

Review

Not peer-reviewed version

How to Design and Analyze Efficient Cluster Randomized Controlled Trials: A Review and Recent Development in Low and Middle-Income Countries

[Amanuel Samago](#) *

Posted Date: 27 December 2023

doi: 10.20944/preprints202312.1826.v1

Keywords: cluster randomized controlled trial; variance inflation factor; interclass correlation; multi-level analysis; public health research; effect modification



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

How to Design and Analyze Efficient Cluster Randomized Controlled Trials: A Review and Recent Development in Low- and Middle Income Countries

Amanuel Yoseph

School of Public Health, College of Medicine and Health Sciences, Hawassa University, Hawassa, Ethiopia; amanuelyoseph45@gmail.com

Abstract: Background: Cluster randomized controlled trials are increasingly becoming popular in public health research interventions and health services research in primary health care design and use in low and middle-income countries. However, the majority of researchers in these trials did not account correctly for the effect of confounders and clustering in the design and analysis stages. **Objectives:** This article aimed to review the main implications of implementing a cluster randomized controlled trials design in public health research and emphasize the practical application of proper analytical methods and the performance of different statistical models. **Methods:** The practical application of various analytical methods is validated via the utilization of experimental data from a public health research case study. This article also demonstrated the performance of hierarchical models and made comparisons in terms of empirically precise effect size estimates. **Results:** Three hierarchical statistical models were provided with highly varied effect sizes during the analysis of the intervention effect on outcome, such as model 1 (AOR: 14.32; 95% CI: 3.75–54.68), model 2 (ARR: 3.52; 95% CI: 2.16–5.77), and model 3 (ARR: 4.79; 95% CI: 2.69–5.52). **Conclusion:** Inaccurate results and potentially misleading conclusions can arise from improper cluster randomized controlled trial designs and analyses. This article has shown that real-world data can be adjusted for the effects of confounders and clustering, and we advocate for the routine use of suitable design and analytical methods. This article concludes that researchers should be careful when selecting hierarchical models to analyze cluster randomized controlled trials.

Keywords: cluster randomized controlled trial; variance inflation factor; interclass correlation; multi-level analysis; public health research; effect modification

Introduction

A study design is a particular protocol or plan for carrying out a study that permits the researcher to convert the theoretical premise into an operating one. It is the study approach to answering the research question [1–3]. The basic epidemiologic research designs are categorized into two categories: descriptive and analytic. A descriptive study is mainly intended to describe the occurrence of a condition in a population by place, person, and time. It is often the first step in an epidemiological investigation. It includes a case report, a case series, a correlational or ecological study, and a cross-sectional study. The analytic study goes further by analyzing the relationship between health outcomes and possible determinants. Analytic studies are further classified into two categories: observational and interventional (experimental). Observational studies permit nature to proceed with its progression, which means the researcher measures but does not interfere. It comprises a case-control and cohort study. But interventional studies incorporate an active effort by the investigator to adjust disease risk factors or the progress of the disease through intervention [1–6].

Choice of study design

The selection of a research design is dependent on multiple situations, such as the research question and objective, researcher skill, availability of time, availability of money, ethical issues, the

status of existing knowledge about the research question, availability of data, nature of the exposure and outcome under study, and duration of the natural history of the disease [5,6].

Ability of different study designs to prove causation

The ability of different study designs to prove causation in increasing order is a case report and series, ecological study, cross-sectional study, case-control study, cohort, a meta-analysis of observational studies, randomized controlled trials (RCTs), and a meta-analysis of RCTs [5–7]. This article review focuses on one of the epidemiology study designs called cluster randomized controlled trials (cRCTs) at a higher level of hierarchy in establishing the causation and becoming the most popular public health research in low and middle-income countries (LMICs). However, most public health researchers do not clearly understand how efficiently and effectively to design and analyze this design, despite its popularity. Therefore, this article starts by highlighting the theoretical understanding of cRCTs and, in detail, its practical application in public health research.

