the working group (KH, MT, JEM, AF, CW, JG) identified items from the original CONSORT statement which required modification; considered whether the modification used in the cluster extension was appropriate; and if not, proposed a modified version for the item. In a modified Delphi process (December 2016), we invited 64 subject experts to consider, rate and comment on the proposed modifications of whom 42 completed the survey. We summarised responses from the survey and circulated a second draft of the proposed modifications in advance

of a one-day consensus meeting (Liverpool May 2017). The CONSORT stepped-wedge consensus group (20 people in total all listed as authors of this statement) consisted of members of the working group and those with expertise in trial design, journal editors (BMJ Open, Trials, Clinical Trials, and BMJ Quality and Safety Improvement), ethicists, statisticians, methodologists, and developers of reporting guidelines (cluster trials, pilot and feasibility trials and equity trials). At the meeting, proposed wording, examples and elaboration text were discussed and amended. The proposed final wording was then circulated and final comments incorporated.

The CONSORT extension for Stepped-Wedge Cluster Randomised Trials

A checklist detailing the 26 items to be reported in the publication of a SW-CRT is presented in Table 3. Some items have not been modified from the original CONSORT statement, some are modified, and some are new. Similar to the CONSORT extension for cluster trials, Item 10 (Implementation of randomisation) has been replaced by Items 10a, 10b and 10c. In recognition of the under-reporting of key ethical aspects of these trials, a new item on Research Ethics Review has been added as Item 26 (as was added to the CONSORT extension for pilot and feasibility studies [Eldridge 2016]). For ease of interpretation in the elaboration that follows, we provide the original CONSORT wording, the wording of the CONSORT extension for cluster randomised trials, as well as the wording for the SW-CRT extension. Table 4 summarises key changes to the original CONSORT statement and substantial deviations from the CONSORT extension for cluster randomised trials. We have provided examples and explanations for most items. Where the item has not been modified or the modification is only minor, readers are referred to the original statements for full explanation and elaboration [Schulz 2010; Campbell 2012]. For some items, which have not been modified, an example or explanation has been provided where this item raises specific nuances under the SW-CRT. Given differences in terminology used to describe the SW-CRT and the significant number of modified items, the items in this statement have been written in such a way so as to replace the original CONSORT items; and therefore, should not be considered extensions to the original items.

Title and abstract

Item 1a Title
Standard CONSORT item: Identification as a randomised trial in the title.
CONSORT cluster extension: Identification as a cluster randomised trial in the title.
Extension for SW-CRTs: Identification as a stepped-wedge cluster randomised trial in the title.

Example: "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." [DAVE Trial]

Explanation: One reason for including the type of study design in the title is to facilitate accurate identification of relevant studies in systematic reviews. A wide variety of different terminology is currently used to describe the SW-CRT. These include the "multiple-period baseline design" and the "wait list design" (although not every multiple-period baseline design and wait list design will be a SW-CRT). Adoption of a single term will improve the identification of these studies and differentiate studies which are not SW-CRTs. Reporting of parallel cluster randomised trials (CRT) improved with the adoption of the single term "cluster" rather than the mix of terms (such as "group randomised" or "field trial") [Ivers 2011]. It can also be useful to report any trial acronym in the title, to aid future searches for the study.

Item 1b: Abstract
Standard CONSORT item: Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts).

CONSORT cluster extension: Abstract See Table (not shown).
Extension for SW-CRTs: Structured summary of trial design, methods, results, and conclusions (Table 5).

For the same rationale as provided in the other CONSORT statements, clear reporting of the trial's objectives, design, methods, main results and conclusions in the abstract is crucial. The primary reason for this is that many readers will base their assessment of the trial from the information available in the abstract [Hopewell 2008]. A review assessing the quality of reporting of abstracts from fully published SW-CRT revealed incomplete reporting of important details [Wang 2017]. A set of items to be reported as a minimum in an abstract of a SW-CRT is included in Table 5. Of some note, the Items recommended to be reported in the abstract results section do not include the summary measures of the outcome under intervention and control conditions, so as to avoid misattributing the unadjusted difference to the treatment effect. A worked example of an abstract according to this template is provided (Table S1, Long-live Mothers Trial).

Introduction

Item 2a: Background
Standard CONSORT: Scientific background and explanation of rationale.
CONSORT cluster extension: Rationale for using a cluster design
Extension for SW-CRTs: Scientific background. Rationale for using a cluster design and rationale for using a stepped-wedge design.

Example 1 (Scientific background): "In 2008, the World Health Organization (WHO) introduced the Surgical Safety Checklist (SSC) designed to improve consistency of care. The pilot pre-/post evaluation of the WHO SSC across 8 countries worldwide, which found reduced morbidity and mortality after SSC implementation, constituted the first scientific evidence of the WHO SSC effects. A number of subsequent studies to date have reported improved patient outcomes with use of checklists. Furthermore, checklists have also been shown to improve communication, preparedness, teamwork, and safety attitudes—findings that have been corroborated by a recent systematic review. Although checklists are becoming a standard of care in surgery, the strength of the available evidence has been criticized as being low because of (i) predominantly pre /post implementation designs without controls; (ii) lack of evidence on effect on length of stay; and (iii) lack of evidence on any associated cost savings. Randomized controlled trials (RCTs) are required…." [Surgical Checklist Trial]

*Example 2 (Rationale for cluster randomisation and stepped-wedge design): "*A stepped wedge cluster randomised controlled design was chosen following piloting to facilitate roll out of the intervention, …, and prevent contamination and disappointment effects in hospitals not randomised to the intervention." [FIT Trial]

Explanation: The need for any randomised evaluation of an intervention, whether randomising clusters or individuals should be justified. This justification should make reference to the best available evidence for similar interventions. Reasons why current evidence is lacking should be articulated (as in Example 1).

As with any trial design, key aspects of the design should be justified. In the SW-CRT, this justification includes the use of cluster randomisation, the need to roll out the intervention to all clusters (where this is the case), and the need for staggered roll-out of the intervention [Hargreaves 2015]. Justifying cluster randomisation is important because cluster randomisation increases the sample size and this, in turn might expose more participants to interventions of unknown effectiveness. Justifying the need for a staggered roll-out of the intervention using a SW-CRT, as opposed to a simple parallel arm implementation, is important because the SW-CRT is more complicated in its design, analysis, and implementation than the parallel CRT. Risks of bias in the SW-CRT may be higher than in a parallel CRT. For example, secular trends may be of concern in a SW-CRT, but not in a parallel design [Hemming

2017]. Risks of bias arising from identification and recruitment of participants may also be higher because in a SW-CRT it may be more difficult to blind people recruiting participants to the cluster's allocation status. The design is consequently viewed by some as potentially providing a lower level of evidence compared to the parallel CRT [Mdege 2011; Kotz 2012; Haines 2017]).

Some possible justifications for adopting the stepped-wedge design include that the intervention will be rolled out regardless of the research study [Prost 2015], availability of an inadequate number of clusters to achieve the target power in a parallel design [Hemming 2016], to increase statistical efficiency [Lawrie 2015; Girling 2016; Zhan 2017], or to facilitate recruitment when engagement of clusters is only forthcoming on some promise of the intervention (as in Example 2).

Although staggering the roll-out may appeal to researchers with limited resources for delivering the intervention simultaneously, this is not in itself a legitimate argument for a SW-CRT [Hemming 2015b]. Providing the intervention to all clusters might also increase the duration of the study (due to the staggering of the roll-out) and will possibly increase the number of clusters (and patients) exposed to the intervention (due to all clusters receiving the intervention). For these reasons, justifying the need to expose all clusters (where this is the case) to the intervention is important. The cluster cross-over design is a more statistically efficient design than the SW-CRT and it might therefore be important to justify why a unidirectional cross-over design has been chosen. However, in practice the use of the cluster cross-over design is restricted to interventions that can be withdrawn from use, and this largely depends on the type of intervention being evaluated.

Item 2b: Objective
Standard CONSORT item: Specific objectives or hypotheses.
CONSORT cluster extension: Whether objectives pertain to the cluster level, the individual participant level or both.
Extension for SW-CRTs: Specific objectives or hypotheses.

Example: "We report a stepped wedge cluster RCT aimed to evaluate the impact of the WHO SSC *(World Health Organisation Surgical Safety Checklist)* on morbidity, mortality, and length of hospital stay (LOS). We hypothesized a reduction of 30 days' in-hospital morbidity and mortality and subsequent LOS post-Checklist implementation." [Surgical Checklist Trial]

Explanation: Having a clear and succinct set of objectives can help summarise the overarching aims of the study. Specification of the objectives gives clarity about the anticipated effects of the intervention being evaluated (as in Example). Sometimes these effects will be anticipated to be on process outcomes (e.g. systems changes, clinician performance), particularly in trials which target health care providers; other times the intervention might target patients and anticipate effects on clinical outcomes. One specific objective which can be of interest in a SW-CRT is to evaluate the effect of the intervention by timing of implementation (e.g. does the effect of the intervention change as the intervention is perhaps refined over time) or time since intervention implementation (e.g. does the intervention create a permanent effect). Also of relevance is whether the study is to show superiority of the intervention condition, non-inferiority or equivalence. For non-inferiority or equivalence authors should also ensure reporting according to the CONSORT extension for non-inferiority and equivalence studies [Piaggio 2012].

Methods: Trial design

Item 3a: Trial design
Standard CONSORT item: Description of trial design (such as parallel, factorial) including allocation ratio.
CONSORT cluster extension: Definition of cluster and description of how the design features apply to the clusters.

Extension for SW-CRTs: Description and diagram of trial design including definition of cluster, number of sequences, number of clusters randomised to each sequence, number of periods, duration of time between each step, and whether the participants assessed in different periods are the same people, different people, or a mixture.

Example 1: "During the DAVE study, the intervention will be rolled out sequentially to 128 rural villages (clusters) over four time periods. The evaluation will consist of data collection at five fixed time points (baseline and following each of the four intervention periods)… The intervention will be fully implemented by the end of the trial, with all 128 villages receiving the intervention: 22 first receiving the intervention at period 2, 36 at period 3, 35 at period 4, and 35 at period 5." [Dave Trial Protocol, Figure S1]

Example 2: This study will use a closed cohort stepped wedge cluster randomised design, which involves a sequential crossover of clusters from the control to the intervention arm, so that every cluster begins in the control condition and eventually receives the intervention, with the order of crossover randomly determined. The study will be conducted in four rural villages…At the start of the study period, baseline (T0) demographic and health data will be collected from each consenting household and baseline hygiene education will be provided. …The second (T1) health survey will start 4 weeks after the initiation of piped untreated river water supply to evaluate the impact of hygiene education combined with improved water quantity compared with baseline (T0). RBF-treated water (intervention arm) will then be sequentially introduced to each village in random order at 12-week intervals (T2–T5), with health surveys performed 4 weeks after the implementation of the intervention to assess the additional effects of improved water quality [Riverbank Filtration Trial, Figure 2]

Explanation: The specific details of the design of the SW-CRT have implications for the type of analysis and sample size calculations required.

Information on the number of sequences and the number of clusters randomised to each sequence is the core of the study design and so should be reported. The number of time periods will often (but not always) be one more than the number of steps (as in Example 1). Definition of cluster (as clearly reported in Example 1) and duration of time periods are also crucial. The duration of the first and last periods can sometimes differ from other periods; if so, this should be reported. The number of clusters allocated to each sequence may vary and, if so, this should be reported.

Information on whether the measurements taken in the different time periods are from the same individuals or different individuals is important for both sample size and analysis. In an open cohort design, participants are repeatedly assessed over series of measurement points and participants can join and leave the cohort; in a closed cohort design, new participants cannot join the study; in a cross-sectional design, different participants are assessed at each measurement occasion. Measurements can also take place at one point in time in each period, or can be continuous throughout the period. This issue is covered in more detail under Item 6a (assessments of outcomes).

A diagram of the trial design can efficiently communicate the details. Key points to depict in the design diagram are the timing of the interventions (Item 3a) and the timing of the data collection (Item 6a). In the Riverbank Filtration Trial, key information about the design was reported in a diagram (Figure 2) and the main text (Example 2).

Item 3b: Changes to trial design
Standard CONSORT item: Important changes to methods after trial commencement (such as eligibility criteria), with reasons.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Important changes to methods after trial commencement (such as eligibility criteria), with reasons.

Example: "…delayed Research and Development registration shortened the baseline pre-randomisation phase from twelve months to nine in the first hospitals randomised to the intervention."[FIT Trial]

Explanation: Changes to key features of the design can have important implications for the interpretation of results. Some changes or deviations may be inevitable. Potential changes in the SW-CRT include modification to the duration between steps (perhaps because of study set up delays as in Example). The timing of any changes is important as they may affect some observations / clusters and not others.

Methods: Participants

Item 4a: Participants
Standard CONSORT item: Eligibility criteria for participants.
CONSORT cluster extension: Eligibility criteria for clusters.
Extension for SW-CRTs: Eligibility criteria for clusters and participants.

Example: "Inclusion criteria: Institution level: At least two units of one *(from each)* nursing home must participate in the study, from which at least 30 residents with dementia can be recruited. The care of the residents must predominantly take place in the respective unit. Resident level: Criteria for inclusion are informed consent obtained from people with dementia or their legal representative; diagnosis of dementia based on the medical diagnosis in the charts and a FAST score > 1); residence for at least 14 days in the unit. Staff level: All of the nursing staff working in one of the two participating wards of the nursing home must provide their informed consent." [FallDem Trial]

Explanation: The SW-CRT is a type of cluster randomised trial and as such, has inclusion and exclusion criteria for both clusters and participants. Furthermore there may be multiple levels of participants. For example, clusters may be general practices that include cluster-level participants (e.g. general practitioners) and individual-level participants (e.g. patients). So, in some trials, there may be multiple levels at which inclusion and exclusion criteria apply (as in the Example). Reporting of eligibility criteria is important so that readers can infer how typical or atypical the clusters and participants are of the population at large [Zwarenstein 2008].

Item 4b: Setting
Standard CONSORT item: Settings and locations where the data were collected.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Settings and locations where the data were collected.

Readers are referred to the CONSORT statement and its extension to CRTs for examples and explanation [Schulz 2010, Campbell 2012].

Methods: Intervention

Item 5: Intervention
Standard CONSORT item: The interventions for each group with sufficient details to allow replication, including how and when they were actually administered.
CONSORT cluster extension: Whether interventions pertain to the cluster level, the individual participant level or both.
Extension for SW-CRTs: The intervention and control conditions with sufficient details to allow replication, including whether the intervention was maintained or repeated, and whether it was delivered at the level of the cluster, the individual, or both.

Example 1 (Description of the intervention condition): "The intervention involves three key modes of delivery: verbally via reception staff, in paper form with a pamphlet, and electronically via a secure, internet-enabled tablet (see Table *(not provided)* for overview of intervention). First, reception staff will verify the organ donor registration status of patients upon their arrival at the clinic on the provincial health card that patients must provide to receive healthcare services from their family physician. As reception staff already request a patient's health card during their visit, this step is designed to fit within existing work routines rather than increasing any workload. Reception staff will provide patients that have not yet registered with an educational pamphlet including a photo and signature of the physicians in the office and office logos and include messages that directly address identified barriers to donor registration. Second, internet-enabled tablets will be provided in each waiting room to give patients the immediate opportunity to register for organ donation online via a secure provincial website. The location of the materials will be tailored according to the family physician office's preferences." *(further details provided in paper)* [RegisterNow-1 Trial]

Example 2 (Description of control condition): "If the participant's medical centre is in the control phase, they will receive usual care. In Australia, usual care would mean the patient would consult their GP as per normal standards for that practice for a patient discharged from hospital. There will be no pharmacist in the medical centre during the control phase. Medication liaison in the form of a discharge medication record may be provided to patients on discharge from hospital and may be included in the hospital discharge summary to the GP." [REMAIN Trial Protocol]

Example 3 (Unit of delivery is individual): "The intervention comprised a therapeutic dose of AQ (10 mg/kg/day for 3 days) combined with one dose of SP on the first day (25mg sulfamethoxypirazyne and 1.25mg pyrimethamine per kg in 2008, 25mg sulfadoxine, 1.25mg pyrimethamine in 2009–10) administered once per month for the last three months of the malaria transmission season (September-November)." [SMC Trial]

Example 4 (Continuously delivered intervention): "It *(the intervention)* comprised bedside placement of alcohol hand-rub, posters and patient empowerment materials encouraging healthcare workers to clean their hands, plus audit and feedback of hand-hygiene compliance at least once every 6 months." [FIT Trial]

Explanation: Clear reporting of the intervention is essential to allow replication and implementation of successful interventions (Example 1). For interventions demonstrated to have little evidence of benefit, reporting of sufficient detail of the intervention helps to avoid evaluating the same intervention again or to identify what aspects of the intervention could be modified. This is especially important for complex interventions – a common type of intervention evaluated in SW-CRTs. We recommend reporting details of the intervention as per the TiDierR guideline [Hoffmann 2014]. As per the original CONSORT statement, it is important to describe all treatment conditions being compared. In SW-CRTs the comparator is often "usual care" which should be described in sufficient detail (Example 2). The control condition should be described in a similar level of detail to the intervention condition [Zwarenstein 2008].

Information on whether the intervention is delivered at the level of the cluster or individual (or perhaps both) is important as it allows identification of whether individuals can avoid the intervention. For example, an intervention which is delivered at the level of the cluster will often mean that it is delivered to all individuals within that cluster (Example 1). In the SMC Trial the intervention was delivered directly to the individual (Example 3). This information is also important as it can inform the degree of penetration of the intervention and it can also be helpful in eliciting what consent procedures should be in place (Items 10c and 26).

In a SW-CRT it is important to be clear about whether the intervention is expected to create an effect that is expected to be immediate (or delayed); and whether the anticipated effects of the intervention are expected to be sustained. This is important because the observations contributing to the analysis will consist of a mixture of observations collected immediately after roll-out of the intervention; and observations collected some time post roll-out.

The effect of any intervention can be delayed; for example, due to a learning effect, one may need to allow for a delay before the effect is fully realised (this might be the case in Example 4). In these situations a transition period might be incorporated into the design. Furthermore the anticipated effects of the intervention might be sustained (in which case an intervention might be designed to have a one-off delivery, as in Example 1) or expected to decay (in which case an intervention might be designed to have repeated delivery, as in Example 4). In some SW-CRTs the exact form of the intervention may evolve over time; reporting this information allows assessment of the level of standardisation of the intervention across the clusters [Zwarenstein 2008].

In Example 1 the intervention being evaluated is formed of several components. Depending on the exact nature of the intervention, there may be a delay before any anticipated effect is realised. The effects of some components may also wane through familiarity. Furthermore some components of an intervention might be continuously delivered (i.e. provision of pamphlets) whereas some components might be delivered just once (i.e. educational components). In Example 4 the educational component of the intervention is re-enforced and so its anticipated effect is less likely to decay.

Methods: Outcomes

Item 6a: Outcomes
Standard CONSORT item: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.
CONSORT cluster extension: Whether outcome measures pertain to the cluster level, the individual participant level or both.
Extension for SW-CRTs: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.

Example 1 (Pre-specified outcomes): "The primary outcome of the study is a 7-day period prevalence of diarrhoea among villagers of all ages. Secondary outcomes include a 7-day period prevalence of other hygiene-related illnesses (respiratory and skin infections), reported changes in hygiene practices, household water usage and water supply preference." [Riverbank Filtration Trial]

Example 2 (Cross-sectional sampling): "Data collection for the evaluation took the form of a postal survey conducted at five fixed time points: baseline (in the month prior to commencement of the first intervention period) and within a week of the end of each of the four intervention periods. A repeated cross-sectional design was employed, in which a random sample of households within each cluster was selected to receive the survey at each period." [DAVE Trial]

Example 3 (Cohort design): "All household members will be eligible for inclusion in the study, regardless of age. …Each household will have the option to participate in up to five subsequent surveys…Outcomes will be measured at each of the six survey visits." [Riverbank Filtration Trial]

Example 4 (Transition period): "A 1-month transition phase is included where the medical centre is not considered as being in control or intervention and does not contribute to analysis. This transition period allows for the time it takes to embed the intervention into a medical centre." [REMAIN Trial]

Example 5 (Time to assessment and source of data): "Participants will be followed up to 12 months from day of hospital discharge. This will be done through collection of routine data from the hospital and medical centre. Demographics and reason for admission at enrolment and subsequent admissions in the 12-month follow-up will be collected through participant hospital records…Medical centre records will be used to identify whether a discharge treatment plan was received and the timeliness and number of GP visits during the 12-month follow-up period for each participant."

Explanation: All outcomes should be completely defined. This should include the pre-specified primary outcome and all secondary outcome measures (Example 1). It is also important to report clearly how and when these measurements were obtained.

SW-CRTs make a series of measurements over time within each cluster. These measurements could be on different participants in each period (i.e. cross-sectional design) as in Example 2; the same participants (i.e. cohort design) as in Example 3; or a mixture, and this will inform the method of analysis and has implications for sample size calculations. Data are rarely collected at the level of the cluster, but knowledge of whether outcomes in each period are at the cluster level (either because of true cluster level outcomes or because of the availability of aggregated data only) or individual level has implications for the method of analysis.

It should be reported whether outcomes are collected at discrete points in time common to all participants (e.g. a survey implemented at several discrete points in time as in Example 3), or at time points specific to each participant (e.g. as they leave hospital as in Example 5). The timing of measurements has implications for the choice of analysis. For example, if the outcomes are collected at discrete time points (as in Example 3), then time effects can be included as categorical effects; whereas if the outcomes are collected continuously (for example as would be the case in a SW-CRT where the outcome was routinely collected mortality data), then time effects could potentially be modelled using parametric or semi-parametric forms.

The reporting of the timing of data collection should also note whether there were periods in which outcomes were not ascertained, for example transition periods immediately after the intervention was rolled out, to allow time for the intervention to realise its full impact (as in Example 4).

In individually and cluster randomised parallel trials outcomes are often assessed at multiple time points (for example 6 and 12 months post randomisation) and it is important to pre-specify the primary follow-up time of interest. This might also be the case in SW-CRTs. Sometimes the outcome assessments will extend beyond the actual study dates. For example, a trial might roll-out the intervention to clusters over a four year period and the primary follow-up time might be 30 years later [Shimakawa 2014]. Clear reporting on the timing of follow-up assessments (as in Example 5) also allows assessment of whether all observations collected under the intervention condition were fully exposed to the intervention, and whether any observations collected under the control condition might have been contaminated by the intervention.

Reporting whether data were collected from routine sources or purposively collected can help ascertain the risk of bias (e.g. from measurement of the outcome) and identify who are the human research participants (see Item 26). SW-CRTs are often implemented in real-world settings and, as such, may rely on routinely collected outcome data (Example 5). Reporting of whether the data collection procedures changed over time is important given the imbalance over time with respect to intervention conditions [Shadish 2002]. It is also important to report any measures which can allow assessment of the reliability and validity of routinely collected data.

Item 6b: Changes to outcomes
Standard CONSORT item: Any changes to trial outcomes after the trial commenced, with reasons.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Any changes to trial outcomes after the trial commenced, with reasons.

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Methods: Sample size

Item 7a: Sample size
Standard CONSORT item: How sample size was determined.
CONSORT cluster extension: Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intra-cluster correlation (ICC or k), and an indication of its uncertainty.
Extension for SW-CRTs: How sample size was determined. Method of calculation and relevant parameters with sufficient detail so the calculation can be replicated (Table 6). Assumptions made about correlations between outcomes of participants from the same cluster.

Example 1 (Sample size): "We would consider an absolute increase of 10**%** in the proportion of patients who are registered organ donors at 7 days post-encounter to be both clinically important and feasible. Our sample size of 6 clusters (10,500 patients in total) achieves 80% power to detect this difference assuming a control proportion of 0.5 using a two-sided test at the 5% level of significance [Hooper 2016]. Our calculation assumes an intra cluster correlation coefficient of 0.06, as calculated from our previous work (19), an average of 250 patient encounters per site in each two-week interval, and a cluster autocorrelation coefficient of 0.8 to allow for a 20% decay in the strength of the correlation in repeated measures over time.(20) The percentage of registered donors in the control condition is conservatively assumed to be 50% to allow for a higher prevalence of registered donors in our participating offices than the provincial average. No adjustment is made for cluster attrition as the risk of attrition is low, and all outcomes will be assessed from routinely collected sources, regardless of any drop-out. Given some uncertainty around parameter estimates required for the stepped wedge sample size calculation, sensitivity of our detectable effect size to a range of alternative assumptions is presented in Table *(not shown)*. The results show that across a range of control arm proportions (from 0.4 to 0.5), average cluster sizes (from 100 to 400), and cluster autocorrelation coefficients (from 0.8 to 0.95), our sample size of 6 practices will achieve 80% power to detect absolute increases between 5% and 11%." [RegisterNow-1 Trial]

Example 2 (Sample size fixed by design): "The study had a fixed sample size by design that could not be modified, so the power calculations did not inform any sample size targets." [Targeted Case Finding Trial]

Explanation:

The method of calculation and all relevant parameters, used in the sample size calculation should be given. Most of the key items to report are listed in Table 6. These have been divided into key items which are essential and likely of relevance to all SW-CRTs; and those which might be considered additional or supplementary information which will only be of relevance to some SW-CRTs. Besides the usual effect size, significance level and power, these may include: the cluster size and whether account of unequal cluster sizes has been made, avoiding any ambiguity between cluster size per measurement period and total cluster size; a within-period intra-cluster correlation (ICC) and assumptions about correlations between outcomes of different participants from the same cluster in different periods (or other assumptions which appropriately reflect the complexity of the design); allowance for repeated

measurement taken from the same participants, with sufficient detail to allow the calculation to be replicated. Often a sensitivity analysis, looking at the effect of relaxing some of the assumptions, may be warranted.

Specifying the method of sample size calculation [Hussey 2016; Hooper 2016], or providing access to sample size calculation code [Baio 2015; Hooper 2016; Hemming 2016] or programmed sample size function [Hemming 2014] can aid replication of the sample size (Example 1 reported they used the Hooper method). Detailed reporting of the sample size method will allow assessment of whether the method has allowed for all features inherent to the particular design (e.g. transition periods, repeated measures on the same participants). Reporting of the sample size calculation will likely include: number of clusters and whether equal or unequal cluster sizes are assumed, cluster size or cluster size per period, number of sequences, and number of clusters per sequence. Reporting of these basic sample size elements is poor in SW-CRTs [Martin 2016]; as is the reporting of basic elements in parallel CRTs [Rutterford 2015].

For clarity it is important to distinguish between total cluster size (across all periods) and cluster sizes per period (Example 1). In a design which repeatedly measures the same participants it would be natural to provide the number of participants in each cluster and the number of repeated measurements per participant; in a design which involves taking repeated, discrete samples with different participants each time it would be natural to provide the number of participants in each cluster in each of these periods; whereas in a design where newly eligible individuals are recruited continuously it might be more appropriate to report the total number of participants expected in each cluster over the duration of recruitment.

In a parallel CRT it is important to report the ICC (the correlation between outcomes of two individuals from the same cluster). The coefficient of variation of cluster rates, proportions or means has been suggested as an alternative parameter in sample size formulae for CRTs [Hayes 1999]. Correlation structures are more complicated in a SW-CRT and there may not be a single ICC, as the strength of correlation might depend additionally on the separation in time [Hooper 2015; Martin 2016b; Kasza 2017]. Such correlation structures could be formalised in a variety of ways, for example using a within-period ICC and a between-period ICC or cluster auto-correlation coefficient (as in Example 1) [Kasza 2017]. In SW-CRTs where the same individuals are assessed repeatedly it may also be important to consider correlations over time within individuals [Hooper 2016].

An indication of the sensitivity of the sample size or power to the assumed parameter values could be provided, for example, by reporting sample size or power at a variety of alternative correlation values. Rationale for the assumed parameter values should be provided (as in Example 1).

In randomised trials the sample size (and so consequently the number of clusters) is often based on the number needed to detect the target difference at a desired level of power and significance [Cook 2017]. SW-CRTs can sometimes have their sample size fixed by the number of clusters, participants, or both, available in a natural setting. Whether the sample size was fixed by factors outside of the control of the experimenters or based on the target difference (as conventionally is the case in a randomised controlled trial) should be reported (as in Example 2). When the sample size is fixed, it can be useful to report what effect size the study was powered to detect. If no power calculation was performed, this should be reported. Retrospective power calculations based on the results of the trial are of little merit [Hoenig 2001; Sculz 2010].

Item 7b: Interim analyses
Standard CONSORT item: When applicable, explanation of any interim analyses and stopping guidelines.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: When applicable, explanation of any interim analyses and stopping guidelines.

Explanation: Interim analyses of outcomes can be used to assess harm, futility, and efficacy. Interim analyses can also be used to monitor recruitment and retention rates, and monitor balance across control and intervention conditions (where trial processes suggest that there may be a risk of differential recruitment or consent).

The relevance of interim analyses of outcomes might be questionable in some SW-CRTs, so careful reporting of motivation is important. For example, if the intervention is being rolled out to all clusters within the fastest time frame possible, then stopping the trial early after demonstrating efficacy does not necessarily mean the intervention can be rolled out to the remaining clusters immediately. In some settings, SW-CRTs evaluate interventions for which safety concerns are likely to be minimal (although this will not always be the case). It might be of interest to consider stopping a SW-CRT for futility, although if there are minimal safety concerns then stopping the trial early for futility may also not be worthwhile. However, other important reasons for considering stopping a trial include that the trial itself is not successful, perhaps because clusters are failing to adhere to the randomisation schedule, because data for outcomes are not forthcoming, or because procedural requirements have delayed the start dates for many clusters [Kristunas 2017]. Dates or times at which any interim analysis will be carried out should be reported together with objectives of such interim analyses.

Of note, in a SW-CRT due to the imbalanced nature of the design, interim analyses for outcomes carried out early in the trial will have a large imbalance between numbers of observations exposed to control and intervention conditions. This imbalance is likely to have power implications [Grayling 2017]; and will make a blinded interim analysis infeasible. The clustered nature of the data will also have implications on power and interim analyses [Zou 2005]. Proposed methods of interim analysis should be outlined. Interim analyses of outcomes might or might not follow the same method of analysis planned for the main results. As with any trial, incorporation of any interim analyses of outcomes (where a decision is to be made about continuation of the trial) should be allowed for in power calculations to control for the over-all Type I error rate.

Methods: Randomisation

Item 8a: Sequence generation
Standard CONSORT item: Method used to generate the random allocation sequence.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Method used to generate the random allocation to the sequences of treatments.

Example: "Eligible schools were randomly assigned to one of the four sequences (3 or 4 schools per sequence) for time of crossover from control to intervention using a computer-generated list of random numbers." [SBP Trial]

Explanation: Random allocation in SW-CRTs takes a different form to that in parallel arm designs. Rather than each cluster being randomly allocated to one of two treatments, allocation is to one of several sequences which define the order with which clusters cross from the control condition to the intervention condition (Example). The term "sequence generation" in a SW-CRT therefore has a slightly different meaning to that of individually randomised trials. In an individually randomised trial "sequence" refers to a sequence of treatments to allocate all participants to either the intervention or control condition.

Furthermore, rather than the randomisation being performed as clusters or individuals present to the trial the randomisation in a SW-CRT is usually done at a single point in time before the trial starts.

Item 8b: Randomisation method
Standard CONSORT: Type of randomisation; details of any restriction (such as blocking and block size).
CONSORT cluster extension: Details of stratification or matching if used
Extension for SW-CRTs: Type of randomisation; details of any constrained randomisation or stratification if used.

Example 1 (Unrestricted): "Nursing-home units were the unit of randomisation... RL (not involved in recruitment) randomly allocated units to one of five groups with computer-generated random numbers…" [Depression Management Trial]

*Example 2 (Stratification): "*All schools are assigned a decile rating, which indicates the extent to which the school draws its students from a range of socioeconomic areas. Decile 1 schools are the 10% of schools with the highest proportion of students from low socioeconomic resource areas (defined according to residents' income, occupation, household crowding, educational qualifications and income support) and decile 10 are the 10% of schools with the highest proportion of students from high socioeconomic areas…. The order of switch-over is determined randomly for each group (*decile*) of clusters" [SBP Trial Protocol]

Example 3 (Covariate constrained randomisation): "The randomization was conducted using a highly restricted randomization design. With this limited number of randomization units, selection of one sequence from the 5.4 $*10^{26}$ completely at random would run the risk of obtaining a sequence that is substantially unbalanced with respect to one or more potentially important covariates. Randomization was done using a highly restricted randomization design to achieve close balance with respect to clinic-level covariates including mean CD4 count, clinic size, average education, tuberculosis treatment levels, existence of a supervised tuberculosis therapy (DOTS) program and geography (reference cited to detailed methods)". [THRio Trial Protocol]

Explanation: In a SW-CRT, rather than the randomisations being done sequentially (as the patient or cluster presents to the trial), the randomisation is usually done at a single point in time before the trial starts. This means that different methods for controlling balance of cluster-level factors can be considered along with methods used in individually randomised trials such as stratification [Ivers 2012]. How the randomisation is restricted is known to have implications for analysis.

There are two common ways in which clusters may be allocated in a SW-CRT. One is simple unrestricted allocation to one of several possible sequences (Example 1); another is stratified allocation with clusters divided into distinct strata prior to random allocation within each stratum (Example 2). For a stratified design the sequences are generated independently within each stratum. This essentially means that separate mini SW-CRTs are conducted in each stratum (Example 2). Yet another method of allocation is covariate constrained allocation which balances key covariate values (such as cluster size) between intervention and control conditions (Example 3) [Moulton 2007].

Item 9: Allocation concealment
Standard CONSORT item: Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned.
CONSORT cluster extension: Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both.
Extension for SW-CRTs: Specification that allocation was based on clusters; description of any methods used to conceal the allocation from the clusters until after recruitment.

*Example 1 (Concealment from cluster)***:** "Once 14 medical centres have provided consent to be involved in the study, each enrolled medical centre will be randomised to a transition step." [REMAIN Trial]

*Example 2 (Concealment of cross-over date)***:**"The allocation sequence will only be made available to two study investigators (ABF and MS). Indian study investigators will be blinded to the allocation sequence with only the next village randomised for rollout being revealed at each intervention implementation time point. Study

participants will be blinded to the allocation sequence and those not yet receiving the intervention will not be aware of the time at which they will have the intervention implemented." [Riverbank Filtration Trial]

Explanation: In a SW-CRT clusters are allocated to a sequence of treatments, so clusters will spend time in the control condition until a particular date when they cross to the intervention condition. This is unlike a parallel arm cluster randomised trial in which clusters are allocated to treatment conditions. Randomisation of all clusters (to sequences) in a SW-CRT will often occur at a single point in time (as in Example 1). Randomisation could in theory also be performed at step-times, where one or more of the remaining clusters will be randomly selected to cross over just prior to the cross-over date (no examples of this have been identified).

It is important to report any method that was used to conceal the allocation from clusters and from those individuals responsible for recruiting clusters, until after recruitment. Reporting of this information allows assessment of the potential for selection bias [Higgins 2016]. One common way of preserving allocation concealment is to perform the randomisation after recruitment of all clusters (as in Example 1).

When randomisation of the clusters occurs at a single point, the cross-over date may be revealed immediately to each cluster, or revealed sequentially to the clusters as they approach the time of cross-over (as in Example 2). Reporting when clusters were told of their cross-over date allows assessment of potential biases. For example, when clusters are informed of their date of cross-over at the beginning of the trial, some clusters (e.g., those randomized to cross over later) may drop-out, leading to differential attrition; yet at the same time a public randomisation at the start of the trial may also prevent subversion of the randomisation process [Higgins 2016]. Knowledge of when a cluster is crossing over could lead to other biases, for example, if individuals within a cluster are aware of the impending cross-over, they may defer enrolling participants into the trial to ensure they receive the intervention.

Full transparency of reporting of the blinding throughout the trial, including the randomisation process, is best reported using a timeline diagram [Caille 2016].

Methods: Implementation of randomisation

As with a parallel CRT, it is important that all steps in the implementation of the randomisation process are clearly described. It is important that this information on the allocation and recruitment process is described for both clusters and participants. Information on the allocation and enrolment of the clusters is described in Item 10a and corresponding information for participants in Item 10b. Enrolment of participants is closely linked to the consent process (for example, differential consent processes can have implications for selective recruitment). Therefore, following the cluster CONSORT extension, Item 10c describes the consent processes.

Of note, we use the term "selection bias" to refer to any process by which there is differential inclusion of participants in the treatment conditions being compared. Sometimes selection bias is used to refer only to differential *inclusion* of clusters by intervention conditions. More specifically, "identification bias" refers to biases which are induced by differential application of the inclusion / exclusion criteria [Higgins 2016]. The term "recruitment bias" refers to biases which are induced by differential recruitment into the trial by the health care practitioner or to biases induced by individuals differentially declining to participate.

Item 10a: Inclusion of clusters
Standard CONSORT item: Not included in original CONSORT statement.
CONSORT cluster extension: Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions.
Extension for SW-CRTs: Who generated the randomisation schedule, who enrolled clusters, and who assigned clusters to sequences.

*Example: "*We will recruit a convenience sample of practices from within our network of family physician office contacts within the London, Ontario and Stratford, Ontario communities. A collaborating family physician will send an introductory email to potential family physician contacts, inviting them and their practice to consider participating. We will then arrange an in-person meeting with family physicians from interested sites to introduce our study and obtain written agreement from family physicians and offices agreeing to participate that meet our eligibility criteria. A statistician blinded to cluster identity and not involved in the intervention delivery will generate the allocation sequence using computer-generated random numbers." [RegisterNow-1 Trial]

Explanation: Knowledge of who implemented the randomisation procedures at the level of the cluster is required for ascertaining if selection biases are possible.

It is important to have a separation of roles between those who generate the randomisation schedule and those who recruit, enrol and assign clusters to the sequence (as in the Example). If the person who generated the randomisation was also responsible for recruiting the clusters, this could mean that there was an increased risk of selection bias. This is best achieved by having a person independent of the trial doing the randomisation. This will be less important in trials where the randomisation takes place after recruitment of all clusters.

Item 10b: Inclusion of participants
Standard CONSORT item: Not included in original CONSORT statement.
CONSORT cluster extension: Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling).
Extension for SW-CRTs: Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration or random sampling; continuous recruitment or ascertainment, or recruitment at a fixed point in time), including who recruited or identified participants.

Example 1 (Complete enumeration with continuous ascertainment): "The study included all patients admitted to 16 acute adult wards of one general hospital over a 32-week period." [Critical Care Outreach Trial]

Example 2 (Random sampling): "Data collection for the evaluation study will focus on adults aged 18 years and over. The study will use a repeated cross-sectional design, in which a random sample of people within each cluster will be surveyed at each stage. A complete list of all households in each of the 128 study villages will be obtained using the Postcode... The order in which households are approached to participate in the survey at each stage will be randomly generated...One adult per household will be randomly selected." [DAVE Trial Protocol]

*Example 3 (Continuous recruitment): "*Then, the leaders of the nursing homes are responsible for the recruitment of the units and the residents according to the inclusion and exclusion criteria of the study. Here, all eligible participants of the participating units are invited to participate. Before the recruitment procedure will commence, each leader of the nursing homes will attend a kick-off meeting held by a senior investigator about the inclusion and exclusion criteria and the planned recruitment strategy. For the participants who drop out of the trial, we are planning to monitor the reasons (for example, death or moving) and perform a sensitivity analysis at the end of the trial to determine whether they differ according to certain characteristics (for example, the prevalence of the challenging behavior or gender). Residents who are newly admitted to clusters during follow up will also be included in the study …" [FallDem Trial]

Explanation: Individual participants can be included in a SW-CRT in many different ways. Sometimes, participants are not recruited into a trial, but rather their data are used from routinely collected sources (Example 1). In this case it is common to take a complete enumeration of the cluster or at least those meeting the eligibility criteria. Alternatively, a sample of individuals from the cluster might be asked to complete data assessments or questionnaires in each

period (Example 2). Alternatively, participants might be recruited to participate in the trial. This recruitment might take place continuously (Example 3) or at a fixed point in time before the start of the trial.

Knowledge of how participants are included in the trial can help assess the likelihood of identification and recruitment bias. Trials with complete enumeration are less likely to suffer from these biases (Example 2). Where participants are identified or recruited after randomisation (as in Examples 1 and 3), either a complete enumeration of the cluster or recruitment/identification by someone who is blind to allocation can help mitigate recruitment and identification biases. Therefore, clear reporting of who recruited or identified participants and whether or not such individuals were blind to allocation is important so readers can determine the risks for bias. Identification and recruitment biases will not occur in designs in which participants are recruited prior to randomisation.

Item 10c: Consent
Standard CONSORT item: Not included in original CONSORT statement.
CONSORT cluster extension: From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation.
Extension for SW-CRTs: Whether, from whom and when consent was sought and for what; whether this differed between treatment conditions.

Example 1 (Individual-level consent): "Written informed assent was obtained from all participating children as well as parental consent. Only children who provided both assent and parental consent were eligible to take part." [SBP Trial]
Example 2 (Cluster and individual-level consent): "Criteria for inclusion are informed consent obtained from people with dementia or their legal representative.…All of the nursing staff working in one of the two participating wards of the nursing home must provide their informed consent" [FallDem Trial]

Explanation: Obtaining informed consent for participation, study interventions, and data collection procedures in clinical trials is an integral principle of research ethics and international human rights law [IEHR 2016; UN 1966]. The process by which consent was obtained can lead to biases [Campbell 2012]. It is important to describe what consent was for (e.g. exposure to the intervention or use of data), whether consent was sought before or after randomisation, and whether the type of consent differed between intervention and control conditions.

In SW-CRTs there can be cluster-level research participants (e.g., health-care practitioners) and individual-level research participants (e.g. patients) [Taljaard 2013]. It is therefore important to identify explicitly from whom consent was obtained in the study (Example 2) or to state that consent was not obtained. Furthermore, in most cluster trials someone provides access to the cluster; such individuals are often called "gatekeepers" or "cluster guardians" [Edwards 1999]. Gatekeeper permission for trial participation is different to consent from cluster-level research participants, such as health providers, for their own participation in the study.

In cluster randomised trials in which the treatment is delivered at the level of the cluster, it may not be possible to obtain consent for exposure to the intervention or control condition as the intervention may be impossible to avoid (as would be the case in Example 1 under Item 10b); however, consent can still be taken for use of data (implied by return of questionnaire data in Example 2 under Item 10b). It is therefore important to clearly report what consent was for. If participants recruited to the control and intervention conditions are given different information when their consent is taken, this can lead to bias [Eldridge 2005]. The information provided about the objectives of the study can itself prompt participants to act differently. For example, participants enrolled in a study of an intervention to increase uptake of HIV screening, who are fully informed about the objectives of the study, might increase uptake of screening irrespective of allocation to the intervention condition. This is known as the Hawthorne effect [McCarney 2007]. Reporting what information was provided to participants can allow readers to judge the risks of

such biases. A recent systematic review found that of the small number of SW-CRTs that reported whether or not consent was obtained, only a small proportion reported explicitly what this consent was for, and none reported when the consent was taken [Taljaard 2017].

Sometimes a research ethics committee might deem it appropriate that the study proceed without the informed consent of research participants (i.e. a waiver of consent) or the research ethics committee may otherwise modify informed consent requirements (i.e. modification of consent). When a waiver or modification of consent has been granted by a research ethics committee, it should be reported and a justification given. It should be clear whose consent was waived and whether the waiver pertains to study participation, data collection, or both. Not all jurisdictions allow for a waiver or modification of consent. Information on data collection procedures in the trial, e.g., whether data are anonymous or pseudo-anonymous, and whether they were routinely collected, can provide clarity around ethical aspects of the trial. When appropriate it can be useful to include any participant consent forms in appendices, which will allow readers to infer precisely the information provided to participants.

Methods: Blinding

Item 11a: Blinding
Standard CONSORT item: If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: If done, who was blinded after assignment to sequences (for example, cluster level participants, individual level participants, those assessing outcomes) and how.

Example 1 (Blinding not possible): "Blinding to the intervention (i.e., the type of water being received) is not possible due to potential differences in turbidity of untreated and RBF *(Riverbank Filtration)*-treated river water." [Riverbank Filtration Trial]

Example 2 (Blinding partially possible): "Residents did not know when the intervention was being implemented or what the programme elements were. Interviewers who administered the outcome questionnaires were masked to intervention implementation or depression treatment, and to previous test results. Data analysts were masked to whether a specific resident had been exposed to the intervention and to when the intervention was implemented in a unit, but were not masked during post-hoc analyses." [Depression Management Trial]

Explanation: SW-CRTs are often used to evaluate interventions for which it is impossible to blind participants or clusters to whether they are in the intervention or control condition, but nonetheless it is important to report clearly whether or not blinding was used and if so, who exactly was blinded to aspects of the trial (Example 1).

Often outcomes are collected at multiple levels (e.g. hospitals (e.g. team climate outcomes), clinicians (e.g. knowledge, skills, practice outcomes), patients (e.g. pain)). The possibility of blinding may be different depending on the level of participants (e.g. clinicians or patients) and may depend on the type of consent required (Item 10c). The degree of blinding should be reported at each level of the trial (e.g. clusters, participants as in Example 2) and whether the blinding differed in control and intervention conditions. Researchers should also specifically report blinding with respect to all outcomes. Blinding of those assessing outcomes should be clearly reported.

A systematic review has found that most SW-CRTs do not report clearly who was blinded and what people were blinded to [Taljaard 2017]. Whether or not and who was blinded, and when, is best reported by the use of a timeline diagram [Caille 2016].

Item 11b: Blinding

Standard CONSORT item: If relevant, description of the similarity of interventions.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: If relevant, description of the similarity of treatments.

Explanation: In trials with a placebo it is important to provide evidence of the similarity of the control condition to the intervention condition (i.e. to provide evidence of the blinding). However, In SW-CRTs it would be unusual to have a placebo and often participants are not blind to their allocation status. Sometimes, a minimal level of intervention is provided in the control condition in an attempt to keep participants blinded to their status as intervention or control participants. When appropriate such minimal level interventions should be described in full.

Methods: Statistical methods

Item 12a: Statistical methods
Standard CONSORT item: Statistical methods used to compare groups for primary and secondary outcomes.
CONSORT cluster extension: How clustering was taken into account.
Extension for SW-CRTs: Statistical methods used to compare treatment conditions for primary and secondary outcomes including how time effects, clustering and repeated measures were taken into account.

Example 1 (Allowance for clustering and secular trends): "A generalised linear mixed model was used for categorical outcomes, and a linear mixed model was used for continuous outcomes, adjusting for age, gender, ethnicity and school terms (i.e., secular trend). The cluster effect by school and correlation between repeated measurements on the same child over time were taken into account in the multilevel analysis." [SBP Trial]

Example 2 (Cluster level analysis): The primary outcome (diarrhoeal prevalence) will be calculated for each cell in the stepped wedge design by aggregating over all individuals surveyed in each village during each time period. Estimation of intervention effects will be obtained from a linear regression of the logarithm of the village-aggregated prevalence adjusting for seasonal effects and incorporating village as a fixed effect. The intervention effect coefficient will be exponentiated to produce an estimated relative reduction (with 95% CIs) in the overall prevalence of diarrhoea in the intervention periods (post-RBF) compared with control periods (piped but unfiltered water). This analysis model controls for both clustering of individuals within villages and for repeated assessments of villages over time... We will use multiple-imputation to impute missing outcomes at the individual person level which will then be aggregated for the village-level analyses." [Riverbank Filtration Trial]

Example 3 (Intention-to-treat analysis): "For the "intention-to-treat" analysis an indicator of whether an observation occurred pre- or post-randomisation was included in the regression model. To allow for delays in implementation a separate "per protocol" analysis was performed with the observations now placed into one of the three categories: "pre-randomisation", "post-randomisation but pre-implementation" and "post-implementation…" [FIT Trial]

Explanation: The statistical methodology should be clearly reported to allow replication. Where possible it can be helpful to provide a reference to the statistical methodology used. In a SW-CRT, clusters are randomised to sequentially initiate the intervention. Observations collected under the control condition are therefore, on average, from an earlier calendar time than observations collected under the intervention condition. Changes external to the trial may create underlying secular trends. Likewise participants, if repeatedly measured over the duration of the study, may get sicker or recover over time. This means that time is a potential confounder. Analysis of a SW-CRT should adjust for time effects [Hussey 2007] irrespective of their statistical significance; failure to do so risks biasing the estimate of the intervention effect, which could lead to declaring an intervention effective when it is ineffective or ineffective when it is effective [Hemming 2017]. It is therefore essential to report if and how time effects were

allowed for. If time is measured continuously, time can be modelled parametrically; if time is measured discreetly then time can be modelled categorically. Furthermore, SW-CRTs typically include only a small number of clusters [Martin 2016] and so pre-specification of important prognostic factors to use in a fully adjusted analysis (in mitigation of the likelihood of imbalance due to sampling variation) might also be undertaken [Senn 1994].

In a parallel CRT, randomisation at the level of the cluster needs to be allowed for at the analysis stage (unless cluster level data are being analysed). In a SW-CRT, as clusters (and possibly individuals) are repeatedly measured over time, there may be some reduction in the strength of correlation between measurements within the same cluster over time [Hooper 2016]. Failure to appropriately model the correlation structure can lead to incorrect estimation of the precision of treatment effects [Thompson 2017]. It is therefore important to clearly describe the correlation structure used in the analysis.

The analysis should also describe how deviations from the randomisation schedule were accommodated (Example 3). A more detailed consideration of this point is given under Item 16 (numbers analysed).

> *Item 12b: Additional statistical methods*
> *Standard CONSORT item:* Methods for additional analyses, such as subgroup analyses and adjusted analyses.
> *CONSORT cluster extension:* No modification suggested.
> *Extension for SW-CRTs:* Methods for additional analyses, such as subgroup analyses and adjusted analyses.
>
> *Example (Time varying effect of intervention):* "Furthermore, a delayed intervention effect of the CCs *(Case Conference i.e. intervention)* is assumed because the nurses need time to implement the procedure. Thus, the duration of the intervention in months must be considered." [FallDem Trial]

Explanation: SW-CRTs, like other trial designs, will commonly investigate subgroup differences and may perform adjusted analyses. In trials with a small number of clusters, investigating sensitivity to model assumptions will be important [Taljaard 2016].

Of some importance in a SW-CRT is time by treatment interactions. Treatment by time interactions are treatment effects which change as the study progresses (not to be confused with secular changes which represent changes in the outcome under the control condition– Table 2 Key concept 1). These changing treatment effects are important because observations contributing to the analysis will comprise a mixture of times since roll-out of the intervention. Interventions delivered at a single occasion (and not repeated to ensure it creates a permanent effect) might have an impact which changes with increasing time since roll-out (for example, the effect of the intervention might be quite large immediately after roll-out and then its impact might start to wane). If interventions are refined over time then their effect will also change over the duration of the study. Few trials if any have clearly investigated these time by treatment interactions [Davey 2015; Martin 2017], although many interventions have been assessed as being at risk of time by treatment interactions [Davey 2015]. The example above makes an acknowledgement of the possibility of a delayed effect, although gives limited detail as to how it will be investigated.

Of particular interest in a SW-CRT might be whether the intervention has a delayed effect (perhaps because its anticipated effect is not expected to materialise immediately (i.e. a lag effect); or if the intervention effect varies by time since exposure (e.g. an effect that decays over time or an effect that improves over time), perhaps because the effect of the intervention might be expected to wane with increasing time since exposure, particularly so in educational type interventions [Hughes 2015]; or perhaps due *to* the intervention being refined over the course of the roll-out.

Also of interest might be whether the effect of the treatment varies between sequences, perhaps because participants get sicker (or recover) with longer duration in the control condition and the treatment is not anticipated to have the same effect in sicker participants [Copas 2015].

Results: Participant flow

Item 13a: Participant flow
Standard CONSORT item: For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome.
CONSORT cluster extension: For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome.
Extension for SW-CRTs: For each treatment condition or allocated sequence, the numbers of clusters and participants who were assessed for eligibility, were randomly assigned, received intended treatments and were analysed for the primary outcome (Figure 3).

Item 13b: Participant attrition
Standard CONSORT item: For each group, losses and exclusions after randomisation, together with reasons
CONSORT cluster extension: For each group, losses and exclusions for both clusters and individual cluster members.
Extension for SW-CRTs: For each treatment condition or allocated sequence, losses and exclusions for both clusters and participants with reasons.

Example Flow chart by treatment condition and sequence (cross-sectional design): Supplementary Figure S2 (Long-live Mothers Trial)

Explanation: Information on the number of clusters and participants who were assessed for eligibility and outcomes along with the number of losses and exclusions (i.e. withdrawals) allows the reader to assess the risk of differential inclusion and attrition.

Any flow chart should allow the reader to examine the nature of any differential inclusion and attrition by allocated sequence, treatment condition, and over time (see Example Figure S2). Because there are many different types of SW-CRTs there is unlikely to be one flow-chart that will be applicable for all SW-CRTs. How the flow chart is constructed will depend on how many sequences and clusters there are, whether participants contribute repeated measures, and whether participants can join and leave the study. This information could be presented by allocated sequence but might also be presented by treatment conditions.