Randomized controlled trials

RCTs are a kind of scientific experiment that aims to investigate the role of some agent in the prevention or treatment of disease [8]. The investigator assigns individuals into two or more groups that either obtain or don't obtain the therapeutic or preventive agent. The group that is assigned the exposure under study is generally called the treatment or experimental group, and the group that is not assigned the exposure under study is called the control or comparison group. Depending on the purpose of the trial, the comparison group may receive no treatment at all, an inactive treatment such as a placebo, or another active treatment (positive control) [8,9]. There can be more than one experimental group or comparison group. The aims of this design are to decrease certain sources of bias while testing the efficacy of new interventions; this is achieved by random allocation of study subjects into 2 or more groups, handling them in a different way, and comparing them with respect to an assessed outcome [9,10]. The active manipulation of the agent (exposure) by the investigator is the hallmark that distinguishes RCT studies from observational ones. In the latter, the investigator acts as a passive observer, merely permitting nature to progress its course. Because RCT studies more strictly look like controlled laboratory studies, the majority of epidemiologists believe that RCT investigations generate more scientifically quality results than do observational studies [10].

Cluster randomized controlled trials

A cRCT is a kind of randomized controlled trial that comprises pre-established groups called clusters, and clusters of study subjects are randomly assigned to treatment and control arms. For instance, naturally-existing clusters can be schools, clinical practices, villages, kebeles, and enumeration areas where the study subjects are schoolchildren, patients, villages, and kebeles inhabitants [11]. CRCT is most commonly used in health services research; however, it can be conducted in several other settings or disciplines as well [12]. The effect of clusters in this type of design has methodological effects for statistical analysis and study design. As a result, clustering frequently leads to correlation or association between observations, which, if not measured and accounted for, could lead to false or spurious inferences about the effectiveness or efficacy of intervention effects on outcomes [13]. It is also a category of RCTs in which sets of study participants—groups of individuals, villages, kebeles, health facilities, or schools (as opposed to individual study subjects)—are randomized in individual RCTs [11]. Due to the above reasons, it is also called cluster randomized trials, place-randomized trials, and group-randomized trials [13,14]. It is utilized when individual-level randomization is impossible due to ethical concerns, logistical management simplicity, and a fear of information cross-contamination between two arms [15].

The gold-standard quality of cRCTs

The "gold standard" quality of cRCTs can be achieved through randomization, the use of placebo, and double-blinding. Randomization is the process of random assignment of study subjects into groups that decreases allocation and selection bias and balances both unknown and known predictive factors between two groups [16]. However, double masking in cRCTs is impractical most of the time due to the nature of some interventions; neither the research team nor study subjects can be blinded, like health education and health promotion interventions in public health practice, to improve certain health outcomes. This type of cRCT is called an open-label trial, but the outcome assessors or data collectors can be blinded. The most common problem in this type of design is the dilution of the intervention effect between two arms [17]. If carefully designed and analyzed, cRCTs are the most valid and reliable type of epidemiological design to demonstrate the causation that impacts healthcare practice and policy. It decreases bias and spurious causality. Findings from RCTs can be pooled in systematic reviews and meta-analyses that are progressively being utilized in evidence-based clinical and public health practices [5,7].

Historical development of cRCTs

cRCTs are a comparatively new study design. However, the method is nowadays well recognized in several kinds of literature. Earlier in the 1980s, there was merely scarce utilization of CRCTs design [11]. However, this design has become more and more popular and formally accepted as one of the study designs, from just seven reported in the 1990s to more than 120 in 2000 [18]. As of June 2008, bibliometric research recognized a rapid increase in the number of scientific publications in medical journals on CRCTs [11]. The popularity and acceptance of the cRCT designs have increased quickly in the last 20 years [13,18]. Similarly, its utilization in public health for interventions delivered at the group or cluster level has become increasingly popular, especially in LMICs. However, the basic concept of proper designs of cRCTs to apply particular interventions is not well understood. Therefore, in this section, this article discusses types of cRCTs.

Types or designs of cRCTs

1. Parallel group design

Parallel group design is the most popular type of design for both cluster and individually randomized controlled trials. The distinguishing feature of this design is that each cluster stays in the arm to which it was assigned at random throughout the study. Thus, all study subjects are randomly assigned to one of two groups, and all study individuals in the allocated group either receive or do not receive an intervention [19,20]. Thus, it is less time-consuming as compared to a crossover design but requires a large number of clusters (resource-consuming).