Including time periods in the flow chart is important to allow for assessment of differential participation over time. When different participants are sampled in each period, each participant will, in theory, be exposed to either the intervention or control condition. In this case, summarising the number of participants by treatment condition is possible. Where the same participant contributes multiple measurements, each participant may provide measurements under both intervention and control conditions. In this case, summarising the number of participants by allocated sequence, along with the average number of measurements contributed by each participant, is more appropriate.

Reporting the number of clusters and participants approached, eligible and included along with the reasons for non-participation is important to allow an assessment of study generalizability, and perhaps even more importantly, of biases due to differential participation between treatment conditions (or sequences). For example, in a parallel CRT without blinding of participants to treatment condition at the time of recruitment, a higher rate of consent among

those recruited to the intervention condition can indicate recruitment bias [Caille 2016]. Information on reasons as to why participants or clusters are not included allows a reader to assess the appropriateness of exclusions.

Results: Recruitment

Item 14a: Recruitment
Standard CONSORT item: Dates defining the periods of recruitment and follow-up.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Dates defining the steps, initiation of intervention and deviations from planned dates. Dates defining recruitment and follow-up for participants.

*Example 1 (Step dates): "*Twenty-two villages received the intervention in the second period (April-June 2011), 36 in the third period (September-November 2011), 35 in the fourth period (April-June 2012), and 35 in the fifth period (September-November 2012)." [DAVE Trial]

Example 2 (Deviations from planned dates): **"**There were 60 study wards in the 16 randomised hospitals, of which 33 (22 ACE and 11 ITU) in 13 hospitals went on to implement the intervention, with a mean (SD) delay in implementation of 5 (4) months …and a mean (SD) duration of implementation of 12 (7) months. Eight wards began implementation very late, and for these the end of the trial was extended to December 31st 2009 to ensure that they had a year of data collection post-implementation." [FIT Trial]

Explanation: Dates defining periods of recruitment of participants can be reported where appropriate; in some designs these dates will be at the beginning of the study before any cross-over of clusters occurs; in other designs recruitment will be continuous throughout the study. In some studies there will be no direct participant recruitment, but identification of data from participants from routine data sources.

Reporting of other key dates are also important in a SW-CRT. These dates include the dates defining when the study was undertaken and dates defining the steps. Dates defining the start and end of the roll-out phase, as well as the dates of the steps are useful to demonstrate if the trial was implemented as planned (Example 1). Dates should be presented so that they can be easily related to the planned timing of the steps as described in Item 3a. Reporting deviations from planned dates is particularly important in the SW-CRT as they demonstrate deviations from the randomised schedule (Example 2).

Dates defining implementation of interventions will allow assessment of when the intervention is fully implemented in each cluster. Dates defining actual implementation of the intervention should be specified. The realised time for an intervention to become fully implemented may differ from that which was planned. This allows assessment of whether all observations collected under the intervention condition were fully exposed to the intervention; it also allows assessment of whether any observations collected under the control condition were likely contaminated by the intervention. Reporting dates also allows inferences about external influences which may have affected secular trends.

Item 14b: Recruitment
Standard CONSORT item: Why the trial ended or was stopped.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Why the trial ended or was stopped.

Explanation: Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010, Campbell 2012].

Results: Baseline data

Item 15: Baseline data
Standard CONSORT: A table showing baseline demographic and clinical characteristics for each group.
CONSORT cluster extension: Baseline characteristics for the individual and cluster levels as applicable for each group.
Extension for SW-CRTs: Baseline characteristics for the individual and cluster levels as applicable for each treatment condition or allocated sequence.

Example 1 Baseline table by treatment condition (cross-sectional design): Supplementary Table S2 (DAVE Trial)
Example 2 Baseline table by allocated sequence (open cohort design): Supplementary Table S3 (Depression Management Trial)

Explanation: In a parallel CRT a summary of the cluster and participant level characteristics at baseline by treatment condition can allow assessment of the success of randomisation and provides a description of the included sample. In trials with post-randomisation recruitment, this table can allow an assessment of potential biases.

The term "baseline" in a SW-CRT can be confusing because of the longitudinal nature of the design. We use the term "baseline characteristic" to mean a characteristic which was either measured before exposure to the control or intervention condition, or which is not expected to be influenced by the treatment conditions (e.g. age). In designs in which observations are made on different participants in each period, these baseline characteristics will often pertain to measurements made just prior to the switch from control to intervention condition (i.e. not at the start of the trial); whereas in designs where participants are repeatedly assessed, these characteristics might be measured prior to randomisation. Cluster level characteristics can often be measured prior to randomisation and are less likely to change over time.

For SW-CRTs in which observations are made on different participants in each period, the summary of baseline characteristics could be presented by treatment condition or by allocated sequence. For example, the DAVE Trial, which measures different participants in each period, reports its baseline table by treatment condition (Table S2).

For SW-CRTs in which the same participants are repeatedly assessed in each of the periods, the baseline characteristics of participants will normally be presented by allocated sequence rather than by treatment condition. This is because most participants will be observed first under the control and then intervention condition. The Depression Management Trial (Table S3) provides summary characteristics by allocated sequence.

Results: Numbers analysed

Item 16: Numbers analysed
Standard CONSORT: For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups.
CONSORT cluster extension: For each group, number of clusters included in each analysis.
Extension for SW-CRTs: The number of observations and clusters included in each analysis for each treatment condition and whether the analysis was according to the allocated schedule.

Example 1 (Numbers by treatment condition): "A total of 5295 surgical procedures were carried out throughout the stepped wedge cluster RCT, that is, 2212 in control and 3083 (of which 2263 had the SSC performed) after implementation of the SSC (*Surgical Safety Checklist*). Patients (14.9%; 667/4475) underwent more than 1 procedure. The control and SSC study steps included 1778 and 2033 unique patients, respectively." [Surgical Checklist Trial]

Example 2 (Intention-to-treat vs. per protocol): "The flow diagram shows there were 60 study wards in the 16 randomised hospitals, of which 33 (22 ACE and 11 ITU) in 13 hospitals went on to implement the intervention… For the primary outcome, intention-to-treat analysis was conducted for the 60 wards randomised into the intervention, and per-protocol analysis was performed for the 33 implementing wards…" [FIT Trial]

Explanation: The number of observations by treatment condition should be reported for analyses of all outcomes (Example 1). For some outcomes this information will be included in a flow chart although not all flow charts for a SW-CRT will give an immediate summary of this information by treatment condition. When the same participants are repeatedly measured across the time periods, each participant will have been exposed to both treatment conditions and so this information can be reported either by giving the total number of observations (by treatment condition) or as the number of participants in the study and average number of assessments per participant under each treatment condition. Where different participants contribute to each measurement period, it might be useful to have information on the number of participants per cluster-period. Such information might be most easily reported in a diagram rather than in text (Figure 3).

Sometimes clusters (and perhaps participants) will not receive the intervention condition as per the randomisation schedule (Example 2). In a parallel trial an intention-to-treat analysis performs the analysis according to the groups to which participants or clusters were originally assigned [Moher 2012]. In a SW-CRT this might be interpreted as analysis of clusters and participants treated as exposed to the intervention according to the *dates* of the randomisation schedule (i.e. according to the planned dates). The application of this principle would mean that clusters are treated as exposed to the intervention if the observation comes from a time period post allocated cross-over date. When a SW-CRT has randomised clusters to actual dates of transitioning from control to intervention, an intention-to-treat analysis following this interpretation is logical.

Alternatively, a SW-CRT might be considered as randomising the order that the clusters transition from control to intervention (although when there are multiple clusters per sequence, several clusters share the same rank-order). In this situation an intention-to-treat analysis might be interpreted as analysis of clusters and participants treated as exposed to the intervention according to the *order* of the randomisation schedule (i.e. according to the planned order of roll-out). The application of this principle would mean that clusters are treated as exposed to the intervention only after the intervention has been implemented in that cluster, provided the order of the allocation did not deviate from that planned.

Providing information on the number of clusters (and participants) contributing to all analyses allows assessment of whether the analysis has been conducted with respect to the randomised cross-over schedule – which might not be in strict accordance with any pre-specified dates; or to the actual cross-over dates that may deviate from planned dates due to delays in implementation.

Sometimes a cluster may drop out from some purposively collected outcome assessments, but still contribute data from routinely collected sources for other outcome variables. If the numbers included in secondary analyses differ from those included in primary analyses, information on differential attrition (or participation) across clusters or periods can be provided in the text (similar to information depicted in the flow chart for the primary outcome (Figure 3).

Results: Outcomes and estimation

Item 17a: Outcomes and estimation
Standard CONSORT item: For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).

CONSORT cluster extension: Results at the individual or cluster level as applicable and a coefficient of intra-cluster correlation (ICC or k) for each primary outcome.
Extension for SW-CRTs: For each primary and secondary outcome, results for each treatment condition, and the estimated effect size and its precision (such as 95% confidence interval); any correlations and time effects estimated in the analysis.

Example 1 (Time adjusted treatment effect): "A total of 321 (10.8%) unexposed patients were started on either antihypertensives or statins, and 577 (19.7%) exposed patients. The time-adjusted mean difference in proportion of patients initiating either treatment was 15.5% (95% CI = 3.9 to 27.1)." [Targeted Case Finding Trial]

Example 2 (Secular trend): Supplementary Figure S3 [FIT Trial]

Example 3 (Correlations): "The ICC in the time-adjusted analysis for initiation of either treatment was 0.014 (95% CI = 0.005 to 0.038)." [Targeted Case Finding Trial]

Explanation: A summary of the findings for each primary and secondary outcome should be provided for each treatment condition. This will allow a description of the severity or prevalence of the outcome in the sample (Example 1). In addition, reporting of results by treatment condition allows estimation of an unadjusted effect of the intervention for comparison with a time adjusted effect (as in Example 1).

Treatment effects should be reported along with 95% Confidence Intervals (CI). A SW-CRT which does not adjust for time is analogous to a simple uncontrolled before-and-after experiment; therefore, it should be clearly reported if the primary and secondary outcomes were adjusted for time (Example 1). To allow an understanding of the potential impact of secular trends it can be helpful to describe the secular trend – either in a figure or as regression coefficients. Ideally this should be done by calendar time and should represent the trend in the clusters yet to be exposed to the intervention (Example 2: Figure S3). In some SW-CRTs participants will be recruited at the very beginning of the trial and measured repeatedly. In chronic conditions these participants may naturally regress over the duration of the study; in acute conditions they may recover. Whilst not a secular trend per se, such effects still may lead to confounding of the intervention effect with time and so time should be adjusted for.

Reporting any estimated coefficients of intra-cluster correlation (ICCs) can be informative for the planning of future trials (Example 3). Correlation structures are more complex than in a parallel cluster trials conducted at a single cross-section in time; therefore, analysis (and reporting) of a single measure of correlation such as the ICC might not be sufficient [Kasza 2017]. Relevant correlation coefficients might include correlations between observations in the same cluster and same time period (within-period ICC); correlations between observations in the same cluster but different time periods (between-period ICC), as well as between-period and within-period correlations on the same *individual* [Hooper 2016]. It is important to be explicit about the types of correlations being reported [Martin 2016b]. Reporting of variance components is an alternative to intra-cluster correlations, particularly for non-continuous outcomes [Hayes 1999]. When intra-cluster correlations are reported for binary outcomes, clearly indicating the scale (e.g. proportions or logistic scale) can help interpretation [Eldridge 2009].

Item 17b: Binary outcomes
Standard CONSORT item: For binary outcomes, presentation of both absolute and relative effect sizes is recommended.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: For binary outcomes, presentation of both absolute and relative effect sizes is recommended.

Explanation: In addition to reporting a relative measure of the effect of the intervention it can be helpful to report an absolute measure of the effect: while absolute measures of effects are more easily understood, relative measures of effects are often more stable across different populations [Ukoumunne 2008].

While reporting relative and absolute measures of effects is recommended, further methodological work is required to determine optimal methods of analysis that yield such estimates. Current approaches include fitting two separate models (for example a binomial model with log link to report the relative risks; and a binomial model with an identity link to report a risk difference) or by fitting one model and using a transformation to report the other measure of treatment effect [Pedroza 2016].

Model based methods for achieving estimates on both scales have been investigated in parallel CRTs in which the model is unadjusted for confounders [Ukoumunne 2008]; and others have evaluated the performance of these models when covariate adjustment is required [Pedroza 2016].

Results: Ancillary analyses

Item 18: Ancillary analyses
Standard CONSORT item: Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory.

Explanation: There are several analyses that can be considered to examine deviation from model assumptions, for example, variations in secular trends across groups of clusters [Hemming 2017]; interactions of the intervention effect with sequence; and whether the effect of the intervention might change with increasing duration of exposure (Item 12b). In the reporting of these ancillary analyses, any limitations due to the assumptions made should be noted.

Results: Harms

Item 19: Harms
Standard CONSORT item: All important harms or unintended effects in each group (for specific guidance see CONSORT for harms).
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Important harms or unintended effects in each treatment condition (for specific guidance see CONSORT for harms).

Explanation: Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Discussion

Item 20: Limitations
Standard CONSORT item: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.

Explanation: Estimated intervention effects from a SW-CRT will almost always be model-based estimates adjusting for time. There is a host of different models which can be used, but all make some assumptions. The assumptions made and potential limitations should be reflected on.

Item 21: Discussion
Standard CONSORT item: Generalisability (external validity, applicability) of the trial findings.
CONSORT cluster extension: Generalisability to clusters and/or individual participants (as relevant)
Extension for SW-CRTs: Generalisability (external validity, applicability) of the trial findings. Generalisability to clusters and/or individual participants (as relevant).

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010, Campbell 2012].

Item 22: Interpretation
Standard CONSORT item: Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.
CONSORT cluster extension: No modification suggested
Extension for SW-CRTs: Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Other information

Item 23: Trial registration
Standard CONSORT item: Registration number and name of trial registry.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Registration number and name of trial registry.

Explanation: The International Committee of Medical Journal Editors (ICMJE) defines a clinical trial "as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the cause-and-effect relationship between a health-related intervention *and* a health outcome" [ICMJE]. The ICMJE states that all medical journal editors should require clinical trials to be registered (prior to the first patient enrolment) as a condition of publication. SW-CRTs of health related interventions meet the ICMJE's definition of a clinical trial and so should wherever possible be registered as a clinical trial prior to the study start date.

Reporting the name of the trial registry and the unique trial registration number facilitates crosschecking with the associated registry entry and allows assessment of whether there are any important changes to the trial design, and the potential for any bias (such as outcome reporting bias). Further, reporting details of the trial registration facilitates linking of multiple publications from the same trial, which is of particular importance for systematic reviews. If the trial has not been registered, this should be stated along with the reason.

Studies examining trial registration rates have found that a large percentage of trials are not registered (e.g. 28% - 44% [Azar 2015; Killeen 2014; Wetering 2012]). Further, in the trials that are registered, not all report the registration details in the trial publication, and not all are prospectively registered. A recent review that examined registration of SW-CRTS found that only 50% of SW-CRTs were prospectively registered [Taljaard 2017].

Item 24: Trial protocol
Standard CONSORT item: Where the full trial protocol can be accessed, if available.
CONSORT cluster extension: No modification suggested
Extension for SW-CRTs: Where the full trial protocol can be accessed, if available.

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Item 25: Funding
Standard CONSORT item: Sources of funding and other support (such as supply of drugs), role of funders.
CONSORT cluster extension: No modification suggested.
Extension for SW-CRTs: Sources of funding and other support (such as supply of drugs), role of funders.

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010, Campbell 2012].

Item 26: Research Ethics Review
Standard CONSORT item: Not included.
CONSORT cluster extension: Not included
Extension for SW-CRTs: Whether the study was approved by a research ethics committee, with identification of the review committee(s). Justification for any waiver or modification of informed consent requirements.

Example 1 (Full review): "The study received ethical approval from the Sport and Health Sciences Ethics Committee at the University of Exeter (February 2011)." [DAVE Trial Protocol]
Example 2 (Waiver of consent): "This study was reviewed by the Regional Committee for Medical and Health Research Ethics (Ref: 2009/561), which advised that use of routinely collected anonymized patient data is clinical service improvement and thus no further approval or patient consent is required."[Surgical Checklist Trial]

Explanation: The original CONSORT statement did not include an item on research ethics approval because it is an existing International Committee of Medical Journal Editors requirement that research "involving human data" should indicate whether the research was reviewed by a research ethics committee [ICMJE]. However, a systematic review found that only 75% of SW-CRTs reported review by a research ethics committee, possibly due to the classification of such studies, by some researchers, as service development or quality improvement. To encourage clear reporting about research ethics review of SW-CRTs we have therefore included this as a new item. This is consistent with the recent extension to the CONSORT statement for pilot studies, which also included this as a new item [Eldridge 2016]. An application number or reference number of the ethical approval should also be reported. If a study is deemed exempt from review by a research ethics committee, this should be reported together with a clear justification for the exemption from review.

Conclusions

The SW-CRT offers an exciting new opportunity to rigorously examine the effects of implementation, policy and service delivery interventions. The design is appealing in many respects, but also provides many challenges. It has noteworthy risks for biases including bias due to temporal trends and within-cluster contamination, as well as methodological complexities such as changes in correlation structures over time. Furthermore, perhaps because the design is being used in situations where researchers are not familiar with standards for reporting or conduct, SW-CRTs have been noted to be particularly prone to inadequacies of ethical reporting, including research ethics review and (in common with many cluster trials) identification of research participants. This extension of the CONSORT

statement for SW-CRTs encourages researchers to reflect on the unique aspects of the SW-CRT and improve the clarity of reporting.

Table 1 Glossary of terms

Term	Explanation
Cluster	The unit of randomisation.
Cluster-period	A grouping of observations by time of measurement and cluster.
Step	A planned point at which a cluster or group of clusters crosses from control to intervention.
Period	A grouping of observations by time of measurement.
Duration of period	Time (e.g. months) between each step.
Sequence of treatments (often abbreviated to sequence or allocated sequence)*	A sequence of codes defining the order of implementation of the treatment conditions for each cluster. More than one cluster can be allocated to each sequence.
Intervention condition*	The treatment under evaluation.
Control condition	The comparator treatment.
Transition period	The time needed to fully embed the intervention. A transition period may have the same or different duration than a measurement period.
Participant	A participant is someone on whom investigators seek to measure the outcome of interest.
Research participant	A research participant denotes a human research subject from the standpoint of ethical considerations.
Open cohort	A study design in which participants are repeatedly assessed over a series of measurement points and can join and leave the study throughout its duration.
Closed cohort	A study design in which participants are repeatedly assessed over a series of measurement points and cannot join the study once it has started.
Cross-sectional	A study design in which different participants are measured at each measurement occasion.
Complex intervention	An intervention that has multiple and interacting parts.
Purposively collected data	Data that are collected for the specific purpose of contributing to the trial (data that are not routinely collected).

*Note the CONSORT statement uses the term "group" to refer to the allocated treatment, but for SW-CRTs we distinguish between the concepts of the allocated sequence and the treatment condition in any given period of that sequence, and avoid terminology such as "group" or "arm". We use the term "treatment" in a generic way to refer to either the active treatment or comparator; and retain the use of the phrase "intervention condition" to refer to the active treatment of the trial; and the "control condition" to refer to the comparator.

Key concept	Detailed Description	Why this is important	Mitigating strategies
Imbalance of the design with respect to time	In a SW-CRT, clusters are randomised to different sequences which dictate the order they initiate the intervention. Observations collected under the control condition are, on average, from an earlier calendar time than observations collected under the intervention condition.	Changes external to the trial may create underlying secular trends. In addition, where the same participants are repeatedly assessed, their health status might improve (or worsen) over the study. Because time is associated with both the treatment condition and the outcome, it means that time is a potential confounder.	Analysis and sample size should allow for the confounding effect of time.
Repeated measures on same clusters and possibly same participants	SW-CRTs make a series of measurements over time within each cluster. These repeated measurements can be on the same participants, different participants, or a mixture of the same and different participants at each measurement.	Correlation structures are more complex than in a parallel cluster trial conducted at a single cross-section in time.	Analysis (and consequently sample size calculations) should allow for the fact that data are not independent and dependencies might vary overtime.
Within cluster contamination	In SW-CRTs, some or all of the clusters will be exposed to both the control and intervention conditions. Participants can either have a relatively short exposure to the intervention (surgical intervention) or long exposure (change in care home policy).	Where duration of exposure is short it is unlikely that individuals will be exposed to both the control and intervention condition. Where the duration of exposure is long, it may be possible that some participants are exposed to both the control condition the intervention condition.	In trials with long exposure, delayed assessment of outcomes should be avoided to prevent participants recruited under the control condition later becoming exposed to the intervention condition.
Delayed treatment effects and transition periods	Sometimes the effect of the intervention is expected to materialise immediately, and sometimes there is a delay before its effect will be realised.	When there is a delay before the effect of the intervention is realised the estimate of effectiveness can be attenuated.	Where there is an expected delay before the effect of the intervention is materialised a transition period can be built into the design of the study.
Time by treatment effect interactions	SW-CRTs can evaluate interventions of many different forms. The intervention can be a one-off delivery involving a "permanent" change to a health care system, or it can be an intervention which may need to be repeated multiple times to ensure its effects are realised such as education of health professionals. Sometimes the intervention may be refined over the duration of the study.	Interventions delivered at a single occasion (and not repeated to ensure it creates a permanent effect) might have an impact which changes with increasing time since roll-out (for example, the effect of the intervention might be quite large immediately after roll-out and then its impact might start to wane). If interventions are refined over time then their effect will also change over the duration of the study.	If interventions are either refined over time or are not expected to create a permanent effect, an analysis examining how the effect of the treatment changes with time should be considered.
Sampling of observations	SW-CRTs can take a complete enumeration of the cluster, a random sample of individuals, or recruit participants into the trial. Furthermore, participants might be continuously recruited into the trial as they present; or all participants might be recruited at the beginning of the trial.	Information on how observations were sampled is important to elicit risks of bias. Studies which take a complete enumeration have lower risks of bias as do studies which recruit all participants at a fixed point in time before randomisation has occurred; studies which continuously recruit participants have higher potential for identification and recruitment biases.	Methods to reduce the risk of bias include taking a complete enumeration of the entire cluster-period, recruiting all participants before randomisation, or recruiting by someone independent to the study.
Continuous or discrete time measurements	Observations may be accrued continuously in time (e.g., as patients present to an emergency department and provide measurements after a follow-up period); or in discrete time (e.g., a survey questionnaire may be implemented at several discrete points in time).	Where observations are accrued in continuous time, outcomes are more likely to be measured in continuous time; where outcomes are accrued in discrete time, outcomes are more likely to be measured in discrete time.	Collecting exact timings of outcomes will ensure the full possible range of analysis methods can be implemented.
Justification of study type	Justifying the need for a staggered roll-out of the intervention using a SW-CRT, as opposed to a simple parallel arm implementation, is important because the SW-CRT is more	Risks of bias in the SW-CRT may be higher than in a parallel CRT. For example, secular trends may be of concern in a SW-CRT, but not in a parallel design.	SW-CRTs should be classified as research and so should be registered as a trial and should be submitted for review to an

Table 2 Key methodological considerations to consider in the reporting of a SW-CRT

	complicated in its design, analysis, and implementation than the parallel CRT. It might also involve exposing a greater number of clusters or participants to the intervention.		approved research ethics committee.

Table 3 Checklist of information to include when reporting a stepped-wedge cluster randomised trial

Section/Topic	Item No	Checklist item	Page Number
Title and abstract			
	1a	Identification as a stepped-wedge cluster randomised trial in the title.	
	1b	Structured summary of trial design, methods, results, and conclusions (see separate SW-CRT checklist for abstracts).	
Introduction			
Background and objectives	2a	Scientific background. Rationale for using a cluster design and rationale for using a stepped-wedge design.	
	2b	Specific objectives or hypotheses.	
Methods			
Trial design	3a	Description and diagram of trial design including definition of cluster, number of sequences, number of clusters randomised to each sequence, number of periods, duration of time between each step and whether the participants assessed in different periods are the same people, different people, or a mixture.	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons.	
Participants	4a	Eligibility criteria for clusters and participants.	
	4b	Settings and locations where the data were collected.	
Interventions	5	The intervention and control conditions with sufficient details to allow replication, including whether the intervention was maintained or repeated, and whether it was delivered at the level of the cluster, the individual, or both.	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined. Method of calculation and relevant parameters with sufficient detail so the calculation can be replicated (see separate checklist for SW-CRT sample size items). Assumptions made about correlations between outcomes of participants from the same cluster.	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation to the sequences of treatments.	

Table 3 Checklist of information to include when reporting a stepped-wedge cluster randomised trial

	8b	Type of randomisation; details of any constrained randomisation or stratification if used.
Allocation concealment mechanism	9	Specification that allocation was based on clusters; description of any methods used to conceal the allocation from the clusters until after recruitment.
Implementation	10a	Who generated the randomisation schedule, who enrolled clusters, and who assigned clusters to sequences.
	10b	Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling; continuous recruitment/ascertainment, or recruitment at a fixed point in time), including who recruited or identified participants.
	10c	Whether, from whom and when consent was sought and for what; whether this differed between treatment conditions.
Blinding	11a	If done, who was blinded after assignment to sequences (for example, cluster level participants, individual level participants, those assessing outcomes) and how.
	11b	If relevant, description of the similarity of treatments.
Statistical methods	12a	Statistical methods used to compare treatment conditions for primary and secondary outcomes including how time effects, clustering and repeated measures were taken into account.
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses.
Results		
Participant flow (a diagram is strongly recommended)	13a	For each treatment condition or allocated sequence the numbers of clusters and participants who were assessed for eligibility, were randomly assigned, received intended treatments and were analysed for the primary outcome.
	13b	For each treatment condition or allocated sequence, losses and exclusions for both clusters and participants with reasons.
Recruitment	14a	Dates defining the steps, initiation of intervention and deviations from planned dates. Dates defining recruitment and follow-up for participants.
	14b	Why the trial ended or was stopped
Baseline data	15	Baseline characteristics for the individual and cluster levels as applicable for each treatment condition or allocated sequence.
Numbers analysed	16	The number of observations and clusters included in each analysis for each treatment condition and whether the analysis was according to the

Table 3 Checklist of information to include when reporting a stepped-wedge cluster randomised trial

		allocated schedule.
Outcomes and estimation	17a	For each primary and secondary outcome, results for each treatment condition, and the estimated effect size and its precision (such as 95% confidence interval); any correlations and time effects estimated in the analysis.
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory
Harms	19	Important harms or unintended effects in each treatment condition (for specific guidance see CONSORT for harms)
Discussion		
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.
Generalisability	21	Generalisability (external validity, applicability) of the trial findings. Generalisability to clusters and/or individual participants (as relevant).
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.
Other information		
Registration	23	Registration number and name of trial registry.
Protocol	24	Where the full trial protocol can be accessed, if available.
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders.
Research Ethics review	26	Whether the study was approved by a research ethics committee, with identification of the review committee(s). Justification for any waiver or modification of informed consent requirements.