2. Crossover design

Another popular design is a cross-over design, which is different from the usual 'parallel design'. Over time, each cluster obtains or does not obtain an intervention in a random order. Each cluster takes each of the treatments under investigation on different occasions (usually called legs); the order of treatment is randomized, i.e., A B or B A. Each cluster acts under its own control. It is particularly needed in the presence of large inter-cluster variations. It is usually more economical: half the number of clusters has to be enrolled than in a parallel group design. It is of limited use when differential carryover effects or interactions are anticipated. It may take a longer time than a parallel study design due to two treatments and a possible washout period. Missing data poses a more serious problem than in the parallel design. Clusters are randomized to interventions, outcomes evaluated, and then groups crossover to other interventions. One of the advantages of this design is that it improves the efficiency and effectiveness of power due to the effect of clustering. The basic limitations of this design are liable to carry-over impacts, such as patient crossover designs and doubles or longer lengths of the trial duration [21–23].

3. Step-wedge design

The Gambia Hepatitis Intervention Study is credited with coining the term "stepped wedge" because a schematic illustration of the design clearly shows a stepped-wedge shape [24]. In this design, the crossover usually occurs from control to intervention, with the treatment remaining in place once implemented. The stepped-wedge design (SWD) entails collecting data during a baseline period when no clusters are exposed to the treatment. Following that, clusters are randomized to get the treatment at regular intervals or steps, and all subjects are measured once more. This process is repeated until the intervention has reached all clusters. After all clusters have received the intervention, one more metric or measurement is taken at the end [25–28]. If determining the effectiveness of a treatment is the main objective of the study, the SWD might not be the best design. It is suitable when the focus of the research is on the treatment's effectiveness rather than its mere existence. Overall, logistical and other practical considerations are thought to be the strongest justifications for using a SWD if the study is pragmatic (i.e., aims primarily to implement a particular policy). Moreover, SWD permits all participants to receive the treatment while permitting it to contrast with a control group if the treatment is anticipated to be beneficial and it would not be morally acceptable to prevent it for some participants. Every participant will have the chance to obtain the treatment provided by the study. Comparisons between and within clusters are feasible because, at the conclusion of the trial, each cluster has received both the control and the treatment condition. This maintains a sample size much smaller than what would be required in a conventional RCT while increasing statistical power. Lastly, time effects can be investigated because each cluster alternates at random between the treatment and control conditions at various points in time. For instance, research on the effects of prolonged or recurrent exposure to experimental stimuli on treatment effectiveness is feasible [25–28]. This design has several disadvantages. First, costs may rise dramatically in SWD because the study period is longer and all subjects receive treatment in the end. Furthermore, downstream analysis is not facilitated by SWD because everyone receives treatment eventually [28,29]. Second, in a SWD, the intervention is applied to a greater number of clusters at a later time than at an earlier one. Consequently, the intervention effect could be confounded by an underlying temporal trend; therefore, the confounding effect of time needs to be taken into consideration in both before-trial computations of power and post-trial analysis. More specifically, it is advised to use generalized estimating equations or generalized linear mixed models in post-trial analysis [28,30]. Finally, compared to other kinds of randomized trials, the design and analysis of stepped-wedge trials are more complicated. Prior systematic reviews have brought attention to the inconsistent analysis of these trials and the inadequate provision of sample size calculations [25,28]. In stepped-wedge studies, where data was gathered at each stage, Hussey and Hughes were the first authors to propose a framework and formula for estimating power [26,30]. This has now been extended to various levels of clustering and designs where observations cannot be collected at every step [30].

4. Three arm design

A cRCT with three or more groups that employ a parallel-group design is called a multi-arm trial. Although the term "multi-arm" is used for this article, the terms "arms" or "groups" can be used interchangeably to describe the intervention groups in cRCT [31]. I assumed the cost and complexity related to cRCTs and the challenge of recruiting adequate clusters to offer a sufficient sample size in each intervention arm. The majority of cRCTs are study designs in which clusters are randomized to merely two intervention arms. The 3-arm trial is occasionally feasible; however, cRCTs with more than three arms are very rare. Though, when utilized, they follow two main methods. First, compare two different treatments with a control arm. Second, it compares the same treatment provided at variable levels of concentration with a control arm to evaluate a dose response analysis [32]. Its advantages include that it can be useful for nearly any disease, any number of clusters can be run concurrently, and arms can be in unconnected locations. However, it is frequently criticized for its high variance resulting from an unreliable control unity, and when there are several treatment groups, statistics can get complicated or challenging to analyze and interpret accurately [31,33].

5. Factorial design

This design demands that each cluster be randomly allocated into a group that obtains a specific combination of treatments or non-treatments. In convention, to evaluate the effect of two treatments, one would need to conduct either two studies or three arm trials, which have the drawback of a small sample size in each study arm. However, this design permits the investigation of the independent effects of two treatments in the same study. This has the merits of decreasing cost and sample size. It takes a 2×2 design, resulting in 4 intervention arms: the first arm obtaining the first treatment, the second arm obtaining the second treatment, the third arm obtaining both treatments, and lastly, a control arm. The design generates 4 intervention arms, but assessing the effect of each treatment will be conducted by comparing an important combination of 2 arms against the remaining 2 arms. For example, group one obtains vitamin Y and vitamin X, group two receives vitamin Y and placebo X, group three receives placebo Y and vitamin X, and group four receives placebo Y and placebo X. This method is merely acceptable if there is no interaction between the treatments. Where interactions are preferred or expected, this design can be utilized to detect the mutual effect of two treatments; however, large sample sizes might be needed [32,34].

Advantages or qualities of unbiased cRCTs

Advantages or qualities of cRCTs over individual-level RCTs [14,35–37] comprise:

- Avoids or minimizes information cross-contamination between study subjects, providers, or patients allocated to different intervention groups.
- Appropriate for group-level treatments or interventions.
- Appropriate when individual-level randomization is difficult.
- The capability to research interventions or treatments that cannot be fixed or directed toward chosen or selected individuals. For example, a pre-recorded audio education regarding lifestyle modifications and the capability to control contamination across persons. That means one person's or part's altering behaviors can influence or guide another person's or part's to do so.
- Cluster randomization decreases the cost of conducting a survey. It can often be inexpensive to select street blocks, kebeles, villages, schools, and health facilities and study all the individuals in clusters to decrease the cost or expense of the surveying of all randomly selected individuals, particularly in sparsely or scaterly distributed populations.
- It provides easy mechanisms for logistics and material distribution, handling, and management during field work.
- Occasionally, because of the availability of data, it is merely promising to conduct cluster sampling. For instance, if it is interesting to study households, it could be that there is a lack of a census list of households because of the privacy limitations of the Central Statistics Agency (CSA) of the country. But there might be a community-recorded street block name and their respective addresses, and these may be utilized for producing the sampling frame.

Disadvantages of CRCTs

Drawbacks of cRCTs as compared with individual-level RCTs include superior or higher complexity in analysis and design and a demand for large sample sizes or study subjects to attain similar statistical power with individual-level RCTs [36]. Utilization of this kind of study design also means that all entities of subjects within the cluster are possibly similar, which leads to correlated findings. This correlation is calculated or measured by using the intraclass correlation (ICC). However, this correlation is a well-known element of cRCTs, and a huge percentage of the cRCTs fail to calculate and report ICC. Failure to calculate and account for the variance infiltration factor negatively influences both the incidence of Type I errors and the statistical power of analysis. It can also be much less efficient and effective than individual-level randomized designs [12,14].

Practical considerations in using cRCTs

Despite the popularity of cRCTs in public health research in LMICs, most scholars poorly design and analyze them, which frequently leads to wrong results and conclusions [13]. In light of this, this

article provides detailed information about cRCTs for researchers to effectively and efficiently design and analyze cRCTs and overcome the currently existing gaps. In the first part, this article focuses on the practical consideration of cRCTs at the design stage, followed by the data analysis stage.

Considerations during the design stage

This article will provide step-by-step information on the common problems researchers encountered in public health at the design stage of cRCTs and methods of minimization in detail.

1. Selecting appropriate design

This article has highlighted the five basic designs of cRCTs, along with their advantages and limitations. Thus, public health researchers need to be careful when choosing the appropriate study design that sufficiently addresses their research questions, is cost-effective, less time-consuming, and has ethical soundness.