Table 4 Noteworthy changes to the CONSORT 2010 statement and the 2012 extension for cluster trials

Noteworthy changes to the CONSORT 2010 Statement	Separate presentation of the CONSORT checklist items for SW-CRTs (see Table 3).
	Modification of Item 2a (Background) to include rationale for use of a stepped-wedge design
	Extension of Item 3a (Design) to include a schematic representation of the design; and clarity over key design aspects (such as number of steps, number of observations per cluster-period).
	Extension of Item 7a and 12a (Sample Size and Statistical Methods) to include reference to the methods used to allow for adjustment for time and assumptions made about correlations.
	Extension of Item 12b (Auxiliary analyses) to include any sensitivity analyses for assumptions made about time effects.
	Extension of Item 13a (Participant flow) to include a modified flow-chart by allocated sequence (see Figure 3).
	Extension of Item 17a (Outcomes and Estimation) to report any adjustment for time effects; and presentation of secular trends (see Figure S2)
	Extended elaboration under Item 18 (Auxiliary analyses) to include reporting of any sensitivity analyses for any model based methods; and extended elaboration under Item 20 (Limitations) to include discussion of any limitations due assumptions made about time effects.
	Extended elaboration under Item 5 (Interventions) to include planned details on timings of interventions; and under Item 6 (Outcomes) timings of outcome assessments. This information, along with the corresponding realised dates under Item 14a (Recruitment Dates) allow determination of the risk of within cluster contamination.
	Addition of Item 26 (Research Ethics Review) to include reporting of ethical review and consent processes.
Noteworthy deviations from the CONSORT 2012 extension for cluster randomised trials	Modification of wording of Item 2b (Objectives) from “Whether objectives pertain to the cluster level, the individual participant level or both.” which was deemed ambiguous to “Specific objectives or hypotheses”.
	Modification of Item 9 (Allocation Concealment) to reference only allocation concealment from the unit of randomisation (i.e. cluster) and not participant (comes under Item 10b).

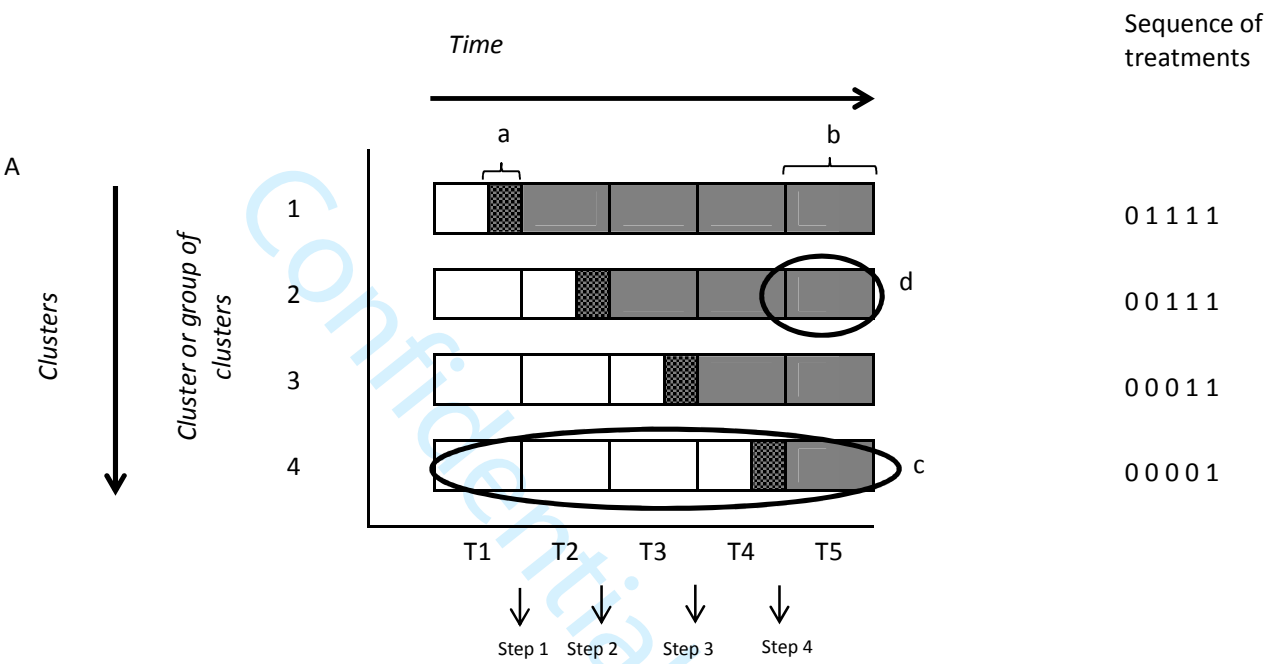
Table 5 Items to report in the journal abstract of a SW-CRT

Abstract Item	Extension for SW-CRTs
Title	Identification of study as a stepped-wedge cluster randomised trial.
Trial design	Description of the trial design (including numbers of sequences and clusters, and whether participants assessed in different periods are the same people, different people, or a mixture).
Methods	
Participants	Eligibility criteria for clusters and participants.
Interventions	The intervention and control conditions.
Objective	Specific objective or hypothesis.
Outcome	Clearly defined primary outcome.
Randomisation	How clusters were allocated to sequence of treatments.
Blinding (masking)	Whether participants, healthcare professionals, those recruiting and those assessing outcomes were blinded.
Results	
Numbers randomised	Number of clusters randomised to each sequence of treatments.
Recruitment	Trial status.
Numbers analysed	Number of observations and clusters included in the analysis.
Outcome	For the primary outcome, the estimated effect size (and CI) and reporting of any adjustment for secular trends.
Harms	Important adverse events or side effects.
Conclusions	General interpretation of the results.
Trial registration	Registration number and name of trial register. Ethical approvals.

Table 6: Essential and additional information to report under sample size calculation (Item 7a)

	Further explanation
Essential information for reporting	
Level of significance	State whether a one or two-sided test was used.
Power	
Target difference	
Variation of outcome	For continuous variables this will be a standard deviation; and for binary variables this will be the control proportion.
Number of clusters	There should be clarity between the total number of clusters and the number of clusters allocated to each sequence. A diagram can be helpful.
Number of sequences	
Average cluster size	There should be clarity between cluster size per measurement period and total cluster size.
The assumed correlation structure	The assumed intra-cluster correlation coefficient (ICC) and whether the ICC is time dependent or time independent. If time-dependent, state the parameters that were assumed to accommodate the time-dependency, for example, the within-period ICC and the between-period ICC or the cluster autocorrelation coefficient, or any variance components. For binary outcomes it is important to report the scale of the correlations or variance components (e.g., proportions scale or logistic scale).
Within person correlations	Where the design includes repeated measurements on the same individual, describe the assumed correlation structure at the individual level, including if any decay in correlation in repeated measures on the same individual has been accounted for (e.g., an individual auto-correlation coefficient).
Additional information for reporting	
Method used	Reference to the methodology used and statistical packages (including details of functions) used for implementation.
Allowance for variation in cluster size	Whether variation in cluster sizes were accommodated and how. This can include variation in total cluster sizes or variation in cluster-period sizes
Allowance for attrition	This can include attrition both at the cluster level and the individual level. If included, provide an explanation of how this was allowed for.
Number of clusters per sequence	If an unequal number of clusters per sequence was used, include information on whether this was accounted for in the sample size calculation.
Allowance for transition periods	State whether any transition periods were allowed for and how. This includes a description of the duration of the transition period and whether these data were excluded from the sample size calculation, or included with alternative coding of the intervention indicator
Sensitivity analysis	This can include sensitivity to all parameters which might vary in the actual trial. A justification should be provided for all assumed sample size parameters

Figure 1 Diagram of the standard stepped-wedge cluster randomised trial



Key

	Control condition
	Transition period
	Intervention condition
a	Duration of transition-period
b	Duration of a time-period
c	Cluster
d	Cluster-period
T1	Time period 1 etc.
0	Control condition
1	Intervention condition

Note that in designs where participants are measured after a follow-up time from their exposure, then the periods and their representation as in Figure 1 are defined based on when an individual was exposed and not when measured.

Figure 2 Example of a diagram of a SW-CRT taken from the Riverbank Filtration Trial

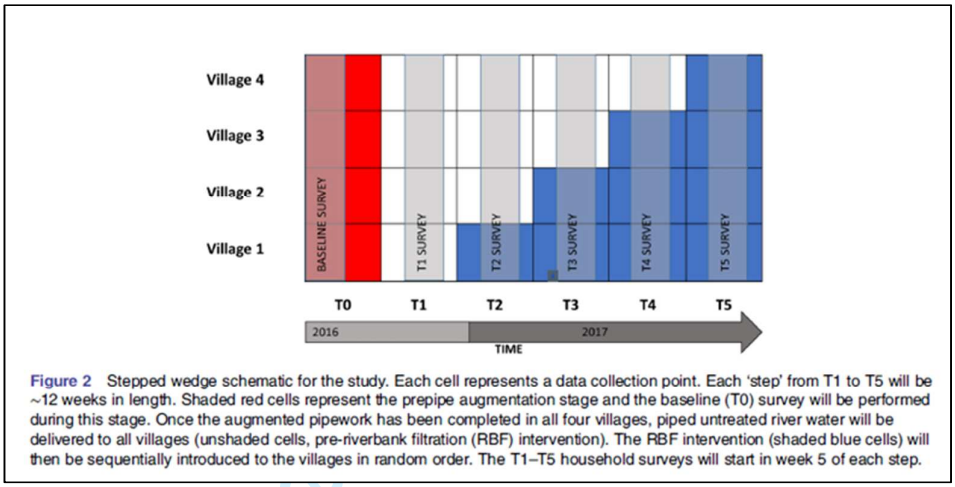
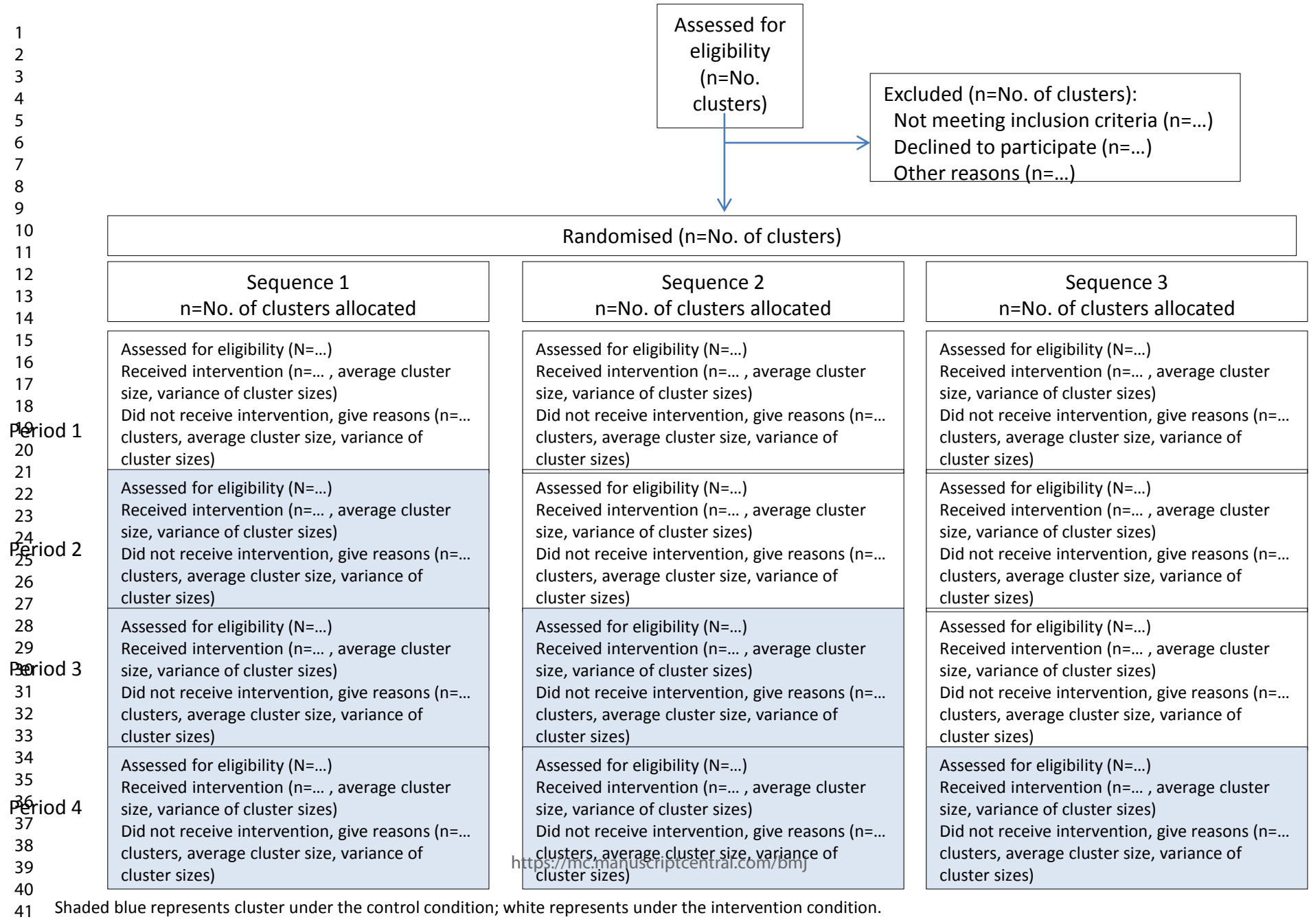


Figure 2 Stepped wedge schematic for the study. Each cell represents a data collection point. Each 'step' from T1 to T5 will be ~12 weeks in length. Shaded red cells represent the prepipe augmentation stage and the baseline (T0) survey will be performed during this stage. Once the augmented pipework has been completed in all four villages, piped untreated river water will be delivered to all villages (unshaded cells, pre-riverbank filtration (RBF) intervention). The RBF intervention (shaded blue cells) will then be sequentially introduced to the villages in random order. The T1–T5 household surveys will start in week 5 of each step.

Taken from Figure 2 in McGuinness SL, O'Toole JE, Boving TB, Forbes AB, Sinclair M, Gautam SK, Leder K. Protocol for a cluster randomised stepped wedge trial assessing the impact of a community-level hygiene intervention and a water intervention using riverbank filtration technology on diarrhoeal prevalence in India. BMJ Open. 2017 Mar 17;7(3):e015036. doi: 10.1136/bmjopen-2016-015036. PubMed PMID: 28314746; PubMed Central PMCID: PMC5372111.

Figure 3 Specimen flow chart for a SW-CRT by allocated sequence and period



Shaded blue represents cluster under the control condition; white represents under the intervention condition.

Figure S1 Example of a diagram of a SW-CRT taken from the DAVE Trial Protocol

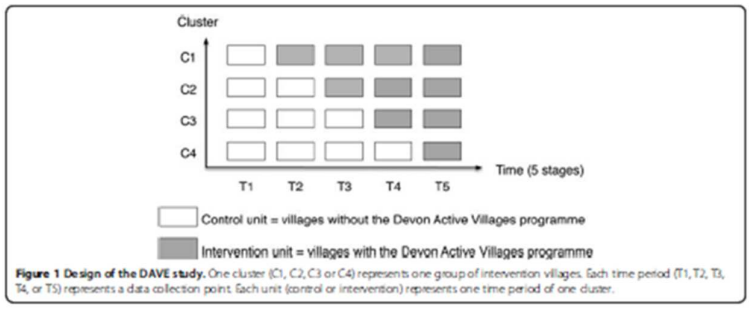
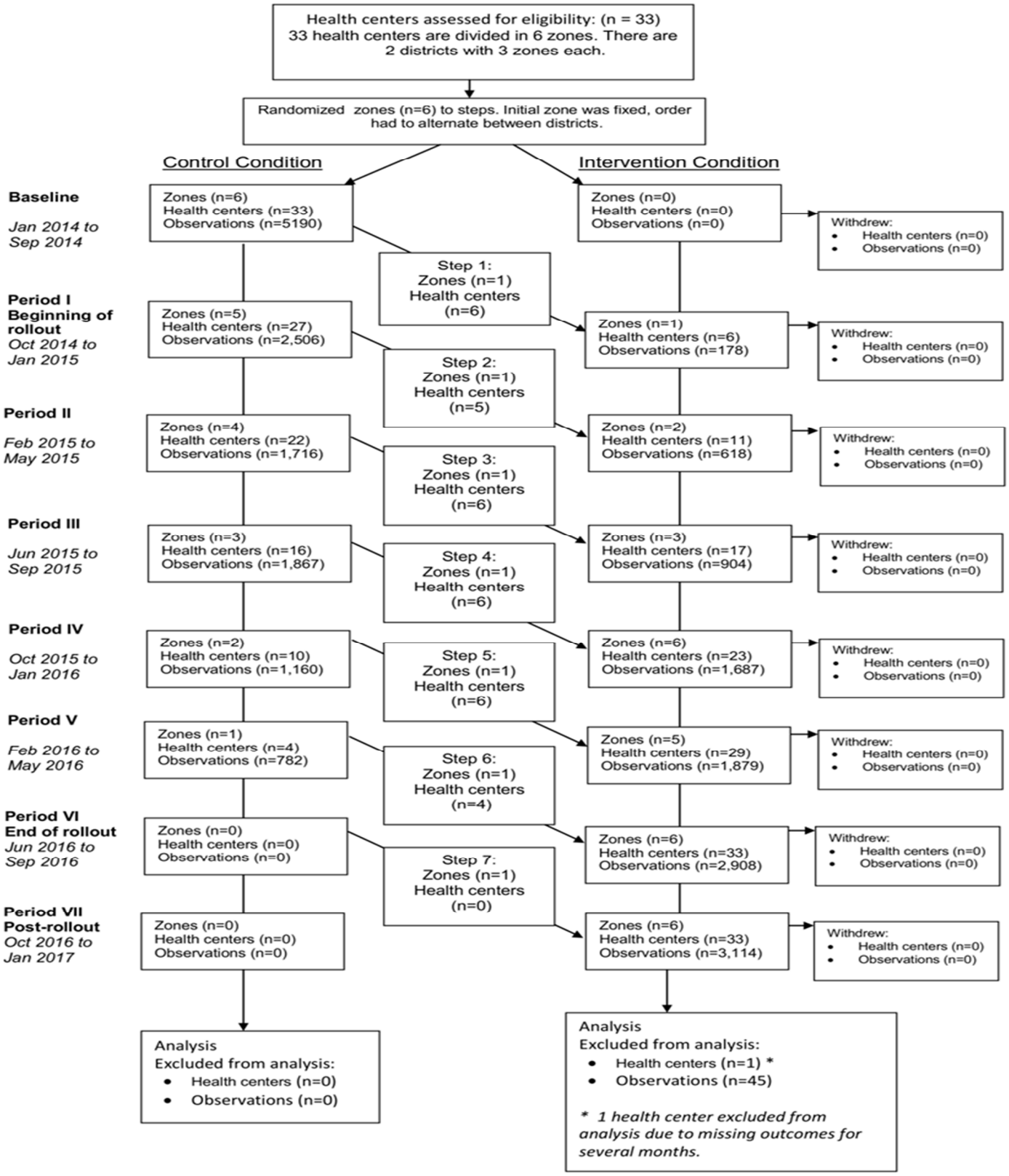


Figure 1 Design of the DAVE study. One cluster (C1, C2, C3 or C4) represents one group of intervention villages. Each time period (T1, T2, T3, T4, or T5) represents a data collection point. Each unit (control or intervention) represents one time period of one cluster.

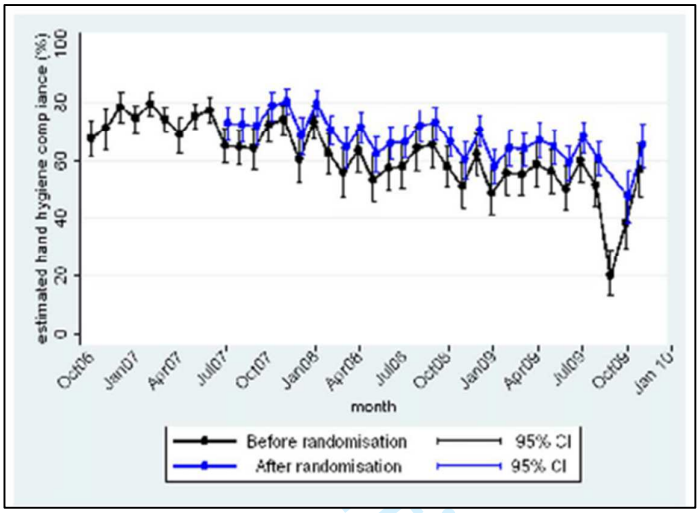
Taken from Figure 1 in Solomon E, Rees T, Ukoumunne OC, Hillsdon M. The Devon Active Villages Evaluation (DAVE) trial: study protocol of a stepped wedge cluster randomised trial of a community-level physical activity intervention in rural southwest England. BMC Public Health. 2012 Aug 1;12:581. doi: 10.1186/1471-2458-12-581. PubMed PMID: 22849310; PubMed Central PMCID: PMC3496564.

Figure S2: Example of a flow-chart for a SW-CRT taken from the Long-live Mothers trial



Taken from Figure 1 in Kestler E. , Ambrosio G. , Hemming K. , Hughes J. Matute J. Moreno M. and Walker D. Scaling up an integrated approach to improve care during delivery in northern Guatemala: a stepped-wedge cluster randomized trial Submitted.

Figure S3 Example of a secular trend taken from the FIT Trial



Note: “Before randomisation” refers to observations under the control condition; and “after randomisation” to observations under the intervention condition.

Taken from Figure 3 in Fuller C, Michie S, Savage J, McAteer J, Besser S, Charlett A, Hayward A, Cookson BD, Cooper BS, Duckworth G, Jeanes A, Roberts J, Teare L, Stone S. The Feedback Intervention Trial (FIT)--improving hand-hygiene compliance in UK healthcare workers: a stepped wedge cluster randomised controlled trial. PloS One. 2012;7(10):e41617. doi: 10.1371/journal.pone.0041617. PubMed PMID: 23110040; PubMed Central PMCID: PMC3479093.

Table S1 Example of an abstract (results are made up)

An integrated approach to improve care during delivery in a low income country: a stepped-wedge cluster randomized trial

Background: Rural communities in low income countries, where most deliveries take place at home under the care of a traditional birth attendant, have high rates of complications. The objective of this study was to evaluate the impact of a package of interventions, with the aim of encouraging women to deliver at health centres and training traditional birth attendants, on adverse maternal and child health indicators.

Methods: The intervention package was implemented in a random order using a stepped-wedge design across the six sub-districts of two purposively selected (high maternal morbidity) districts of the country over the period January-2014 to January-2017. The intervention was implemented in sequentially with one of the six sub-districts transitioning to the intervention every four months. The randomisation was stratified by the two participating districts with one sub-district randomly selected to be allocated first in the order. Data on outcomes were collected on all births in all 33 health centres within the two districts from nine months before the first implementation until four months after the last implementation.

The intervention encompassed three components. The first component consisted of the distribution of promotional materials encouraging health centre delivery. The second educational component sought to raise awareness among health centre personnel of the importance of the participation of traditional birth attendants and increase knowledge on the appropriate management of obstetric emergencies. The third training component focused on building capacity among health personnel. Main outcomes were number of health centre deliveries and maternal and perinatal morbidity; and perinatal mortality. Usual care continued over the control periods. Women, health care professionals and data collection were unblended to the intervention.

Results: There were a total of 24,464 deliveries over the study period. Health centre deliveries per 100 live births showed an overall increase over the study period, although the adjusted (for secular trends and clustering) relative risk (aRR) was not statistically significant ((aRR 1.06, [CI: 0.94 - 1.32, p = 0.17]). . Furthermore, maternal morbidity decreased (aRR 0.78 [CI: 0.60 – 1.02, p = 0.07]), as well as perinatal morbidity (aRR 0.65 [CI: 0.55 - 1.15, p = 0.12]) and mortality (aRR 0.86 [CI: 0.65 - 1.29, p = 0.29]).

Conclusions: This study found no statistically significant effect of an integrated approach to promote health centre delivery. The intervention holds some promise for decreasing maternal, perinatal morbidity and mortality.

Trial registration: ClinicalTrials.gov, NCTXXXXX; ethical approval: National Institutional Review Board

Table S2 Example of a baseline table by control and intervention conditions taken from the DAVE Trial

Table 1 Sample characteristics by trial mode

Variable	Trial mode	
	Intervention (N = 4693)	Control (N = 5719)
Male, %	39.8	38.0
Age in years, mean (SD)	58.7 (15.3)	58.1 (15.3)
Education		
16 and under, %	36.5	38.1
17/18, %	25.8	26.3
19 and over, %	37.7	35.6
Car ownership		
No car	3.9	4.4
One car	37.8	39.2
Two or more cars	58.3	56.4
Indices of multiple deprivation score (quintiles, %)		
1 (lowest)	25.7	21.3
2	20.9	16.8
3	19.8	19.2
4	17.8	20.4
5 (highest)	15.8	22.2

Taken from Table 1 in 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 Jul 18

Table S3 Example of a baseline table by allocated sequence taken from the Depression Management Trial

	Group 1	Group 2	Group 3	Group 4	Group 5	Total
Total units	7	6	7	6	7	33
Dementia units	4	3	2	4	3	16
Nursing staff per unit	26·1 (14·5)	28·3 (12·5)	27·3 (7·1)	24·7 (9·5)	25·1 (9·3)	26·3 (10·3)
Residents per unit	28·1 (15·1)	30·6 (11·6)	26·7 (5·7)	27·3 (5·3)	25·4 (7·4)	27·6 (9·6)
Residents enrolled at baseline per unit*	15·4 (5·4)	16·0 (8·0)	15·3 (4·3)	16·8 (6·8)	13·0 (5·0)	15·2 (5·2)
Proportion enrolled	60·1% (16·6)	52·5% (17·7)	58·2% (13·4)	59·3% (17·1)	51·4% (13·7)	56·3% (15·1)
Residents enrolled at baseline†	108	96	107	101	91‡	503
In dementia units	55	65	27	79	43‡	269
Age (years)	84·1 (1·4)	78·0 (8·6)	81·4 (2·5)	80·7 (3·8)	78·0 (10·7)	80·5 (6·5)
Women§	73 (70·6% [15·6])	64 (67·0% [16·9])	70 (65·9% [11·1])	67 (64·8% [9·3])	64 (71·7% [13·0])	338 (68·2% [12·9])
Score on mini-mental state examination	14·2 (7·5; n=82)	13·3 (6·8; n=84)	14·9 (6·0; n=78)	14·0 (7·4; n=57)	15·7 (8·8; n=79)	14·5 (6·9; n=380)
With CSDD depression§	39 (57·1% [23·8]; n=74)	49 (56·4% [18·7]; n=78)	44 (54·0% [11·8]; n=84)	53 (50·9% [16·5]; n=96)	41 (47·9% [23·9]; n=86)	226 (53·2% [18·7]; n=418)
With CSDD severe depression§	20 (30·0% [18·1]; n=74)	22 (20·6% [17·2]; n=78)	19 (24·5% [14·6]; n=84)	21 (19·3% [11·0]; n=96)	19 (21·1% [14·5]; n=86)	101 (23·3% [14·9]; n=418)
CSDD score	9·7 (2·8; n=74)	8·9 (2·4; n=78)	8·7 (2·0; n=84)	7·7 (1·8; n=96)	8·4 (2·3; n=86)	8·7 (2·2; n=418)
With GDS8 depression§	12 (19·4% [17·1]; n=51)	19 (46·7% [32·7]; n=33)	14 (34·9% [26·8]; n=47)	11 (49·1% [44·8]; n=27)	14 (28·2% [25·8]; n=42)	70 (34·9% [30·3]; n=200)
With GDS8 severe depression§	3 (5·7% [9·8]; n=51)	11 (26·5% [20·9]; n=33)	8 (22·2% [31·3]; n=47)	9 (38·9% [37·5]; n=27)	7 (15·3% [16·6]; n=42)	38 (21·1% [25·8%]; n=200)
GDS8 score	1·2 (0·9; n=51)	2·4 (1·6; n=33)	2·1 (1·5; n=47)	3·1 (2·7; n=27)	1·7 (1·2; n=42)	2·1 (1·7; n=200)
Quality-of-life score	67·1 (11·3; n=55)	61·7 (9·2; n=61)	60·6 (10·2; n=67)	65·4 (11·0; n=46)	69·0 (12·9; n=50)	64·8 (10·9; n=279)
Residents enrolled after baseline†	74	52	59	59	46	290
In dementia units	43	16	13	40	22	134
Age (years)	82·1 (4·9)	80·7 (12·5)	78·8 (4·2)	81·2 (2·9)	75·9 (12·2)	79·7 (8·2)
Women§	57 (79·2% [16·4])	32 (54·7% [32·2])	33 (49·7% [29·6])	42 (69·4% [6·3])	28 (65·5% [28·9])	192 (63·8 [25·6])
Score on mini-mental state examination	15·5 (6·0; n=57)	16·4 (5·4; n=43)	15·2 (4·5; n=51)	13·6 (5·9; n=42)	18·4 (7·0; n=33)	15·9 (5·7; n=226)
With CSDD depression§	20 (40·5% [31·2]; n=71)	15 (29·9% [20·2]; n=48)	21 (39·4% [19·9]; n=58)	33 (53·9% [30·4]; n=54)	13 (26·6% [18·9]; n=45)	102 (37·8% [24·9]; n=276)
With CSDD severe depression§	9 (26·7% [34·6]; n=71)	7 (13·1% [16·2]; n=48)	4 (6·2% [8·7]; n=58)	17 (27·9% [16·6]; n=54)	3 (10·1% [14·0]; n=45)	40 (16·6% [21·0]; n=276)
CSDD score	8·3 (3·9; n=71)	5·5 (3·1; n=48)	6·4 (1·5; n=58)	9·2 (2·8; n=54)	5·9 (2·6; n=45)	7·0 (3·1; n=276)
With GDS8 depression§	7 (14·0% [21·9]; n=30)	7 (19·7% [24·5]; n=23)	13 (42·1% [11·7]; n=31)	10 (45·8% [26·7] n=18)	14 (42·0% [37·1]; n=28)	51 (33·7% [27·8]; n=130)
With GDS8 severe depression§	3 (7·3% [10·1]; n=30)	6 (17·6% [24·5]; n=23	7 (19·4% [17·3]; n=31)	6 (30·6% [24·5]; n=18)	8 (26·4% [36·1]; n=28)	30 (20·9% [24·6]; n=130)
GDS8 score	0·8 (1·0; n=30)	1·4 (1·5; n=23)	2·4 (0·4; n=31)	2·8 (1·7; n=18)	2·6 (1·9; n=28)	2·1 (1·5; n=130)
Quality-of-life score	64·4 (9·0; n=39)	61·6 (14·9; n=33)	62·0 (10·3; n=39)	63·6 (18·7; n=30)	64·1 (12·4; n=29)	63·1 (12·4; n=170)

Data are n, mean (SD), % (SD), or n (%, SD). Mean, SDs, and percentages calculated at the level of units. CSDD depression was defined as a CSDD score >7. CSDD severe depression was defined as a CSDD score >11. GDS8 depression was defined as a GDS8 score >2. GDS8 severe depression was defined as a GDS8 score >4. Quality of life was assessed with a visual analogue scale of Euroqol-5D. CSDD=Cornell scale for depression in dementia. GDS8=eight-item geriatric depression scale. *Residents were recruited when informed consent had been obtained. †Residents were deemed to be enrolled when they had been assessed during at least one assessment. ‡One dementia unit with 15 residents was enrolled in group 5 after baseline; data for these residents at time of unit inclusion is reported as at baseline. §Percentages are prevalence after adjustment for clustering.

Table 1: **Baseline characteristics**

Note that in this trial the authors have used the phrase “group” to refer to what we mean by “sequence of treatments”

Taken from Table 1 in the Depression Management Trial: Leontjevas R, Gerritsen DL, Smalbrugge M, Teerenstra S, Vernooij-Dassen MJ, Koopmans RT. A structural multidisciplinary approach t depression management in nursing-home residents: a multicentre, stepped-wedge cluster-randomised trial. Lancet. 2013 Jun 29;381(9885):2255-64.

References

[Azar 2015] Azar M, Riehm KE, McKay D, Thombs BD. Transparency of Outcome Reporting and Trial Registration of Randomized Controlled Trials. PLOS ONE. 2015 Nov 18;10:e0142894.

[Baio 2015] Baio G, Copas A, Ambler G, Hargreaves J, Beard E, Omar RZ. Sample size calculation for a stepped wedge trial. Trials. 2015 Aug 17;16:354.

[Barker 2016] 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 Jun 6;16:69.

[Beard 2015] 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 Aug 17;16:353.

[Begg 1996] Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996 Aug 28;276(8):637-9.

[Brown 2006] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol. 2006 Nov 8;6:54.

[Caille 2016] Caille A, Kerry S, Tavernier E, Leyrat C, Eldridge S, Giraudeau B. Timeline cluster: a graphical tool to identify risk of bias in cluster randomised trials. BMJ. 2016 Aug 16;354:i4291.

[Campbell 2004] Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ. 2004 Mar 18;328:702.

[Campbell 2012] Campbell MK, Piaggio G, Elbourne DR, Altman DG; for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012 Sep 4;345:e5661.

[Copas 2015] 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 Aug 17;16:352.

[Cook 2017] Cook JA, Julious SA, Sones W, Rothwell JC, Ramsay CR, Hampson LV, et al. Choosing the target difference ('effect size') for a randomised controlled trial - DELTA(2) guidance protocol. Trials. 2017 Jun 12;18(1):271.

[Davey 2015] Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, et al. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. Trials. 2015 Aug 17;16:358.

[Doussea 2016] Doussau A, Grady C. Deciphering assumptions about stepped wedge designs: the case of Ebola vaccine research. J Med Ethics. 2016 Dec 1;42(12):797-804.

[Edwards 1999] Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ. 1999 May 22;318(7195):1407-9.

[Eldridge 2005] Eldridge SM, Ashby D, Feder GS. Informed patient consent to participation in cluster randomized trials: an empirical exploration of trials in primary care. Clin Trials. 2005 Apr 1;2(2):91-8.

[Eldridge 2009] Eldridge SM, Ukoumunne OC, Carlin JB. The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions. Int Stat Rev. 2009 Oct 29;77(3):378-94.

[Eldridge 2016] Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud. 2016 Oct 21;2:64.

[Girling 2016] Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. Stat Med. 2016 Jun 15;35(13):2149-66.

[Grayling 2007] Grayling MJ, Wason JM, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. Trials. 2017 Jan 21;18(1):33.

[Haines 2017] Haines TP, Hemming K. Stepped-wedge cluster-randomised trials: level of evidence, feasibility and reporting. J Physiother. 2018 Jan;64(1):63-66. doi: 10.1016/j.jphys.2017.11.008. Epub 2017 Dec 27.

[Hargreaves 2015] Hargreaves JR, Copas AJ, Beard E, Osrin D, Lewis JJ, Davey C, et al. Five questions to consider before conducting a stepped wedge trial. Trials. 2015 Aug 17;16:350.

[Hayes 1999] Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999 Apr 1;28(2):319-26.

[Hemming 2014] Hemming K, Girling A. A menu driven facility for sample size for power and detectable difference calculations in stepped wedge randomised trials. Stata J. 2014;14(2):363-80.

[Hemming 2015] Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015 Feb 6;350:h391.380.

[Hemming 2015b] Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med. 2015 Jan 30;34(2):181-96.

[Hemming 2015c] Hemming K, Girling AJ, Haines T, Lilford, R. Protocol: Consort extension to stepped wedge cluster randomised controlled trials. Equator network. http://www.equator-network.org/wp-content/uploads/2009/02/Consort-SW-Protocol-V1.pdf.

[Hemming 2016] Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. J Clin Epidemiol. 2016 Jan;69:137-46.

[Hemming 2017] Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. Trials. 2017 Mar 4;18(1):101. doi: 10.1186/s13063-017-1833-7.

[Higgins 2016] Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie J, Boutron I, Welch V (editors). Cochrane Methods. Cochrane Database of Systematic Reviews. 2016;10(Suppl 1).

[Hoenig 2001] Hoenig JM, Heisey DM. The abuse of power. Am Stat. 2001;55(1):19-24.

[Hoffmann 2014] Hoffmann T, Glasziou P, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.

[Hooper 2015] Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. BMJ. 2015 Jun 8;350:h2925.

[Hooper 2016] Hooper R, Teerenstra S, de Hoop E, Eldridge S. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Stat Med. 2016 Nov 20;35(26):4718-28.

[Hopewell 2008] Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. Lancet. 2008 Jan 26;371(9609):281-3.

[Hughes 2015] Hughes JP, Granston TS, Heagerty PJ. Current issues in the design and analysis of stepped wedge trials. Contemp Clin Trials. 2015 Nov;45(Pt A):55-60.

[Hussey 2007] Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials. 2007 Feb;28(2):182-91.

[ICMJE] International Committee of Medical Journal Editors [http://www.icmje.org/]. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals [16/05/2017] Available from: http://www.ICMJE.org.

[Ivers 2012] Ivers NM, Halperin IJ, Barnsley J, Grimshaw JM, Shah BR, Tu K, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. Trials. 2012 Aug 1;13:120.

[Kasza 2017] Kasza J, Hemming K, Hooper R, Matthews J, Forbes AB; ANZICS Centre for Outcomes & Resource Evaluation (CORE) Committee. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Stat Methods Med Res. 2017 Jan 1:962280217734981.

[Killeen 2014] Killeen SMDF, Sourallous PM, Hunter IAPF, Hartley JEMDBF, Grady HLOMDF. Registration Rates, Adequacy of Registration, and a Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials Published in Surgery Journals. Ann Surg. 2014:259(1):193-6.

[Kotz 2012] Kotz D, Spigt M, Arts IC, Crutzen R, Viechtbauer W. Researchers should convince policy makers to perform a classic cluster randomized controlled trial instead of a stepped wedge design when an intervention is rolled out. J Clin Epidemiol. 2012 Dec;65(12):1255-6.

[Kristunas 2017] Kristunas CA, Hemming K, Eborall HC, Gray LJ. The use of feasibility studies for stepped-wedge cluster randomised trials: a protocol for a review of impact and scope. BMJ Open. 2017;7:e017290.

[Lawrie 2015] Lawrie J, Carlin JB, Forbes AB. Optimal stepped wedge designs. Stat Probabil Lett. 2015;99:210-4.

[Mathieu 2009] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA. 2009;302(9):977-84.

[Martin 2016] Martin J, Taljaard M, Girling A, Hemming K. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. BMJ Open. 2016 Feb 4;6(2):e010166.

[Martin 2016b] Martin J, Girling A, Nirantharakumar K, Ryan R, Marshall T, Hemming K. Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in UK primary care. Trials. 2016 Aug 15;17:402.

[Martin 2017] Martin J. Advancing knowledge in stepped-wedge cluster randomised trials (Unpublished doctoral thesis). University of Birmingham, UK. 2017.

[McCarney 2007] McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol. 2007 Jul 3;7:30.

[Mdege 2011] 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 Sep;64(9):936-48.

[Moher 2010] Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010 Feb 16;7(2):e1000217.

[Moulton 2007] Moulton LH, Golub JE, Durovni B, Cavalcante SC, Pacheco AG, Saraceni V, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention. Clin Trials. 2007;4(2):190-9.

[Pedroza 2016] Pedroza C, Thanh Truong VT. Performance of models for estimating absolute risk difference in multicenter trials with binary outcome. BMC Med Res Methodol. 2016 Aug 30;16(1):113.

[Piaggio] Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012 Dec 26;308(24):2594-604.

[Prost 2015] 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 Aug 17;16:351.

[Taljaard 2013] Taljaard M, Weijer C, Grimshaw JM, Eccles MP; Ottawa Ethics of Cluster Randomised Trials Consensus Group. The Ottawa Statement on the ethical design and conduct of cluster randomised trials: precis for researchers and research ethics committees. BMJ. 2013 May 9;346:f2838.

[Taljaard 2016] Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. Clin Trials. 2016 Aug;13(4):459-63.

[Taljaard 2017] Taljaard M, Hemming K, Shah L, Giraudeau B, Grimshaw JM, Weijer C. Inadequacy of ethical conduct and reporting of stepped wedge cluster randomized trials: Results from a systematic review. Clin Trials. 2017 Aug;14(4):333-341.

[Thompson 2017] 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 Oct 15;36(23):3670-3682.

[Rutterford 2015] 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 Jun;68(6):716-23.

[Rennie 2001] Rennie D. CONSORT revised--improving the reporting of randomized trials. JAMA. 2001 Apr 18;285(15):2006-7.

[Senn 1994] Senn S. Testing for baseline balance in clinical trials. Stat Med. 1994 Sep 15;13(17):1715-26.

[Schulz 2010] Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.

[Shadish 2002] Shadish WR, Cook TD, Campbell D T. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Wadsworth Cangage Learning. 2002.

[Ukoumunne 2008] Ukoumunne OC, Forbes AB, Carlin JB, Gulliford MC. Comparison of the risk difference, risk ratio and odds ratio scales for quantifying the unadjusted intervention effect in cluster randomized trials. Stat Med. 2008 Nov 10;27(25):5143-55.

[Wang 2017] Wang M, Jin Y, Hu ZJ, Thabane A, Dennis B, Gajic-Veljanoski O, et al. The reporting quality of abstracts of stepped wedge randomized trials is suboptimal: A systematic survey of the literature. Contemp Clin Trials Comm. 2017 Dec;8:1-10.

[Wetering 2012] van de Wetering FT, Scholten RJPM, Haring T, Clarke M, Hooft L. Trial Registration Numbers Are Underreported in Biomedical Publications. PLOS ONE. 2012;7(11):e49599.

[UN 1966] United Nations. International Covenant on Civil and Political Rights. 1966.

[Zhan 2017] Zhan Z, de Bock GH, van den Heuvel ER. Statistical methods for unidirectional switch designs: Past, present, and future. Stat Methods Med Res. 2017 Jan 1.

[Zou 2005] Zou GY, Donner A, Klar N. Group sequential methods for cluster randomization trials with binary outcomes. Clin Trials. 2005;2(6):479-87.

[Zwarenstein 2008] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ. 2008;337:a2390.

References Key Examples

[DAVE Trial] 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 Jul 18;11:94.

[DAVE Trial protocol] Solomon E, Rees T, Ukoumunne OC, Hillsdon M. The Devon Active Villages Evaluation (DAVE) trial: study protocol of a stepped wedge cluster randomised trial of a community-level physical activity intervention in rural southwest England. BMC Public Health. 2012 Aug 1;12:581.

[Depression Management Trial] Leontjevas R, Gerritsen DL, Smalbrugge M, Teerenstra S, Vernooij-Dassen MJ, Koopmans RT. A structural multidisciplinary approach to depression management in nursing-home residents: a multicentre, stepped-wedge cluster-randomised trial. Lancet. 2013 Jun 29;381(9885):2255-64.

[FallDem Trial] Reuther S, Holle D, Buscher I, Dortmann O, Müller R, Bartholomeyczik S, et al. Effect evaluation of two types of dementia-specific case conferences in German nursing homes (FallDem) using a stepped-wedge design: study protocol for a randomized controlled trial. Trials. 2014 Aug 12;15:319.

[FIT Trial] Fuller C, Michie S, Savage J, McAteer J, Besser S, Charlett A, Hayward A, Cookson BD, Cooper BS, Duckworth G, Jeanes A, Roberts J, Teare L, Stone S. The Feedback Intervention Trial (FIT)--improving hand-hygiene compliance in UK healthcare workers: a stepped wedge cluster randomised controlled trial. PLOS ONE. 2012;7(10):e41617.

[Long-live Mothers Trial] Kestler E. , Ambrosio G. , Hemming K., Hughes J. Matute J. Moreno M. and Walker D. Scaling up an integrated approach to improve care during delivery in northern Guatemala: a stepped-wedge cluster randomized trial Submitted.

[RegisterNow-1 Trial] Li A, Garg A, Prakash V, Grimshaw J, Taljaard M, Mitchell J, Matti D, Linklater S, Naylor K, Dixon S, Faulds C, Bevan R, Getchell L, Knoll G, Kim J, Sontrop J, Bjerre L, Tong A, Presseau J. Promoting deceased organ and tissue donation registration in family physician waiting rooms (RegisterNow-1 Trial): Study Protocol for a pragmatic stepped-wedge cluster randomized controlled registry trial. Trials 2017; 18:610. ~~Li AH, Garg AX, Prakash V, Mitchell J, Grimshaw J, Taljaard M, et al. Promoting deceased organ and tissue donation registration in family physician waiting rooms: Protocol for a pragmatic stepped-wedge cluster-randomized controlled registry trial (RegisterNow-1 Trial). Submitted to Trials.~~

[REMAIN Trial protocol] Foot H, Freeman C, Hemming K, Scott I, Coombes ID, Williams ID, et al. Reducing Medical Admissions into Hospital through Optimising Medicines (REMAIN HOME) Study: protocol for a stepped-wedge, cluster-randomised trial. BMJ Open. 2017 Apr 13;7(4):e015301.

[Riverbank Filtration Trial] McGuinness SL, O'Toole JE, Boving TB, Forbes AB, Sinclair M, Gautam SK, et al. Protocol for a cluster randomised stepped wedge trial assessing the impact of a community-

Formatted: Font: Not Bold

Formatted: Font: Not Italic

level hygiene intervention and a water intervention using riverbank filtration technology on diarrhoeal prevalence in India. BMJ Open. 2017 Mar 17;7(3):e015036.

[SBP Trial] Mhurchu CN, Gorton D, Turley M, Jiang Y, Michie J, Maddison R, et al. Effects of a free school breakfast programme (SBP) on children's attendance, academic achievement and short-term hunger: results from a stepped-wedge, cluster randomised controlled trial. J Epidemiol Community Health. 2013 Mar;67(3):257-64.

[SBP Trial Protocol] Ni Mhurchu C, Turley M, Gorton D, Jiang Y, Michie J, Maddison R, et al. Effects of a free school breakfast programme on school attendance, achievement, psychosocial function, and nutrition: a stepped wedge cluster randomised trial. BMC Public Health. 2010 Nov 29;10:738.

[SMC Trial] NDiaye JL, Cissé B, Ba EH, Gomis JF, Ndour CT, Molez JF, et al. 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 Oct 20;11(10):e0162563.

[Surgical Checklist Trial] Haugen AS, Søfteland E, Almeland SK, Sevdalis N, Vonen B, Eide GE, et al. Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. Ann Surg. 2015 May;261(5):821-8.

[Targeted Case Finding Trial] Hemming K, Ryan R, Gill P, Westerby P, Jolly K, Marshall T. Targeted case finding in the prevention of cardiovascular disease: a stepped wedge cluster randomised controlled trial. Br J Gen Pract. 2016 Oct;66(651):e758-67.

[THRio Trial Protocol] Moulton L, Golub JE, Durovni B, Cavalcante SC, Pacheco AG, Saraceni V, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention. Clin Trials. 2007;4:190-9.

Reporting of ~~The CONSORT extension for~~ Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration

K Hemming[1], M Taljaard[2], JE McKenzie[3], R Hooper[4], A Copas[5], JA Thompson[5 6], M Dixon-Woods[7], A Aldcroft[8], A Doussau[9], M Grayling[10], C Kristunas[11], CE Goldstein[12], MK Campbell[13], A Girling[14], S Eldridge[15], MJ Campbell[16], RJ Lilford[17], C Weijer[18], A Forbes[19], JM Grimshaw[2 20]

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. k.hemming@bham.ac.uk;

[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, 1053 Carling Avenue, Ottawa, Ontario, Canada; and School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Canada. mtaljaard@ohri.ca;

[3] School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. Joanne.mckenzie@monsh.edu;

[4]Pragmatic Clinical Trials Unit, Centre for Primary Care & Public Health, Queen Mary University of London, London, UK. r.l.hooper@qmul.ac.uk;

[5]London Hub for Trials Methodology Research, MRC Clinical Trials Unit at University College London, London, UK. a.copas@ucl.ac.uk;

[6]Department for Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. jennifer.thompson@lshtm.ac.uk;

[7]THIS Institute~~Cambridge Centre for Health Services Research, Department of Public Health and Primary Care~~, University of Cambridge, Cambridge Biomedical Campus, Bay 13 Clifford Allbutt Building, Cambridge CB2 OAH. md753@medschl.cam.ac.uk;

[8]BMJ Publishing Group, London, UK. aaldcroft@bmj.com

[9]Biomedical Ethics Unit, McGill University School of Medicine, Montreal, Canada. Adelaide.doussau@mail.mcgill.ca;

[10]MRC Biostatistics Unit, Cambridge, UK. michael.grayling@mrc-bsu.cam.ac.uk;

[11]Department of Health Sciences, University of Leicester, Leicester, UK. Cak21@le.ac.uk;

[12]Rotman Institute of Philosophy, Western University, London, Canada. cgoldst2@uwo.ca;

[13] Health Services Research Unit, University of Aberdeen, Aberdeen, UK. m.k.campbell@abdn.ac.uk;

[14]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. a.j.girling@bham.ac.uk;

[15]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK. s.eldridge@qmul.ac.uk;

[16]ScHARR, University of Sheffield, Sheffield, UK. m.j.campbell@sheffield.ac.uk;

[14]University of Warwick, Coventry, UK. R.J.Lilford@warwick.ac.uk;

[18]Rotman Institute of Philosophy, Western University, London, Canada. cweijer@uwo.ca;

[19]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. Andrew.Forbes@monash.edu;

Formatted: Default Paragraph Font, Font color: Text 1

Formatted: Font color: Text 1

[20] Department of Medicine University of Ottawa, Ottawa, Canada. jgrimshaw@ohri.ca.

Acknowledgements

With acknowledgement to those who participated in the Delphi survey and Peter Chilton who provided administrative support.

Author contributions

KH led the development of the project, the Delphi survey, the consensus meeting, drafting of the items; and wrote the first draft of the paper. MT, JG, AF, CW and JM made a substantial contribution to all stages of the project. CW and MT gave insight into the ethical aspects of the project. KH, MT, JM, CW and AF contributed to the development of the items. SE and MJC gave critical insights into reporting guidelines. AF and JMG provided project leadership and guidance. JMG facilitated the consensus meeting. RL provided critical insight into the early stages of the project. All authors participated in the consensus meeting and commented on the draft paper.

Funding

This research was funded by the Australian National Health and Medical Research Council (NHMRC) project grant (1108283) and also partly funded by the UK NIHR Collaborations for Leadership in Applied Health Research and Care West Midlands initiative. Mary Dixon-Woods is funded by a Welcome Trust Senior Investigator award WT097899. Jennifer A Thompson is funded by the Medical Research Council Network of Hubs for Trials Methodology Research (MR/L004933/1-P27). Jeremy Grimshaw holds a Canada Research Chair in Health Knowledge Transfer and Uptake. Charles Weijer holds a Canada Research Chair. Joanne E McKenzie holds an NHMRC Australian Public Health Fellowship (1072366). Karla Hemming holds an NIHR Senior Research Fellowship (SRF-2017-002).

Competing Interests

We have read and understood the BMJ Group policy on declaration of interests and declare the following interests: none.

Summary

This document presents the Consolidated Standards Of Reporting Trials (CONSORT) extension for the stepped-wedge cluster randomised trial (SW-CRT). The SW-CRT involves randomisation of clusters to different sequences that dictate the order (or timing) at which each cluster will switch to the intervention condition. The development of this statement was motivated by the unique design characteristics of this study ~~e stepped wedge~~ design, including the need to allow for time effects and because the design is increasingly being used. The guideline was developed using a Delphi survey and consensus meeting; and is informed by the CONSORT statements for individually and cluster randomised trials. Reporting items along with explanations and examples are provided. We include a glossary of terms, and explore the key properties of the SW-CRT which require special consideration in their reporting.

Introduction

The CONSORT (Consolidated Standards Of Reporting Trials) statement, initially published in 1996 and updated in 2001 and 2010, outlines essential items to be reported in a parallel arm individually randomised trial [Begg 1996; Rennie 2001; Schulz 2010]. The CONSORT extension for cluster randomised trials, initially published in 2004 and updated in 2012, extended this guidance for trials in which groups of individuals (clusters – for a full glossary of terms see Table 1) are randomised to different treatment conditions [Campbell 2004; Campbell 2012]. In recent years, a novel type of cluster randomized design - the stepped-wedge cluster randomised trial (SW-CRT) - has become increasingly popular [Brown 2006; Mdege 2011, Martin 2017]. The SW-CRT involves randomisation of clusters to different sequences. These sequences dictate the order (or timing) with which each cluster will switch to the intervention condition.

The basic components of the design, as well as illustrative examples of studies which have used this design, have been described previously [Hemming 2015]. The unit of randomisation in these trials is the cluster with clusters (or groups of clusters) allocated to different sequences (as opposed to different "arms" in a parallel trial). These sequences dictate the number of time periods spent in the control condition and the number of time periods in the intervention condition. In Figure 1, for example, there are four clusters allocated to four different sequences. Each cluster contributes data to the analysis from each measurement period. In the example in Figure 1 there are five measurement periods. The point at which a cluster switches to the intervention condition is called a "step". Sometimes a transition period is built into the design, during which the intervention is implemented in the cluster.

This design has numerous methodological complexities, including potential confounding with time [Hemming 2017]; changes in correlation structures over time [Girling 2016; Hooper 2016; Kasza 2017]; the possibility of within cluster contamination over time [Copas 2015]; the possibility of time varying treatment effects [Davey 2015, Hemming 2017]; and different design variations [Prost 2015; Hargreaves 2015], all of which increase the complexity of reporting [Hemming 2015]. Perhaps unsurprisingly, systematic reviews examining the adequacy of reporting of SW-CRTs have revealed numerous inadequacies, including absence of essential details of the design, inconsistent use of terminology [Brown 2006; Mdege 2011; Martin 2016; Grayling 2017; Taljaard 2017]; frequent lack of clarity in reporting of adjustment for time effects [Hemming 2017; Martin 2017]; as well as frequent failure to report ethical review and trial registration [Taljaard 2017]. These findings suggest there is a need for a specific reporting guideline for this trial design. Here we report the results of a consensus process to develop an extension to the CONSORT statement for use with SW-CRTs. The ultimate goal of this extension is to improve the standards of reporting of this important and increasingly used research design.

Scope of this statement

This reporting statement should be followed when reporting results from any SW-CRT. In line with other CONSORT statements this guideline includes the minimum set of items that should be reported; it is not intended to be a comprehensive list of all possible items that could be reported.

A wide variety of terminology has been used to describe aspects of the SW-CRT design. For the purpose of this reporting statement, the key components of the design are defined in Figure 1 and a glossary of terms is provided in Table 1. ~~with cin in to the analysis from implemented~~ Generally, SW-CRTs ~~stepped wedge trials~~ have a minimum of 3 sequences. Trials with ~~2~~two sequences and three~~3~~ periods, for example, a two-arm ~~before and after~~ cluster randomi~~z~~sed trial~~s~~ in which both arms are initially observed under the control condition and in addition~~,~~ ~~in which~~ the control arm adopts the intervention during a third measurement period ~~and there is a third measurement period in the intervention condition in both arms~~ might also technically be considered a SW-CRT~~stepped wedge trial~~. The statement was developed for comparisons of two treatment conditions. So as to take a broader perspective on the range of designs that can be included, we are not restricting our definition to designs with all

clusters initiating in the control condition and ending up in the intervention condition [Hooper 2016], so include recent proposed dog-leg designs and variations [Hooper 2015].

A wide variety of terminology has been used to describe aspects of the SW-CRT design. For the purpose of this reporting statement, the key components of the design are defined in Figure 1 and a glossary of terms is provided in Table 1.

Extending the CONSORT statement to SW-CRTs

We developed this extension using methods recommended for developing reporting guidelines [Moher 2010]. We registered our protocol on the EQUATOR website in July 2015 [Hemming 2015c] and identified relevant and related reporting guidelines. We conducted several systematic reviews of published SW-CRTs examining aspects of reporting and methodological conduct and undertook a consensus process.

Results from systematic reviews examining SW-CRT methods and reporting

We conducted several systematic reviews in advance of the consensus process [Martin 2016; Taljaard 2017; Grayling 2017; Martin 2017]. Martin et al. (2016) found that the SW-CRT is increasingly being used and that the majority of trials are conducted in advanced economies and in healthcare settings; although a significant minority are conducted in lower middle income settings; with most trials having less than 20 clusters and a smaller number of time periods [Martin 2016].

Reviews of the quality of reporting of sample size and analysis methods revealed incomplete or inadequate reporting overall, and specifically, lack of reporting of how time effects and extended correlation structures were incorporated both at the design and analysis stages [Davey 2015; Martin 2016; Grayling 2017; Martin 2017]. Reviews of the ethical conduct and reporting revealed that many SW-CRTs do not report research ethics review; do not clearly identify from whom and for what consent was obtained; and a significant number do not pre-register with a trial registration database [Taljaard 2017]. Reviews of the methodological literature have identified several key aspects of the SW-CRT which are associated with bias [Barker 2016; Martin 2017]. Clear reporting of these aspects is essential to facilitate interpretation of trial results in published reports.

Firstly, time is a potential confounder in a SW-CRT and requires special consideration both at the design and analysis stage [Hughes 2007; Hemming 2017]. Secondly, as the SW-CRT is a longitudinal and clustered study, correlation structures are more complex than those of a parallel CRT carried out at a single cross-section in time [Hooper 2016]. Thirdly, some SW-CRTs are at risk of within-cluster contamination. Within-cluster contamination can arise either when outcomes in the intervention condition are obtained from participants who are yet to be exposed to the intervention, or alternatively, when outcome assessments in the control condition are from participants already exposed to the intervention [Copas 2015]. Contamination arising from observations yet to be fully exposed to the intervention condition can be allowed for by building in transition periods into the design; or by modelling these effects (referred to as lag effects) [Hughes 2015]. Interactions between time and treatment can also arise. These time varying effects are more likely to arise when the intervention is not continuously delivered, does not create a permanent change, or where its impact might waiine or grow over time [Davey 2015].

These complexities differ according to the many different ways that a SW-CRT can be conducted, including whether the same or different participants are repeatedly assessed, whether participants are continuously recruited and the duration of their exposure, and whether a complete enumeration of the cluster is taken [Hemming 2015; Copas 2015]. With practical and ethical considerations also in play, the adoption of this design requires careful justification [Prost 2015; Doussau 2016]. A summary of key methodological issues which need extra consideration when reporting a SW-CRT is presented in Table 2.

Consensus process

Members of the working group (KH, MT, JEM, AF, CW, JG) identified items from the original CONSORT statement which required modification; considered whether the modification used in the cluster extension was appropriate; and if not, proposed a modified version for the item. In a modified Delphi process (December 2016), we invited 64 subject experts to consider, rate and comment on the proposed modifications of whom 42 completed the survey. We summarised responses from the survey and circulated a second draft of the proposed modifications in advance of a one-day consensus meeting (Liverpool May 2017). The CONSORT stepped-wedge consensus group (20 people in total all listed as authors of this statement) consisted of members of the working group and those with expertise in trial design, journal editors (BMJ Open, Trials, Clinical Trials, and BMJ Quality and Safety Improvement), ethicists, statisticians, methodologists, and developers of reporting guidelines (cluster trials, pilot and feasibility trials and equity trials). At the meeting, proposed wording, examples and elaboration text were discussed and amended. The proposed final wording was then circulated~~,~~ and final comments incorporated.

The CONSORT extension for Stepped-Wedge Cluster Randomised Trials

A checklist detailing the 26 items to be reported in the publication of a SW-CRT is presented in Table 3. Some items have not been modified from the original CONSORT statement, some are modified, and some are new. Similar to the CONSORT extension for cluster trials, Item 10 (Implementation of randomisation) has been replaced by Items 10a, 10b and 10c. In recognition of the under-reporting of key ethical aspects of these trials, a new item on Research Ethics Review has been added as Item 26 (as was added to the CONSORT extension for pilot and feasibility studies [Eldridge 2016]). For ease of interpretation in the elaboration that follows, we provide the original CONSORT wording, the wording of the CONSORT extension for cluster randomised trials, as well as the wording for the SW-CRT extension. Table 4 summarises key changes to the original CONSORT statement and substantial deviations from the CONSORT extension for cluster randomised trials. We have provided examples and explanations for most items. Where the item has not been modified or the modification is only minor, readers are referred to the original statements for full explanation and elaboration [Schulz 2010; Campbell 2012]. For some items, which have not been modified, an example or explanation has been provided where this item raises specific nuances under the- SW-CRT~~stepped-wedge design~~. Given differences in terminology used to describe the SW-CRT and the significant number of modified items, the items in this statement have been written in such a way so as to replace the original CONSORT items; and therefore, should not be considered extensions to the original items.

Title and abstract

Item 1a Title

Standard CONSORT item: Identification as a randomised trial in the title.

CONSORT cluster extension: Identification as a cluster randomised trial in the title.

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Identification as a stepped-wedge cluster randomised trial in the title.

Example: "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." [DAVE Trial]

Explanation: One reason for including the type of study design in the title is to facilitate accurate identification of relevant studies in systematic reviews. A wide variety of different terminology is currently used to describe the SW-CRT. These include the "multiple-period baseline design" and the "wait list design" (although not every multiple-period baseline design and wait list design will be a SW-CRT). Adoption of a single term will improve the

identification of these studies and differentiate studies which are not SW-CRTs. Reporting of parallel cluster randomised trials (CRT) improved with the adoption of the single term "cluster" rather than the mix of terms (such as "group randomised" or "field trial") [Ivers 2011]. It can also be useful to report any trial acronym in the title, to aid future searches for the study.

Formatted: Not Highlight

Item 1b: Abstract
Standard CONSORT item: Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts).
CONSORT cluster extension: Abstract See Table (not shown).
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs: Structured summary of trial design, methods, results, and conclusions (Table 5).*

For the same rationale as provided in the other CONSORT statements, clear reporting of the trial's objectives, design, methods, main results and conclusions in the abstract is crucial. The primary reason for this is that many readers will base their assessment of the trial from the information available in the abstract [Hopewell 2008]. A review assessing the quality of reporting of abstracts from fully published SW-CRT revealed incomplete reporting of important details [Wang 2017]. A set of items to be reported as a minimum in an abstract of a SW-CRT is included in Table 5. Of some note, the Items recommended to be reported in the abstract results section do not include the summary measures of the outcome under intervention and control conditions, so as to avoid misattributing the unadjusted difference to the treatment effect. A worked example of an abstract according to this template is provided (Table S1, Long-live Mothers Trial).

Introduction

Item 2a: Background
Standard CONSORT: Scientific background and explanation of rationale.
CONSORT cluster extension: Rationale for using a cluster design
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Scientific background. Rationale for using a cluster design and rationale for using a stepped-wedge design.

Example 1 (Scientific background): "In 2008, the World Health Organization (WHO) introduced the Surgical Safety Checklist (SSC) designed to improve consistency of care. The pilot pre-/post evaluation of the WHO SSC across 8 countries worldwide, which found reduced morbidity and mortality after SSC implementation, constituted the first scientific evidence of the WHO SSC effects. A number of subsequent studies to date have reported improved patient outcomes with use of checklists. Furthermore, checklists have also been shown to improve communication, preparedness, teamwork, and safety attitudes—findings that have been corroborated by a recent systematic review. Although checklists are becoming a standard of care in surgery, the strength of the available evidence has been criticized as being low because of (i) predominantly pre /post implementation designs without controls; (ii) lack of evidence on effect on length of stay; and (iii) lack of evidence on any associated cost savings. Randomized controlled trials (RCTs) are required…." [Surgical Checklist Trial]

*Example 2 (Rationale for cluster randomisation and stepped-wedge design): "*A stepped wedge cluster randomised controlled design was chosen following piloting to facilitate roll out of the intervention, …, and prevent contamination and disappointment effects in hospitals not randomised to the intervention." [FIT Trial]

Explanation: The need for any randomised evaluation of an intervention, whether randomising clusters or individuals should be justified. This justification should make reference to the best available evidence for similar interventions. Reasons why current evidence is lacking should be articulated (as in Example 1).

As with any trial design, key aspects of the design should be justified. In the SW-CRT, this justification includes the use of cluster randomisation, the need to roll out the intervention to all clusters (where this is the case), and the need for staggered roll-out of the intervention [Hargreaves 2015]. Justifying cluster randomisation is important because cluster randomisation increases the sample size and this, in turn might expose more participants to interventions of unknown effectiveness. Justifying the need for a staggered roll-out of the intervention using a SW-CRT, as opposed to a simple parallel arm implementation, is important because the SW-CRT is more complicated in its design, analysis, and implementation than the parallel CRT. R~~. R~~isks of bias in the SW-CRT may be higher than in a parallel CRT. For example, secular trends may be of concern in a SW-CRT, but not in a parallel design [Hemming 2017]. Risks of bias arising from identification and recruitment of participants may also be higher because in a SW-CRT it may be more difficult to blind people recruiting participants to the cluster's allocation status. The design is consequently viewed by some as potentially providing a lower level of evidence compared to the parallel CRT [Mdege 2011; Kotz 2012; Haines 2017]).

Some possible justifications for adopting the stepped-wedge design include that the intervention will be rolled out regardless of the research study [Prost 2015], availability of an inadequate number of clusters to achieve the target power in a parallel design [Hemming 2016], to increase statistical efficiency [Lawrie 2015; Girling 2016; Zhan 2017], or to facilitate recruitment when engagement of clusters is only forthcoming on some promise of the intervention (as in Example 2).

Although staggering the roll-out may appeal to researchers with limited resources for delivering the intervention simultaneously, this is not in itself a legitimate argument for a SW-CRT [Hemming 2015b]. Providing the intervention to all clusters might also increase the duration of the study (due to the staggering of the roll-out) and will possibly increase the number of clusters (and patients) exposed to the intervention (due to all clusters receiving the intervention). For these reasons, justifying the need to expose all clusters (where this is the case) to the intervention is important. The cluster cross-over design is a more statistically efficient design than the SW-CRT and it might therefore be important to justify why a unidirectional cross-over design has been chosen. However, in practice the use of the cluster cross-over design is restricted to interventions that can be withdrawn from use, and this largely depends on the type of intervention being evaluated.

Item 2b: Objective
Standard CONSORT item: Specific objectives or hypotheses.
CONSORT cluster extension: Whether objectives pertain to the cluster level, the individual participant level or both.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Specific objectives or hypotheses.

Example: "We report a stepped wedge cluster RCT aimed to evaluate the impact of the WHO SSC *(World Health Organisation Surgical Safety Checklist)* on morbidity, mortality, and length of hospital stay (LOS). We hypothesized a reduction of 30 days' in-hospital morbidity and mortality and subsequent LOS post-Checklist implementation." [Surgical Checklist Trial]

Explanation: Having a clear and succinct set of objectives can help summarise the overarching aims of the study. Specification of the objectives gives clarity about the anticipated effects of the intervention being evaluated (as in Example). Sometimes these effects will be anticipated to be on process outcomes (e.g. systems changes, clinician performance), particularly in trials which target health care providers; other times the intervention might target patients and anticipate effects on clinical outcomes. One specific objective which can be of interest in a SW-CRT is to evaluate the effect of the intervention by timing of implementation (e.g. does the effect of the intervention change as the intervention is perhaps refined over time) or time since intervention implementation (e.g. does the intervention create a permanent effect). Also of relevance is whether the study is to show superiority of the

intervention condition, non-inferiority or equivalence. For non-inferiority or equivalence authors should also ensure reporting according to the CONSORT extension for non-inferiority and equivalence studies [Piaggio 2012].

Methods: Trial design

Item 3a: Trial design
Standard CONSORT item: Description of trial design (such as parallel, factorial) including allocation ratio.
CONSORT cluster extension: Definition of cluster and description of how the design features apply to the clusters.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Description and diagram of trial design including definition of cluster, number of sequences, number of clusters randomised to each sequence, number of periods, duration of time between each step, and whether the participants assessed in different periods are the same people, different people, or a mixture.

Example 1: "During the DAVE study, the intervention will be rolled out sequentially to 128 rural villages (clusters) over four time periods. The evaluation will consist of data collection at five fixed time points (baseline and following each of the four intervention periods)… The intervention will be fully implemented by the end of the trial, with all 128 villages receiving the intervention: 22 first receiving the intervention at period 2, 36 at period 3, 35 at period 4, and 35 at period 5." [Dave Trial Protocol, Figure S1]

Example 2: This study will use a closed cohort stepped wedge cluster randomised design, which involves a sequential crossover of clusters from the control to the intervention arm, so that every cluster begins in the control condition and eventually receives the intervention, with the order of crossover randomly determined. The study will be conducted in four rural villages…At the start of the study period, baseline (T0) demographic and health data will be collected from each consenting household and baseline hygiene education will be provided. …The second (T1) health survey will start 4 weeks after the initiation of piped untreated river water supply to evaluate the impact of hygiene education combined with improved water quantity compared with baseline (T0). RBF-treated water (intervention arm) will then be sequentially introduced to each village in random order at 12-week intervals (T2–T5), with health surveys performed 4 weeks after the implementation of the intervention to assess the additional effects of improved water quality [Riverbank Filtration Trial, Figure 2]

Explanation: The specific details of the design of the SW-CRT have implications for the type of analysis and sample size calculations required.

Information on the number of sequence~~step~~s and the number of clusters randomised to each sequence is the core of the study design and so should be reported. The number of time periods will often (but not always) be one more than the number of steps (as in Example 1). Definition of cluster (as clearly reported in Example 1) and duration of time periods are also crucial. The duration of the first and last periods can sometimes differ from other periods; if so, this should be reported. The number of clusters allocated to each sequence may vary and, if so, this should be reported. ~~Also of relevance is whether the design is to show superiority of the intervention condition, non-inferiority or equivalence.~~

Information on whether the measurements taken in the different time periods are from the same individuals or different individuals is important for both sample size and analysis. In an open cohort design, participants are repeatedly assessed over series of measurement points and participants can join and leave the cohort; in a closed cohort design, new participants cannot join the study; in a cross-sectional design, different participants are assessed at each measurement occasion. Measurements can also take place at one point in time in each period, or can be continuous throughout the period. This issue is covered in more detail under Item 6a (assessments of outcomes).

A diagram of the trial design can efficiently communicate the details. Key points to depict in the design diagram are the timing of the interventions (Item 3a) and the timing of the data collection (Item 6a). In the Riverbank Filtration Trial, key information about the design was reported in a diagram (Figure 2) and the main text (Example 2).

Item 3b: Changes to t~~T~~rial design
Standard CONSORT item: Important changes to methods after trial commencement (such as eligibility criteria), with reasons.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Important changes to methods after trial commencement (such as eligibility criteria), with reasons.
~~No modification suggested.~~

Example: "…delayed Research and Development registration shortened the baseline pre-randomisation phase from twelve months to nine in the first hospitals randomised to the intervention."[FIT Trial]

Explanation: Changes to key features of the design can have important implications for the interpretation of results. Some changes or deviations may be inevitable. Potential changes in the SW-CRT include modification to the duration between steps (perhaps because of study set up delays as in Example). The timing of any changes is important as they may affect some observations / clusters and not others.

Methods: Participants

Item 4a: Participants
Standard CONSORT item: Eligibility criteria for participants.
CONSORT cluster extension: Eligibility criteria for clusters.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Eligibility criteria for clusters and participants.

Example: "Inclusion criteria: Institution level: At least two units of one *(from each)* nursing home must participate in the study, from which at least 30 residents with dementia can be recruited. The care of the residents must predominantly take place in the respective unit. Resident level: Criteria for inclusion are informed consent obtained from people with dementia or their legal representative; diagnosis of dementia based on the medical diagnosis in the charts and a FAST score > 1); residence for at least 14 days in the unit. Staff level: All of the nursing staff working in one of the two participating wards of the nursing home must provide their informed consent." [FallDem Trial]

Explanation: The SW-CRT is a type of cluster randomised trial and as such, has inclusion and exclusion criteria for both clusters and participants. Furthermore there may be multiple levels of participants. For example, clusters may be general practices that include cluster-level participants (e.g. general practitioners) and individual-level participants (e.g. patients). So, in some trials, there may be multiple levels at which inclusion and exclusion criteria apply (as in the Example). Reporting of eligibility criteria is important so that readers can infer how typical or atypical the clusters and participants are of the population at large [Zwarenstein 2008].

Item 4b: ~~Participant~~Settings~~s~~
Standard CONSORT item: Settings and locations where the data were collected.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Settings and locations where the data were collected.
~~No modification suggested.~~

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Readers are referred to the CONSORT statement and its extension to CRTs for examples and explanation [Schulz 2010, Campbell 2012].

Methods: Intervention

Item 5: Intervention
Standard CONSORT item: The interventions for each group with sufficient details to allow replication, including how and when they were actually administered.
CONSORT cluster extension: Whether interventions pertain to the cluster level, the individual participant level or both.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* The intervention and control conditions with sufficient details to allow replication, including- if maintained or repeated~~how and when they were administered~~; whether the intervention was delivered at the level of the cluster, the individual, or both.

Example 1 (Description of the intervention condition): "The intervention involves three key modes of delivery: verbally via reception staff, in paper form with a pamphlet, and electronically via a secure, internet-enabled tablet (see Table *(not provided)* for overview of intervention). First, reception staff will verify the organ donor registration status of patients upon their arrival at the clinic on the provincial health card that patients must provide to receive healthcare services from their family physician. As reception staff already request a patient's health card during their visit, this step is designed to fit within existing work routines rather than increasing any workload. Reception staff will provide patients that have not yet registered with an educational pamphlet including a photo and signature of the physicians in the office and office logos and include messages that directly address identified barriers to donor registration. Second, internet-enabled tablets will be provided in each waiting room to give patients the immediate opportunity to register for organ donation online via a secure provincial website. The location of the materials will be tailored according to the family physician office's preferences." *(further details provided in paper)* [RegisterNow-1 Trial]

Example 2 (Description of control condition): "If the participant's medical centre is in the control phase, they will receive usual care. In Australia, usual care would mean the patient would consult their GP as per normal standards for that practice for a patient discharged from hospital. There will be no pharmacist in the medical centre during the control phase. Medication liaison in the form of a discharge medication record may be provided to patients on discharge from hospital and may be included in the hospital discharge summary to the GP." [REMAIN Trial Protocol]

Example 3 (Unit of delivery is individual): "The intervention comprised a therapeutic dose of AQ (10 mg/kg/day for 3 days) combined with one dose of SP on the first day (25mg sulfamethoxypirazyne and 1.25mg pyrimethamine per kg in 2008, 25mg sulfadoxine, 1.25mg pyrimethamine in 2009–10) administered once per month for the last three months of the malaria transmission season (September-November)." [SMC Trial]

Example 4 (Continuously delivered intervention): "It *(the intervention)* comprised bedside placement of alcohol hand-rub, posters and patient empowerment materials encouraging healthcare workers to clean their hands, plus audit and feedback of hand-hygiene compliance at least once every 6 months." [FIT Trial]

Explanation: Clear reporting of the intervention is essential to allow replication and implementation of successful interventions (Example 1). For interventions demonstrated to have little evidence of benefit, reporting of sufficient detail of the intervention helps to avoid evaluating the same intervention again or to identify what aspects of the intervention could be modified. This is especially important for complex interventions – a common type of intervention evaluated in SW-CRTs. We recommend reporting details of the intervention as per the TiDierR guideline

[Hoffmann 2014]. As per the original CONSORT statement, it is important to describe all treatment conditions being compared. In SW-CRTs the comparator is often "usual care" which should be described in sufficient detail (Example 2). The control condition should be described in a similar level of detail to the intervention condition [Zwarenstein 2008].

Information on whether the intervention is delivered at the level of the cluster or individual (or perhaps both) is important as it allows identification of whether individuals can avoid the intervention. For example, an intervention which is delivered at the level of the cluster will often mean that it is delivered to all individuals within that cluster (Example 1). In the SMC Trial the intervention was delivered directly to the individual (Example 3). This information is also important as it can inform the degree of penetration of the intervention and it can also be helpful in eliciting what consent procedures should be in place (Items 10c and 26).

In a SW-CRT it is important to be clear about whether the intervention is expected to create an effect that is expected to be immediate (or delayed); and whether the anticipated effects of the intervention are expected to be sustained. This is important because the observations contributing to the analysis will consist of a mixture of observations collected immediately after roll-out of the intervention; and observations collected some time post roll-out.

The effect of any intervention can be delayed; for example, due to a learning effect, one may need to allow for a delay before the effect is fully realised (this might ~~perhaps~~ be the case in Example 4). In these situations a transition period might be incorporated into the design. Furthermore the anticipated effects of the intervention might be sustained (in which case an intervention might be designed to have a one-off delivery, as in Example 1) or expected to decay (in which case an intervention might be designed to have repeated delivery, as in Example 4). In some SW-CRTs the exact form of the intervention may evolve over time; reporting this information allows assessment of the level of standardisation of the intervention across the clusters [Zwarenstein 2008].

In Example 1 the intervention being evaluated is formed of several components. Depending on the format of the different components, in some studies~~This~~ this might mean there may be both a delay before any anticipated effect is realised; and it might be the case that the effects of some components might wa~~i~~ne through familiarity. Furthermore some ~~some~~ components of an ~~the~~ intervention ~~in Example 1~~ might be~~are~~ continuously delivered (i.e. provision of pamphlets) whereas some components might be ~~are~~ delivered just once (i.e. educational components). In Example 4 ~~it is~~ the educational component of the intervention is re-enforced and so its anticipated effect is less likely to decay.

Methods: Outcomes

Item 6a: Outcomes

Standard CONSORT item: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.

CONSORT cluster extension: Whether outcome measures pertain to the cluster level, the individual participant level or both.

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.

*Example 1 (Pre-specified outcomes)***:** "The primary outcome of the study is a 7-day period prevalence of diarrhoea among villagers of all ages. Secondary outcomes include a 7-day period prevalence of other hygiene-related illnesses (respiratory and skin infections), reported changes in hygiene practices, household water usage and water supply preference." [Riverbank Filtration Trial]

Example 2 (Cross-sectional sampling): "Data collection for the evaluation took the form of a postal survey conducted at five fixed time points: baseline (in the month prior to commencement of the first intervention period) and within a week of the end of each of the four intervention periods. A repeated cross-sectional design was employed, in which a random sample of households within each cluster was selected to receive the survey at each period." [DAVE Trial]

Example 3 (Cohort design): "All household members will be eligible for inclusion in the study, regardless of age. …Each household will have the option to participate in up to five subsequent surveys…Outcomes will be measured at each of the six survey visits." [Riverbank Filtration Trial]

Example 4 (Transition period): "A 1-month transition phase is included where the medical centre is not considered as being in control or intervention and does not contribute to analysis. This transition period allows for the time it takes to embed the intervention into a medical centre." [REMAIN Trial]

Example 5 (Time to assessment and source of data): "Participants will be followed up to 12 months from day of hospital discharge. This will be done through collection of routine data from the hospital and medical centre. Demographics and reason for admission at enrolment and subsequent admissions in the 12-month follow-up will be collected through participant hospital records…Medical centre records will be used to identify whether a discharge treatment plan was received and the timeliness and number of GP visits during the 12-month follow-up period for each participant."

Explanation: All outcomes should be completely defined. This should include the pre-specified primary outcome and all secondary outcome measures (Example 1). It is also important to report clearly how and when these measurements were obtained.

SW-CRTs make a series of measurements over time within each cluster. These measurements could be on different participants in each period (i.e. cross-sectional design) as in Example 2; the same participants (i.e. cohort design) as in Example 3; or a mixture, and this will inform the method of analysis and has implications for sample size calculations. Data are rarely collected at the level of the cluster, but knowledge of whether outcomes in each period are at the cluster level (either because of true cluster level outcomes or because of the availability of aggregated data only) or individual level has implications for the method of analysis.

It should be reported whether outcomes are collected at discrete points in time common to all participants (e.g. a survey implemented at several discrete points in time as in Example 3), or at time points specific to each participant (e.g. as they leave hospital as in Example 5). The timing of measurements has implications for the choice of analysis. For example, if the outcomes are collected at discrete time points (as in Example 3), then time effects can be included as categorical effects; whereas if the outcomes are collected continuously (for example as would be the case in a SW-CRT where the outcome was routinely collected mortality data), then time effects could potentially be modelled using parametric or semi-parametric forms.

The reporting of the timing of data collection should also note whether there were periods in which outcomes were not ascertained, for example transition periods immediately after the intervention was rolled out, to allow time for the intervention to realise its full impact (as in Example 4).

In individually and cluster randomised parallel trials outcomes are often assessed at multiple time points (for example 6 and 12 months post randomisation) and it is important to pre-specify the primary follow-up time of interest. This might also be the case in SW-CRTs. Sometimes the outcome assessments will extend beyond the actual study dates. For example, a trial might roll-out the intervention to clusters over a four year period and the primary

follow-up time might be 30 years later [Shimakawa 2014]. Clear reporting on the timing of follow-up assessments (as in Example 5) also allows assessment of whether all observations collected under the intervention condition were fully exposed to the intervention, and whether any observations collected under the control condition might have been contaminated by the intervention.

Reporting whether data were collected from routine sources or purposively collected can help ascertain the risk of bias (e.g. from measurement of the outcome) and identify who are the human research participants (see Item 26). SW-CRTs are often implemented in real-world settings and, as such, may rely on routinely collected outcome data (Example 5). Reporting of whether the data collection procedures changed over time is important given the imbalance over time with respect to intervention conditions [Shadish 2002]. It is also important to report ~~, and~~ any measures which can allow assessment of the reliability and validity of routinely collected data.

~~**Methods Outcomes**~~

Item 6b: Changes to outcomes
Standard CONSORT item: Any changes to trial outcomes after the trial commenced, with reasons.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Any changes to trial outcomes after the trial commenced, with reasons.
~~No modification suggested.~~

Formatted: Font: Not Italic

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Methods: Sample size

Item 7a: Sample size
Standard CONSORT item: How sample size was determined.
CONSORT cluster extension: Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intra-cluster correlation (ICC or k), and an indication of its uncertainty.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* How sample size was determined. Method of calculation and relevant parameters with sufficient detail so the calculation can be replicated (Table 6). Assumptions made about correlations between outcomes of participants from the same cluster.

Example 1 (Sample size): "We would consider an absolute increase of 10**%** in the proportion of patients who are registered organ donors at 7 days post-encounter to be both clinically important and feasible. Our sample size of 6 clusters (10,500 patients in total) achieves 80% power to detect this difference assuming a control proportion of 0.5 using a two-sided test at the 5% level of significance [Hooper 2016]. Our calculation assumes an intra cluster correlation coefficient of 0.06, as calculated from our previous work (19), an average of 250 patient encounters per site in each two-week interval, and a cluster autocorrelation coefficient of 0.8 to allow for a 20% decay in the strength of the correlation in repeated measures over time.(20) The percentage of registered donors in the control condition is conservatively assumed to be 50% to allow for a higher prevalence of registered donors in our participating offices than the provincial average. No adjustment is made for cluster attrition as the risk of attrition is low, and all outcomes will be assessed from routinely collected sources, regardless of any drop-out. Given some uncertainty around parameter estimates required for the stepped wedge sample size calculation, sensitivity of our detectable effect size to a range of alternative assumptions is presented in Table *(not shown)*. The results show that across a range of control arm proportions (from 0.4 to 0.5), average cluster sizes (from 100 to 400), and

cluster autocorrelation coefficients (from 0.8 to 0.95), our sample size of 6 practices will achieve 80% power to detect absolute increases between 5% and 11%." [RegisterNow-1 Trial]

Example 2 (Sample size fixed by design): "The study had a fixed sample size by design that could not be modified, so the power calculations did not inform any sample size targets." [Targeted Case Finding Trial]

Explanation:

The method of calculation and all relevant parameters, used in the sample size calculation should be given. Most of the key items to report are listed in Table 6. These have been divided into key items which are essential and likely of relevance to all SW-CRTs; and those which might be considered additional or supplementary information which will only be of relevance to some SW-CRTs. Besides the usual effect size, significance level and power, these may include: the cluster size and whether account of unequal cluster sizes has been made, avoiding any ambiguity between cluster size per measurement period and total cluster size; a within-period intra-cluster correlation (ICC) and assumptions about correlations between outcomes of different participants from the same cluster in different periods (or other assumptions which appropriately reflect the complexity of the design); allowance for repeated measurement taken from the same participants, with sufficient detail to allow the calculation to be replicated. Often a sensitivity analysis, looking at the effect of relaxing some of the assumptions, may be warranted.

Specifying the method of sample size calculation [Hussey 2016; Hooper 2016], or providing access to sample size calculation code [Baio 2015; Hooper 2016; Hemming 2016] or programmed sample size function [Hemming 2014] can aid replication of the sample size (Example 1 reported they used the Hooper method). Detailed reporting of the sample size method will allow assessment of whether the method has allowed for all features inherent to the particular design (e.g. transition periods, repeated measures on the same participants). Reporting of the sample size calculation will likely include: number of clusters and whether equal or unequal cluster sizes are assumed, cluster size or cluster size per period, number of sequences, and number of clusters per sequence. Reporting of these basic sample size elements is poor in SW-CRTs [Martin 2016]; as is the reporting of basic elements in parallel CRTs [Rutterford 2015].

For clarity it is important to distinguish between total cluster size (across all periods) and cluster sizes per period (Example 1). In a design which repeatedly measures the same participants it would be natural to provide the number of participants in each cluster and the number of repeated measurements per participant; in a design which involves taking repeated, discrete samples with different participants each time it would be natural to provide the number of participants in each cluster in each of these periods; whereas in a design where newly eligible individuals are recruited continuously it might be more appropriate to report the total number of participants expected in each cluster over the duration of recruitment.

In a parallel CRT it is important to report the ~~intra-cluster correlation coefficient (~~ICC~~)~~ (the correlation between outcomes of two individuals from the same cluster). The coefficient of variation of cluster rates, proportions or means has been suggested as an alternative parameter in sample size formulae for CRTs [Hayes 1999]. Correlation structures are more complicated in a SW-CRT and there may not be a single ICC, as the strength of correlation might depend additionally on the separation in time [Hooper 2015; Martin 2016b; Kasza 2017]. Such correlation structures could be formalised in a variety of ways, for example using a within-period ICC and a between-period ICC or cluster auto-correlation coefficient (as in Example 1) [Kasza 2017]. In SW-CRTs where the same individuals are assessed repeatedly it may also be important to consider correlations over time within individuals [Hooper 2016].

An indication of the sensitivity of the sample size or power to the assumed parameter values could be provided, for example, by reporting sample size or power at a variety of alternative correlation values. Rationale for the assumed parameter values should be provided (as in Example 1).

In randomised trials the sample size (and so consequently the number of clusters) is often based on the number needed to detect the target difference at a desired level of power and significance [Cook 2017]. SW-CRTs can sometimes have their sample size fixed by the number of clusters, participants, or both, available in a natural setting. Whether the sample size was fixed by factors outside of the control of the experimenters or based on the target difference (as conventionally is the case in a randomised controlled trial) should be reported (as in Example 2). When the sample size is fixed, it can be useful to report what effect size the study was powered to detect. If no power calculation was performed, this should be reported. Retrospective power calculations based on the results of the trial are of little merit [Hoenig 2001; Sculz 2010].

Item 7b: Interim analyses
Standard CONSORT item: When applicable, explanation of any interim analyses and stopping guidelines.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs*: When applicable, explanation of any interim analyses and stopping guidelines.
~~No modification suggested.~~

Formatted: Font: Not Italic

Explanation: Interim analyses of outcomes can be used to assess harm, futility, and efficacy. Interim analyses can also be used to monitor recruitment and retention rates, and monitor balance across control and intervention conditions (where trial processes suggest that there may be a risk of differential recruitment or consent).

The relevance of interim analyses of outcomes might be questionable in some SW-CRTs, so careful reporting of motivation is important. For example, if the intervention is being rolled out to all clusters within the fastest time frame possible, then stopping the trial early after demonstrating efficacy does not necessarily mean the intervention can be rolled out to the remaining clusters immediately. In some settings, SW-CRTs evaluate interventions for which safety concerns are likely to be minimal (although this will not always be the case). It might be of interest to consider stopping a SW-CRT for futility, although if there are minimal safety concerns then stopping the trial early for futility may also not be worthwhile. However, other important reasons for considering stopping a trial include that the trial itself is not successful, perhaps because clusters are failing to adhere to the randomisation schedule, because data for outcomes are not forthcoming, or because procedural requirements have delayed the start dates for many clusters [Kristunas 2017]. Dates or times at which any interim analysis will be carried out should be reported together with objectives of such interim analyses.

Of note, in a SW-CRT due to the imbalanced nature of the design, interim analyses for outcomes carried out early in the trial will have a large imbalance between numbers of observations exposed to control and intervention conditions. This imbalance is likely to have power implications [Grayling 2017]; and will make a blinded interim analysis infeasible. The clustered nature of the data will also have implications on power and interim analyses [Zou 2005]. Proposed methods of interim analysis should be outlined. Interim analyses of outcomes might or might not follow the same method of analysis planned for the main results. As with any trial, incorporation of any interim analyses of outcomes (where a decision is to be made about continuation of the trial) should be allowed for in power calculations to control for the over-all Type I error rate.

Methods: Randomisation ~~**– Sequence generation**~~

Item 8a: Sequence generation
Standard CONSORT item: Method used to generate the random allocation sequence.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs*: Method used to generate the random allocation to the sequences of treatments.

Example: "Eligible schools were randomly assigned to one of the four sequences (3 or 4 schools per sequence) for time of crossover from control to intervention using a computer-generated list of random numbers." [SBP Trial]

Explanation: Random allocation in SW-CRTs takes a different form to that in parallel arm designs. Rather than each cluster being randomly allocated to one of two treatments, allocation is to one of several sequences which define the order with which clusters cross from the control condition to the intervention condition (Example). The term "sequence generation" in a SW-CRT therefore has a slightly different meaning to that of individually randomised trials. In an individually randomised trial "sequence" refers to a sequence of treatments to allocate all participants to either the intervention or control condition.

Furthermore, rather than the randomisation being performed as clusters or individuals present to the trial the randomisation in a SW-CRT is usually done at a single point in time before the trial starts.

~~**Methods Randomisation – Sequence generation**~~

Item 8b: Randomisation method~~Sequence generation~~
Standard CONSORT: Type of randomisation; details of any restriction (such as blocking and block size).
CONSORT cluster extension: Details of stratification or matching if used
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Type of randomisation; details of any constrained randomisation or stratification if used.

Example 1 (Unrestricted): "Nursing-home units were the unit of randomisation... RL (not involved in recruitment) randomly allocated units to one of five groups with computer-generated random numbers…" [Depression Management Trial]

*Example 2 (Stratification): "*All schools are assigned a decile rating, which indicates the extent to which the school draws its students from a range of socioeconomic areas. Decile 1 schools are the 10% of schools with the highest proportion of students from low socioeconomic resource areas (defined according to residents' income, occupation, household crowding, educational qualifications and income support) and decile 10 are the 10% of schools with the highest proportion of students from high socioeconomic areas…. The order of switch-over is determined randomly for each group (*decile*) of clusters" [SBP Trial Protocol]

Example 3 (Covariate constrained randomisation): "The randomization was conducted using a highly restricted randomization design. With this limited number of randomization units, selection of one sequence from the 5.4 $*10^{26}$ completely at random would run the risk of obtaining a sequence that is substantially unbalanced with respect to one or more potentially important covariates. Randomization was done using a highly restricted randomization design to achieve close balance with respect to clinic-level covariates including mean CD4 count, clinic size, average education, tuberculosis treatment levels, existence of a supervised tuberculosis therapy (DOTS) program and geography (reference cited to detailed methods)". [THRio Trial Protocol]

Explanation: In a SW-CRT, rather than the randomisations being done sequentially (as the patient or cluster presents to the trial), the randomisation is usually done at a single point in time before the trial starts. This means that different methods for controlling balance of cluster-level factors can be considered along with methods used in individually randomised trials such as ~~minimisation and~~ stratification [Ivers 2012]. How the randomisation is restricted is known to have implications for analysis.

There are two common ways in which clusters may be allocated in a SW-CRT. One is simple unrestricted allocation to one of several possible sequences (Example 1); another is stratified allocation with clusters divided into distinct

strata prior to random allocation within each stratum (Example 2). For a stratified design the sequences are generated independently within each stratum. This essentially means that separate mini SW-CRTs are conducted in each stratum (Example 2). Yet another method of allocation is covariate constrained allocation which balances key covariate values (such as cluster size) between intervention and control conditions (Example 3) [Moulton 2007].

~~**Methods Randomisation – Allocation concealment**~~

Item 9: Allocation concealment
Standard CONSORT item: Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned.
CONSORT cluster extension: Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both.
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Specification that allocation was based on clusters; description of any methods used to conceal the allocation from the clusters until after recruitment.

*Example 1 (Concealment from cluster)***:** "Once 14 medical centres have provided consent to be involved in the study, each enrolled medical centre will be randomised to a transition step." [REMAIN Trial]

*Example 2 (Concealment of cross-over date)***:**"The allocation sequence will only be made available to two study investigators (ABF and MS). Indian study investigators will be blinded to the allocation sequence with only the next village randomised for rollout being revealed at each intervention implementation time point. Study participants will be blinded to the allocation sequence and those not yet receiving the intervention will not be aware of the time at which they will have the intervention implemented." [Riverbank Filtration Trial]

Explanation: In a SW-CRT clusters are allocated to a sequence of treatments, so clusters will spend time in the control condition until a particular date when they cross to the intervention condition. This is unlike a parallel arm cluster randomised trial in which clusters are allocated to treatment conditions. Randomisation of all clusters (to sequences) in a SW-CRT will often occur at a single point in time (as in Example 1). Randomisation could in theory also be performed at step-times, where one or more of the remaining clusters will be randomly selected to cross over just prior to the cross-over date (no examples of this have been identified).

It is important to report any method that was used to conceal the allocation from clusters and from those individuals responsible for recruiting clusters, until after recruitment. Reporting of this information allows assessment of the potential for selection bias [Higgins 2016]. One common way of preserving allocation concealment is to perform the randomisation after recruitment of all clusters (as in Example 1).

When randomisation of the clusters occurs at a single point, the cross-over date may be revealed immediately to each cluster, or revealed sequentially to the clusters as they approach the time of cross-over (as in Example 2). Reporting when clusters were told of their cross-over date allows assessment of potential biases. For example, when clusters are informed of their date of cross-over at the beginning of the trial, some clusters (e.g., those randomized to cross over later) may drop-out, leading to differential attrition; yet at the same time a public randomisation at the start of the trial may also prevent subversion of the randomisation process [Higgins 2016]. Knowledge of when a cluster is crossing over could lead to other biases, for example, if individuals within a cluster are aware of the impending cross-over, they may defer enrolling participants into the trial to ensure they receive the intervention.

Full transparency of reporting of the blinding throughout the trial, including the randomisation process, is best reported using a timeline diagram [Caille 2016].

Methods: ~~**Methods Randomisation – Implementation**~~

~~Item 10:~~ Implementation of randomisation

~~*Standard CONSORT item:* Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions.~~

~~*CONSORT cluster extension:* Replace by 10a, 10b and 10c.~~

~~*Extension for stepped-wedge trials:* Replace by 10a, 10b and 10c.~~

Formatted: Font: Bold, Not Italic

Formatted: Indent: Left: 0"

~~*Explanation:*~~ As with a parallel CRT, it is important that all steps in the implementation of the randomisation process are clearly described. It is important that this information on the allocation and recruitment process is described for both clusters and participants. Information on the allocation and enrolment of the clusters is described in Item 10a and corresponding information for participants in Item 10b. Enrolment of participants is closely linked to the consent process (for example, differential consent processes can have implications for selective recruitment). Therefore, following the cluster CONSORT extension, Item 10c describes the consent processes.

Of note, we use the term “selection bias” to refer to any process by which there is differential inclusion of participants in the treatment conditions being compared. Sometimes selection bias is used to refer only to differential *inclusion* of clusters by intervention conditions. More specifically, “identification bias” refers to biases which are induced by differential application of the inclusion / exclusion criteria [Higgins 2016]. The term "recruitment bias" refers to biases which are induced by differential recruitment into the trial by the health care practitioner or to biases induced by individuals differentially declining to participate.

~~**Methods Randomisation – Implementation**~~

*Item 10a: I*nclusion of clusters~~*mplementation*~~

Standard CONSORT item: Not included in original CONSORT statement.

CONSORT cluster extension: Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions.

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Who generated the randomisation schedule, who enrolled clusters, and who assigned clusters to sequences.

*Example: “*We will recruit a convenience sample of practices from within our network of family physician office contacts within the London, Ontario and Stratford, Ontario communities. A collaborating family physician will send an introductory email to potential family physician contacts, inviting them and their practice to consider participating. We will then arrange an in-person meeting with family physicians from interested sites to introduce our study and obtain written agreement from family physicians and offices agreeing to participate that meet our eligibility criteria. A statistician blinded to cluster identity and not involved in the intervention delivery will generate the allocation sequence using computer-generated random numbers.” [RegisterNow-1 Trial]

Explanation: Knowledge of who implemented the randomisation procedures at the level of the cluster is required for ascertaining if selection biases are possible.

It is important to have a separation of roles between those who generate the randomisation schedule and those who recruit, enrol and assign clusters to the sequence (as in the Example). If the person who generated the randomisation was also responsible for recruiting the clusters, this could mean that there was an increased risk of selection bias. This is best achieved by having a person independent of the trial doing the randomisation. This will be less important in trials where the randomisation takes place after recruitment of all clusters.

~~**Methods Randomisation – Implementation**~~

Item 10b: Inclusion of participants~~mplementation~~
Standard CONSORT item: Not included in original CONSORT statement.
CONSORT cluster extension: Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling).
~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration or random sampling; continuous recruitment or ascertainment, or recruitment at a fixed point in time), including who recruited or identified participants.

Example 1 (Complete enumeration with continuous ascertainment): "The study included all patients admitted to 16 acute adult wards of one general hospital over a 32-week period." [Critical Care Outreach Trial]

Example 2 (Random sampling): "Data collection for the evaluation study will focus on adults aged 18 years and over. The study will use a repeated cross-sectional design, in which a random sample of people within each cluster will be surveyed at each stage. A complete list of all households in each of the 128 study villages will be obtained using the Postcode... The order in which households are approached to participate in the survey at each stage will be randomly generated...One adult per household will be randomly selected." [DAVE Trial Protocol]

*Example 3 (Continuous recruitment): "*Then, the leaders of the nursing homes are responsible for the recruitment of the units and the residents according to the inclusion and exclusion criteria of the study. Here, all eligible participants of the participating units are invited to participate. Before the recruitment procedure will commence, each leader of the nursing homes will attend a kick-off meeting held by a senior investigator about the inclusion and exclusion criteria and the planned recruitment strategy. For the participants who drop out of the trial, we are planning to monitor the reasons (for example, death or moving) and perform a sensitivity analysis at the end of the trial to determine whether they differ according to certain characteristics (for example, the prevalence of the challenging behavior or gender). Residents who are newly admitted to clusters during follow up will also be included in the study …" [FallDem Trial]

Explanation: Individual participants can be included in a SW-CRT in many different ways. Sometimes, participants are not recruited into a trial, but rather their data are used from routinely collected sources (Example 1). In this case it is common to take a complete enumeration of the cluster or at least those meeting the eligibility criteria. Alternatively, a sample of individuals from the cluster might be asked to complete data assessments or questionnaires in each period (Example 2). Alternatively, participants might be recruited to participate in the trial. This recruitment might take place continuously (Example 3) or at a fixed point in time before the start of the trial.

Knowledge of how participants are included in the trial can help assess the likelihood of identification and recruitment bias. Trials with complete enumeration are less likely to suffer from these biases (Example 2). Where participants are identified or recruited after randomisation (as in Examples 1 and 3), either a complete enumeration of the cluster or recruitment/identification by someone who is blind to allocation can help mitigate recruitment and identification biases. Therefore, clear reporting of who recruited or identified participants and whether or not such individuals were blind to allocation is important so readers can determine the risks for bias. Identification and recruitment biases will not occur in designs in which participants are recruited prior to randomisation.

~~**Methods Randomisation – Implementation**~~

Item 10c: Consent ~~Implementation~~
Standard CONSORT item: Not included in original CONSORT statement.

CONSORT cluster extension: From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Whether, from whom and when consent was sought and for what; whether this differed between treatment conditions.

Example 1 (Individual-level consent): "Written informed assent was obtained from all participating children as well as parental consent. Only children who provided both assent and parental consent were eligible to take part." [SBP Trial]
Example 2 (Cluster and individual-level consent): "Criteria for inclusion are informed consent obtained from people with dementia or their legal representative….All of the nursing staff working in one of the two participating wards of the nursing home must provide their informed consent" [FallDem Trial]

Explanation: Obtaining informed consent for participation, study interventions, and data collection procedures in clinical trials is an integral principle of research ethics and international human rights law [IEHR 2016; UN 1966]. The process by which consent was obtained can lead to biases [Campbell 2012]. It is important to describe what consent was for (e.g. exposure to the intervention or use of data), whether consent was sought before or after randomisation, and whether the type of consent differed between intervention and control conditions.

In SW-CRTs there can be cluster-level research participants (e.g., health-care practitioners) and individual-level research participants (e.g. patients) [Taljaard 2013]. It is therefore important to identify explicitly from whom consent was obtained in the study (Example 2) or to state that consent was not obtained. Furthermore, in most cluster trials someone provides access to the cluster; such individuals are often called "gatekeepers" or "cluster guardians" [Edwards 1999]. Gatekeeper permission for trial participation is different to consent from cluster-level research participants, such as health providers, for their own participation in the study.

In cluster randomised trials in which the treatment is delivered at the level of the cluster, it may not be possible to obtain consent for exposure to the intervention or control condition as the intervention may be impossible to avoid (as would be the case in Example 1 under Item 10b); however, consent can still be taken for use of data (implied by return of questionnaire data in Example 2 under Item 10b). It is therefore important to clearly report what consent was for. If participants recruited to the control and intervention conditions are given different information when their consent is taken, this can lead to bias [Eldridge 2005]. The information provided about the objectives of the study can itself prompt participants to act differently. For example, participants enrolled in a study of an intervention to increase uptake of HIV screening, who are fully informed about the objectives of the study, might increase uptake of screening irrespective of allocation to the intervention condition. This is known as the Hawthorne effect [McCarney 2007]. Reporting what information was provided to participants can allow readers to judge the risks of such biases. A recent systematic review found that of the small number of SW-CRTs that reported whether or not consent was obtained, only a small proportion reported explicitly what this consent was for, and none reported when the consent was taken [Taljaard 2017].

Sometimes a research ethics committee might deem it appropriate that the study proceed without the informed consent of research participants (i.e. a waiver of consent) or the research ethics committee may otherwise modify informed consent requirements (i.e. modification of consent). When a waiver or modification of consent has been granted by a research ethics committee, it should be reported and a justification given. It should be clear whose consent was waived and whether the waiver pertains to study participation, data collection, or both. Not all jurisdictions allow for a waiver or modification of consent. Information on data collection procedures in the trial, e.g., whether data are anonymous or pseudo-anonymous, and whether they were routinely collected, can provide clarity around ethical aspects of the trial. When appropriate it can be useful to include any participant consent forms in appendices, which will allow readers to infer precisely the information provided to participants.

Formatted: Not Highlight

Methods: Blinding

Item 11a: Blinding

Standard CONSORT item: If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.

CONSORT cluster extension: No modification suggested.

~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* If done, who was blinded after assignment to sequences (for example, cluster level participants, individual level participants, those assessing outcomes) and how.

Example 1 (Blinding not possible): "Blinding to the intervention (i.e., the type of water being received) is not possible due to potential differences in turbidity of untreated and RBF *(Riverbank Filtration)*-treated river water." [Riverbank Filtration Trial]

Example 2 (Blinding partially possible): "Residents did not know when the intervention was being implemented or what the programme elements were. Interviewers who administered the outcome questionnaires were masked to intervention implementation or depression treatment, and to previous test results. Data analysts were masked to whether a specific resident had been exposed to the intervention and to when the intervention was implemented in a unit, but were not masked during post-hoc analyses." [Depression Management Trial]

Explanation: SW-CRTs are often used to evaluate interventions for which it is impossible to blind participants or clusters to whether they are in the intervention or control condition, but nonetheless it is important to report clearly whether or not blinding was used and if so, who exactly was blinded to aspects of the trial (Example 1).

Often outcomes are collected at multiple levels (e.g. hospitals (e.g. team climate outcomes), clinicians (e.g. knowledge, skills, practice outcomes), patients (e.g. pain)). The possibility of blinding may be different depending on the level of participants (e.g. clinicians or patients) and may depend on the type of consent required (Item 10c). The degree of blinding should be reported at each level of the trial (e.g. clusters, participants as in Example 2) and whether the blinding differed in control and intervention conditions. Researchers should also specifically report blinding with respect to all outcomes. Blinding of those assessing outcomes should be clearly reported.

A systematic review has found that most SW-CRTs do not report clearly who was blinded and what people were blinded to [Taljaard 2017]. Whether or not and who was blinded, and when, is best reported by the use of a timeline diagram [Caille 2016].

Item 11b: Blinding

Standard CONSORT item: If relevant, description of the similarity of interventions.

CONSORT cluster extension: No modification suggested.

~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* If relevant, description of the similarity of treatments.

~~No modification suggested.~~

Formatted: Not Highlight

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Explanation: In trials with a placebo it is important to provide evidence of the similarity of the control condition to the intervention condition (i.e. to provide evidence of the blinding). However, In SW-CRTs it would be unusual to have a placebo and often participants are not blind to their allocation status. Sometimes, a minimal level of intervention is provided in the control condition in an attempt to keep participants blinded to their status as intervention or control participants. When appropriate such minimal level interventions should be described in full.

Methods: Statistical methods

Item 12a: Statistical methods
Standard CONSORT item: Statistical methods used to compare groups for primary and secondary outcomes.
CONSORT cluster extension: How clustering was taken into account.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Statistical methods used to compare treatment conditions for primary and secondary outcomes including how time effects, clustering and repeated measures were taken into account.

Example 1 (Allowance for clustering and secular trends): "A generalised linear mixed model was used for categorical outcomes, and a linear mixed model was used for continuous outcomes, adjusting for age, gender, ethnicity and school terms (i.e., secular trend). The cluster effect by school and correlation between repeated measurements on the same child over time were taken into account in the multilevel analysis." [SBP Trial]

Example 2 (Cluster level analysis): The primary outcome (diarrhoeal prevalence) will be calculated for each cell in the stepped wedge design by aggregating over all individuals surveyed in each village during each time period. Estimation of intervention effects will be obtained from a linear regression of the logarithm of the village-aggregated prevalence adjusting for seasonal effects and incorporating village as a fixed effect. The intervention effect coefficient will be exponentiated to produce an estimated relative reduction (with 95% CIs) in the overall prevalence of diarrhoea in the intervention periods (post-RBF) compared with control periods (piped but unfiltered water). This analysis model controls for both clustering of individuals within villages and for repeated assessments of villages over time... We will use multiple-imputation to impute missing outcomes at the individual person level which will then be aggregated for the village-level analyses." [Riverbank Filtration Trial]

Example 3 (Intention-to-treat analysis): "For the "intention-to-treat" analysis an indicator of whether an observation occurred pre- or post-randomisation was included in the regression model. To allow for delays in implementation a separate "per protocol" analysis was performed with the observations now placed into one of the three categories: "pre-randomisation", "post-randomisation but pre-implementation" and "post-implementation…" [FIT Trial]

Explanation: The statistical methodology should be clearly reported to allow replication. Where possible it can be helpful to provide a reference to the statistical methodology used. In a SW-CRT, clusters are randomised to sequentially initiate the intervention. Observations collected under the control condition are therefore, on average, from an earlier calendar time than observations collected under the intervention condition. Changes external to the trial may create underlying secular trends. Likewise participants, if repeatedly measured over the duration of the study, may get sicker or recover over time. This means that time is a potential confounder. Analysis of a SW-CRT should adjust for time effects [Hussey 2007] irrespective of their statistical significance; failure to do so risks biasing the estimate of the intervention effect, which could lead to declaring an intervention effective when it is ineffective or ineffective when it is effective [Hemming 2017]. It is therefore essential to report if and how time effects were allowed for. If time is measured continuously, time can be modelled parametrically; if time is measured discreetly then time can be modelled categorically. Furthermore, SW-CRTs typically include only a small number of clusters [Martin 2016] and so pre-specification of important prognostic factors to use in a fully adjusted analysis (in mitigation of the likelihood of imbalance due to sampling variation) might also be undertaken [Senn 1994].

In a parallel CRT, randomisation at the level of the cluster needs to be allowed for at the analysis stage (unless cluster level data are being analysed). In a SW-CRT, as clusters (and possibly individuals) are repeatedly measured over time, there may be some reduction in the strength of correlation between measurements within the same cluster over time [Hooper 2016]. Failure to appropriately model the correlation structure can lead to incorrect

estimation of the precision of treatment effects [Thompson 2017]. It is therefore important to clearly describe the correlation structure used in the analysis.

The analysis should also describe how deviations from the randomisation schedule were accommodated (Example 3).~~-~~ A more detailed consideration of this point is given under Item 16 (numbers analysed)~~In the context of a parallel design, an intention-to-treat analysis is defined as an analysis according to allocated group; the analogous definition in a SW-CRT is an analysis which treats all observations taken after the allocated cross-over date as exposed to the intervention.~~

Item 12b: Additional s~~S~~tatistical methods
Standard CONSORT item: Methods for additional analyses, such as subgroup analyses and adjusted analyses.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Methods for additional analyses, such as subgroup analyses and adjusted analyses.

~~No modification suggested.~~

Example (Time varying effect of intervention): "Furthermore, a delayed intervention effect of the CCs *(Case Conference i.e. intervention)* is assumed because the nurses need time to implement the procedure. Thus, the duration of the intervention in months must be considered." [FallDem Trial]

Explanation: SW-CRTs, like other trial designs, will commonly investigate subgroup differences and may perform adjusted analyses. In trials with a small number of clusters, investigating sensitivity to model assumptions will be important [Taljaard 2016].

Of some importance in a SW-CRT is time by treatment interactions. Treatment by time interactions are treatment effects which change as the study progresses (not to be confused with secular changes which represent changes in the outcome under the control condition– Table 2 Key concept 1). These changing treatment effects are important ,because ~~since~~ observations contributing to the analysis will comprise a mixture of times since roll-out of the intervention. ~~F~~Interventions delivered at a single occasion (and not repeated to ensure it creates a permanent effect) might have an impact which changes with increasing time since roll-out (for example, the effect of the intervention might be quite large immediately after roll-out and then its impact might start to wane). If interventions are refined over time then their effect will also change over the duration of the study. Few trials if any have clearly investigated these time by treatment interactions [Davey 2015; Martin 2017], although many interventions have been assessed as being at risk of time by treatment~~intervention~~ interactions [Davey 2015]. The example above makes an acknowledgement of the possibility of a delayed effect, although gives limited detail as to how it will be investigated.

Of particular interest in a SW-CRT might be whether the ~~effect of the~~ intervention has a delayed effect (perhaps because its anticipated effect is not expected to materialise immediately (i.e. a lag effect); or if the intervention effect varies by time since exposure (e.g. an effect that decays over time or an effect that improves over time), perhaps because the effect of the intervention might be expected to wane with increasing time since exposure, particularly so in educational type interventions [Hughes 2015]; or perhaps due *to* the intervention being refined over the course of the roll-out.

Also of interest might be whether the effect of the treatment varies between sequences, perhaps because participants get sicker (or recover) with longer duration in the control condition and the treatment is not anticipated to have the same effect in sicker participants [Copas 2015].

Results: P~~P~~articipant flow

Formatted: Not Highlight
Formatted: Highlight
Formatted: Font: Not Italic

Item 13a: Participant flow

Standard CONSORT item: For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome.

CONSORT cluster extension: For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome.

~~*Extension for stepped wedge trials*~~*Extension for SW-CRTs:* For each treatment condition or allocated sequence, the numbers of clusters and participants who were assessed for eligibility, were randomly assigned, received intended treatments and were analysed for the primary outcome (Figure 3).

Item 13b: Participant attrition ~~*Participant flow*~~

Standard CONSORT item: For each group, losses and exclusions after randomisation, together with reasons

CONSORT cluster extension: For each group, losses and exclusions for both clusters and individual cluster members.

Extension for SW-CRTs~~*stepped wedge trial*~~*:* For each treatment condition or allocated sequence, losses and exclusions for both clusters and participants with reasons.

Example Flow chart by treatment condition and sequence (cross-sectional design): Supplementary Figure S2 (Long-live Mothers Trial)

Explanation: Information on the number of clusters and participants who were assessed for eligibility and outcomes along with the number of losses and exclusions (i.e. withdrawals) allows the reader to assess the risk of differential inclusion and attrition.

Any flow chart should allow the reader to examine the nature of any differential inclusion and attrition by allocated sequence, treatment condition, and over time (see Example Figure S2). Because there are many different types of SW-CRTs there is unlikely to be one flow-chart that will be applicable for all SW-CRTs. How the flow chart is constructed will depend on how many sequences and clusters there are, whether participants contribute repeated measures, and whether participants can join and leave the study. This information could be presented by allocated sequence but might also be presented by treatment conditions.

Including time periods in the flow chart is important to allow for assessment of differential participation over time. When different participants are sampled in each period, each participant will, in theory, be exposed to either the intervention or control condition. In this case, summarising the number of participants by treatment condition is possible. Where the same participant contributes multiple measurements, each participant may provide measurements under both intervention and control conditions. In this case, summarising the number of participants by allocated sequence, along with the average number of measurements contributed by each participant, is more appropriate.

Reporting the number of clusters and participants approached, eligible and included along with the reasons for non-participation is important to allow an assessment of study generalizability, and perhaps even more importantly, of biases due to differential participation between treatment conditions (or sequences). For example, in a parallel CRT without blinding of participants to treatment condition at the time of recruitment, a higher rate of consent among those recruited to the intervention condition can indicate recruitment bias [Caille 2016]. Information on reasons as to why participants or clusters are not included allows a reader to assess the appropriateness of exclusions.

Results: Recruitment

Item 14a: Recruitment

Standard CONSORT item: Dates defining the periods of recruitment and follow-up.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Dates defining the steps, initiation of intervention and deviations from planned dates. Dates defining recruitment and follow-up for participants.

Example 1 (Step dates): "Twenty-two villages received the intervention in the second period (April-June 2011), 36 in the third period (September-November 2011), 35 in the fourth period (April-June 2012), and 35 in the fifth period (September-November 2012)." [DAVE Trial]

Example 2 (Deviations from planned dates): **"**There were 60 study wards in the 16 randomised hospitals, of which 33 (22 ACE and 11 ITU) in 13 hospitals went on to implement the intervention, with a mean (SD) delay in implementation of 5 (4) months …and a mean (SD) duration of implementation of 12 (7) months. Eight wards began implementation very late, and for these the end of the trial was extended to December 31st 2009 to ensure that they had a year of data collection post-implementation." [FIT Trial]

Explanation: Dates defining periods of recruitment of participants can be reported where appropriate; in some designs these dates will be at the beginning of the study before any cross-over of clusters occurs; in other designs recruitment will be continuous throughout the study. In some studies there will be no direct participant recruitment, but identification of data from participants from routine data sources.

Reporting of other key dates are also important in a SW-CRT. These dates include the dates defining when the study was undertaken and dates defining the steps. Dates defining the start and end of the roll-out phase, as well as the dates of the steps are useful to demonstrate if the trial was implemented as planned (Example 1). Dates should be presented so that they can be easily related to the planned timing of the steps as described in Item 3a. Reporting deviations from planned dates is particularly important in the SW-CRT as they demonstrate deviations from the randomised schedule (Example 2).

Dates defining implementation of interventions will allow assessment of when the intervention is fully implemented in each cluster. Dates defining actual implementation of the intervention should be specified. The realised time for an intervention to become fully implemented may differ from that which was planned. This allows assessment of whether all observations collected under the intervention condition were fully exposed to the intervention; it also allows assessment of whether any observations collected under the control condition were likely contaminated by the intervention. Reporting dates also allows inferences about external influences which may have affected secular trends.

Item 14b: Recruitment
Standard CONSORT item: Why the trial ended or was stopped.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Why the trial ended or was stopped.
~~No modification suggested.~~

Formatted: Font: Not Italic

Explanation: Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010, Campbell 2012].

Results: Baseline data

Item 15: Baseline data
Standard CONSORT: A table showing baseline demographic and clinical characteristics for each group.

CONSORT cluster extension: Baseline characteristics for the individual and cluster levels as applicable for each group.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Baseline characteristics for the individual and cluster levels as applicable for each treatment condition or allocated sequence.

Example 1 Baseline table by treatment condition (cross-sectional design): Supplementary Table S2~~1~~ (DAVE Trial)
Example 2 Baseline table by allocated sequence (open cohort design): Supplementary Table S3~~2~~ (Depression Management Trial)

Explanation: In a parallel CRT a summary of the cluster and participant level characteristics at baseline by treatment condition can allow assessment of the success of randomisation and provides a description of the included sample. In trials with post-randomisation recruitment, this table can allow an assessment of potential biases.

The term "baseline" in a SW-CRT can be confusing because of the longitudinal nature of the design. We use the term "baseline characteristic" to mean a characteristic which was either measured before exposure to the control or intervention condition, or which is not expected to be influenced by the treatment conditions (e.g. age). In designs in which observations are made on different participants in each period, these baseline characteristics will often pertain to measurements made just prior to the switch from control to intervention condition ~~at that period~~ (i.e. not at the start of the trial); whereas in designs where participants are repeatedly assessed, these characteristics might be measured prior to randomisation. Cluster level characteristics can often be measured prior to randomisation and are less likely to change over time.

For SW-CRTs in which observations are made on different participants in each period, the summary of baseline characteristics could be presented by treatment condition or by allocated sequence. For example, the DAVE Trial, which measures different participants in each period, reports its baseline table by treatment condition (Table S2~~1~~).

For SW-CRTs in which the same participants are repeatedly assessed in each of the periods, the baseline characteristics of participants will normally be presented by allocated sequence rather than by treatment condition. This is because most participants will be observed first under the control and then intervention condition. The Depression Management Trial (Table S3~~2~~) provides summary characteristics by allocated sequence.

Results: Numbers analysed

Item 16: Numbers analysed
Standard CONSORT: For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups.
CONSORT cluster extension: For each group, number of clusters included in each analysis.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* The number of observations and clusters included in each analysis for each treatment condition and whether the analysis was according to the allocated schedule.

Example 1 (Numbers by treatment condition): "A total of 5295 surgical procedures were carried out throughout the stepped wedge cluster RCT, that is, 2212 in control and 3083 (of which 2263 had the SSC performed) after implementation of the SSC (*Surgical Safety Checklist*). Patients (14.9%; 667/4475) underwent more than 1 procedure. The control and SSC study steps included 1778 and 2033 unique patients, respectively." [Surgical Checklist Trial]

Example 2 (Intention-to-treat vs. per protocol): "The flow diagram shows there were 60 study wards in the 16 randomised hospitals, of which 33 (22 ACE and 11 ITU) in 13 hospitals went on to implement the intervention…

For the primary outcome, intention-to-treat analysis was conducted for the 60 wards randomised into the intervention, and per-protocol analysis was performed for the 33 implementing wards…" [FIT Trial]

Explanation: The number of observations by treatment condition should be reported for analyses of all outcomes (Example 1). For some outcomes this information will be included in a flow chart although not all flow charts for a SW-CRT will give an immediate summary of this information by treatment condition. When the same participants are repeatedly measured across the time periods, each participant will have been exposed to both treatment conditions and so this information can be reported either by giving the total number of observations (by treatment condition) or as the number of participants in the study and average number of assessments per participant under each treatment condition. Where different participants contribute to each measurement period, it might be useful to have information on the number of participants per cluster-period. Such information might be most easily reported in a diagram rather than in text (Figure 3).

Sometimes clusters (and perhaps participants) will not receive the intervention condition as per the randomisation schedule (Example 2). In a parallel trial an~~The~~ intention-to-treat analysis performs the analysis according to the groups to which participants or clusters were originally assigned [Moher 2012]. In a SW-CRT this might be interpreted as analysis of ~~treats~~ clusters and participants treated as ~~as~~ exposed to the intervention according to the *dates* of the randomisation schedule (i.e. according to the planned date~~s of being considered exposed to intervention~~). ~~In a SW-CRT, T~~the application of this principale would ~~would~~ mean that clusters are ~~would be~~ treated as exposed to the intervention if the observation comes from a time period post allocated cross-over date. When a SW-CRT has randomised clusters to actual dates of ~~to~~ transitioning from control to intervention, an intention-to-treat analysis following this interpretation is logical.

Alternatively, a SW-CRT might be considered as randomising the order that the clusters transition from control to intervention (although when there are multiple clusters per sequence, several clusters share the same rank-order). In this situation an intention-to-treat analysis might be interpreted as analysis of clusters and participants treated as exposed to the intervention according to the *order* of the randomisation schedule (i.e. according to the planned order of roll-out). The application of this principle would mean that clusters are treated as exposed to the intervention only after the intervention has been implemented ~~roll-out~~ in that cluster, provided the order of the allocation did not deviate from that planned.

Providing information on the number of clusters (and participants) contributing to all ~~the intention-to-treat and other~~ analyses allows assessment of whether the analysis has been conducted with respect to the randomised cross-over schedule – which might not be in strict accordance with any pre-specified dates; or to ~~and not to~~ the actual cross-over dates that may deviate from planned dates due to delays in implementation.

Sometimes a cluster may drop out from some purposively collected outcome assessments, but still contribute data from routinely collected sources for other outcome variables. If the numbers included in secondary analyses differ from those included in primary analyses, information on differential attrition (or participation) across clusters or periods can be provided in the text (similar to information depicted in the flow chart for the primary outcome (Figure 3).

Results: Outcomes and estimation

Item 17a: Outcomes and estimation
Standard CONSORT item: For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).
CONSORT cluster extension: Results at the individual or cluster level as applicable and a coefficient of intra-cluster correlation (ICC or k) for each primary outcome.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Not Highlight

Formatted: Not Highlight

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* For each primary and secondary outcome, results for each treatment condition, and the estimated effect size and its precision (such as 95% confidence interval); any correlations~~-~~ and time effects estimated in the analysis.

Example 1 (Time adjusted treatment effect): "A total of 321 (10.8%) unexposed patients were started on either antihypertensives or statins, and 577 (19.7%) exposed patients. The time-adjusted mean difference in proportion of patients initiating either treatment was 15.5% (95% CI = 3.9 to 27.1)." [Targeted Case Finding Trial]

Example 2 (Secular trend): Supplementary Figure S3~~2~~ [FIT Trial]

Example 3 (Correlations): "The ICC in the time-adjusted analysis for initiation of either treatment was 0.014 (95% CI = 0.005 to 0.038)." [Targeted Case Finding Trial]

Explanation: A summary of the findings for each primary and secondary outcome should be provided for each treatment condition. This will allow a description of the severity or prevalence of the outcome in the sample (Example 1). In addition, reporting of results by treatment condition allows estimation of an unadjusted effect of the intervention for comparison with a time adjusted effect (as in Example 1).

Treatment effects should be reported along with 95% Confidence Intervals (CI). A SW-CRT which does not adjust for time is analogous to a simple uncontrolled before-and-after experiment; therefore, it should be clearly reported if the primary and secondary outcomes were adjusted for time (Example 1). To allow an understanding of the potential impact of secular trends it can be helpful to describe the secular trend – either in a figure or as regression coefficients. Ideally this should be done by calendar time and should represent the trend in the clusters yet to be exposed to the intervention (Example 2: Figure S3~~2~~). In some SW-CRTs participants will be recruited at the very beginning of the trial and measured repeatedly. In chronic conditions these participants may naturally regress over the duration of the study; in acute conditions they may recover. Whilst not a secular trend per se, such effects still may lead to confounding of the intervention effect with time and so time should be adjusted for.

Reporting any estimated coefficients of intra-cluster correlation (ICCs) can be informative for the planning of future trials (Example 3) ~~[Hooper 2016]~~. Correlation structures are more complex than in a parallel cluster trials conducted at a single cross-section in time; therefore, analysis (and reporting) of a single measure of correlation such as the ICC might not be sufficient [Kasza 2017]. Relevant correlation coefficients ~~Types of correlations~~ might include correlations between observations in the same cluster and same time period (within-period ICC); correlations between observations in the same cluster but different time periods (between-period ICC), as well as between-period and within-period correlations on the same *individual* [Hooper 2016]. It is important to be explicit about the types of correlations being reported [Martin 2016b]. Reporting of variance components is an alternative to intra-cluster correlations, particularly for non-continuous outcomes [Hayes 1999]. When intra-cluster correlations are reported for binary outcomes, clearly indicating the scale (e.g. proportions or logistic scale) can help interpretation [Eldridge 2009].

~~Types of correlations might include correlations between observations in the same cluster and same time period (within-period ICC); correlations between observations in the same cluster but different time periods (between-period ICC), as well as between-period and within-period correlations on the same *individual*. It is important to be explicit about the type of correlations being reported [Martin 2016b].~~

~~**Results: Outcomes and estimation**~~

Item 17b: *Binary outcomes*~~*Outcomes and estimation*~~

Standard CONSORT item: For binary outcomes, presentation of both absolute and relative effect sizes is recommended.

CONSORT cluster extension: No modification suggested.

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* For binary outcomes, presentation of both absolute and relative effect sizes is recommended.

~~No modification suggested.~~

Formatted: Font: Not Italic

Explanation: In addition to reporting a relative measure of the effect of the intervention it can be helpful to report an absolute measure of the effect: while absolute measures of effects are more easily understood, relative measures of effects are often more stable across different populations [Ukoumunne 2008].

While reporting relative and absolute measures of effects is recommended, further methodological work is required to determine optimal methods of analysis that yield such estimates. Current approaches include fitting two separate models (for example a binomial model with log link to report the relative risks; and a binomial model with an identity link to report a risk difference) or by fitting one model and using a transformation to report the other measure of treatment effect [Pedroza 2016].

Model based methods for achieving estimates on both scales have been investigated in parallel CRTs in which the model is unadjusted for confounders [Ukoumunne 2008]; and~~. Although~~ others have evaluated the performance of these models when covariate adjustment is required [Pedroza 2016].~~In SW-CRTs these models would further need to adjust for the confounding effect of time.~~

Results: Ancillary analyses

Item 18: Ancillary analyses

Standard CONSORT item: Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory.

CONSORT cluster extension: No modification suggested.

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory.

~~No modification suggested.~~

Formatted: Font: Not Italic

Explanation: There are several analyses that can be considered to examine deviation from model assumptions, for example, variations in secular trends across groups of clusters [Hemming 2017]; interactions of the intervention effect with sequence; and whether the effect of the intervention might change with increasing duration of exposure (Item 12b). In the reporting of these ancillary analyses, any limitations due to the assumptions made should be noted.

Results: Harms

Item 19: Harms

Standard CONSORT item: All important harms or unintended effects in each group (for specific guidance see CONSORT for harms).

CONSORT cluster extension: No modification suggested.

~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Important harms or unintended effects in each treatment condition (for specific guidance see CONSORT for harms).

Explanation: Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Discussion~~:~~

Item 20: Limitations
Standard CONSORT item: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.
~~No modification suggested.~~

Explanation: Estimated intervention effects from a SW-CRT will almost always be model-based estimates adjusting for time. There is a host of different models which can be used, but all make some assumptions. The assumptions made and potential limitations should be reflected on.

Item 21: Discussion
Standard CONSORT item: Generalisability (external validity, applicability) of the trial findings.
CONSORT cluster extension: Generalisability to clusters and/or individual participants (as relevant)
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Generalisability (external validity, applicability) of the trial findings. Generalisability to clusters and/or individual participants (as relevant).

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010, Campbell 2012].

Item 22: Interpretation
Standard CONSORT item: Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.
CONSORT cluster extension: No modification suggested
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs:* Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.
~~No modification suggested.~~

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Other information

Item 23: Trial registration
Standard CONSORT item: Registration number and name of trial registry.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs*: Registration number and name of trial registry.
~~No modification suggested.~~

Explanation: The International Committee of Medical Journal Editors (ICMJE) defines a clinical trial "as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the cause-and-effect relationship between a health-related intervention *and* a health outcome" [ICMJE]. The ICMJE states that all medical journal editors should require clinical trials to be registered (prior to the first patient enrolment) as a condition of publication. SW-CRTs of health related

Formatted: Font: Not Italic

Formatted: Font: Not Italic

interventions meet the ICMJE's definition of a clinical trial and so ~~therefore~~ should wherever possible be registered as a clinical trial prior to the study start date.

Reporting the name of the trial registry and the unique trial registration number facilitates crosschecking with the associated registry entry and allows assessment of whether there are any important changes to the trial design, and the potential for any bias (such as outcome reporting bias). Further, reporting details of the trial registration facilitates linking of multiple publications from the same trial, which is of particular importance for systematic reviews. If the trial has not been registered, this should be stated along with the reason.

Studies examining trial registration rates have found that a large percentage of trials are not registered (e.g. 28% - 44% [Azar 2015; Killeen 2014; Wetering 2012]). Further, in the trials that are registered, not all report the registration details in the trial publication, and not all are prospectively registered. A recent review that examined registration of SW-CRTS found that only 50% of SW-CRTs were prospectively registered [Taljaard 2017].

Item 24: Trial protocol
Standard CONSORT item: Where the full trial protocol can be accessed, if available.
CONSORT cluster extension: No modification suggested
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs*: Where the full trial protocol can be accessed, if available.
~~No modification suggested.~~

Formatted: Font: Not Italic

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010; Campbell 2012].

Item 25: Funding
Standard CONSORT item: Sources of funding and other support (such as supply of drugs), role of funders.
CONSORT cluster extension: No modification suggested.
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs*: Sources of funding and other support (such as supply of drugs), role of funders.
~~No modification suggested.~~

Formatted: Font: Not Italic

Readers are referred to the CONSORT statement and the extension to the CONSORT statement for examples and explanation [Schulz 2010, Campbell 2012].

Item 26: Research Ethics Review
Standard CONSORT item: Not included.
CONSORT cluster extension: Not included
~~*Extension for stepped-wedge trials*~~*Extension for SW-CRTs*: Whether the study was approved by a research ethics committee, with identification of the review committee(s). Justification for any waiver or modification of informed consent requirements.

Example 1 (Full review): "The study received ethical approval from the Sport and Health Sciences Ethics Committee at the University of Exeter (February 2011)." [DAVE Trial Protocol]
Example 2 (Waiver of consent): "This study was reviewed by the Regional Committee for Medical and Health Research Ethics (Ref: 2009/561), which advised that use of routinely collected anonymized patient data is clinical service improvement and thus no further approval or patient consent is required."[Surgical Checklist Trial]

Explanation: The original CONSORT statement did not include an item on research ethics approval because it is an existing International Committee of Medical Journal Editors requirement that research "involving human data" should indicate whether the research was reviewed by a research ethics committee [ICMJE]. However, a systematic review found that only 75% of SW-CRTs reported review by a research ethics committee, possibly due to the classification of such studies, by some researchers, as service development or quality improvement. To encourage clear reporting about research ethics review of SW-CRTs we have therefore included this as a new item. This is consistent with the recent extension to the CONSORT statement for pilot studies, which also included this as a new item [Eldridge 2016]. An application number or reference number of the ethical approval should also be reported. If a study is deemed exempt from review by a research ethics committee, this should be reported together with a clear justification for the exemption from review.

Conclusions

The SW-CRT offers an exciting new opportunity to rigorously examine the effects of implementation, policy and service delivery interventions. The design is appealing in many respects, but also provides many challenges. It has noteworthy risks for biases including bias due to temporal trends and within-cluster contamination, as well as methodological complexities such as changes in correlation structures over time. Furthermore, perhaps because the design is being used in situations where researchers are not familiar with standards for reporting or conduct, SW-CRTs have been noted to be particularly prone to inadequacies of ethical reporting, including research ethics review and (in common with many cluster trials) identification of research participants. This extension of the CONSORT statement for SW-CRTs encourages researchers to reflect on the unique aspects of the SW-CRT and improve the clarity of reporting.