2. Apply appropriate randomization to minimize bias arising from the randomization procedure (bias in randomization).

The randomization procedure is expected to balance baseline covariates between two arms or cancel the effect of either unknown or known confounders between two arms [38–40]. This means that, in most cases, both intervention groups have the same predictive factors prior to the start of intervention [40]. However, if the randomization procedure is not carefully carried out by blinded individuals or statisticians, the estimated outcome of the intervention may be biased by 'confounding,' which occurs when the sources of intervention arm allocation and outcome are shared [38,39]. Similar to individual-level randomization, bias originating from the randomization procedure also operates at the cluster level in cRCTs [41]. The bias caused by the randomization procedure may be less in cRCTs than in individually randomized trials [42]. In cRCTs, the most likely source of subversion is methodologists (who usually carry out the randomization procedure), who are likely to have knowledge or motives that would lead them to subvert the procedure [43]. To minimize bias arising during the randomization, different methods of randomization are used. Simple, unrestricted, or blocked randomization: in a simple randomization procedure, the clusters are allocated randomly to the control and treatment groups. In the case of a small number of clusters of variable size, this can produce extensive inconsistencies in sample size [44]. Matching: to avoid obvious mismatching of the clusters in the control and intervention arms from the start, the clusters involved in the study are paired concerning predictors like sex, age, socioeconomic status, occupation, and cultural background [41]. One cluster will be randomly chosen for the intervention from each cluster pair, thus ensuring that the two groups of the study are balanced. But this means that when a cluster discontinues the study or a loss of follow-up happens, the paired clusters have to be omitted. To relieve this problem, the matching can be set separately at the data analysis phase [45]. Stratification is the procedure of dividing the study subject into separate groups ("strata"). Then each stratum is expected to be homogeneous concerning important features; however, the strata could differ very broadly from one another [44]. Minimization: the minimization technique signifies a compromise between true randomization and balance [41,44]. Co-variable randomization: Another method is co-variable-restricted randomization, in which clusters are assigned to the study groups in equal numbers based on the distribution of important or fundamental variables [46–48].

3. Minimize the bias arising from the recruitment of participants into clusters

To minimize recruitment bias, it is often important to obtain community and individual-level consent and recruit study subjects before randomization. If this is impossible, blinding the recruiters and individuals who conduct randomization is a critical step to minimize this type of bias. In many cluster trials, clusters are identified and randomized before recruiting participants. In these instances, the allocation of the cluster is typically known to the recruiter and will influence who is approached to require part within the trial and therefore the application of any inclusion and exclusion criteria, opposing the randomization procedure [42,43]. This might have led to differential recruitment for the two conditions, leading to post-randomization selection bias. Again, participants may refuse consent despite the knowledge of the intervention they received, resulting in differences in selection between

the two groups [43,44,48,49]. One symptom of biased recruitment is differential recruitment rates. Even when recruitment rates appear similar between treatment arms, selection biases are often introduced [49,50]. Blinding the participants and research teams during a CRT is typically not feasible. This might end in differing motivations and thus become a source of recruitment bias [49]. To prevent recruitment bias, include the identification and recruitment of participants before cluster randomization or recruitment by a blinded and independent person [48]. If prior identification can't be achieved, then the person recruiting participants should be masked or blind to the cluster allocation, or at the very least have some independence between the enrollment and treatment decisions [41,49]. If strategies, like blinding participants and/or recruiters to their cluster allocation, were employed, then recruitment bias would be taken into account [41,51]. Generally, the justification for selecting a cluster design should be made clear, and efforts made to scale back bias should be reported [51].

4. Minimizing bias due to deviations from intended interventions

The biases that originate when there are changes from the intended interventions. Such differences might be the administration of additional interventions that are inconsistent with the trial protocol, failure to implement the protocol interventions as intended, or non-adherence by trial participants to their assigned intervention, depending on their role and interest [41,44,49,52]. The biases that originate due to deviations from intended interventions are sometimes called performance biases [49]. The consequence of this bias is a dilution of the intervention effect, which might be underestimating or overestimating intervention effects. Thus, it is very important to minimize this bias by taking measures like maintaining adherence to intended interventions, which can be achieved by avoiding the administration of additional interventions that are inconsistent with the trial protocol, implementing the protocol interventions as intended, encouraging adherence by trial participants to their assigned intervention, double blinding, and using appropriate analysis.

The role of blinding

Deviations from intended interventions can sometimes be reduced or avoided by implementing mechanisms that make sure the participants, careers, and trial personnel (i.e., personnel providing the interventions) are unfamiliar with the interventions obtained [53]. This is commonly mentioned as 'blinding', although in some areas the term 'masking' is preferred [54]. Blinding is often difficult or impossible in some contexts, for instance, during a trial comparing a surgical with a non-surgical intervention. Non-blinded ('open') trials may take other measures to avoid deviations from the intended intervention, like treating patients consistent with strict criteria that prevent the administration of non-protocol interventions [41]. Lack of blinding of participants, careers, or people delivering the interventions may cause bias if it results in deviations from intended interventions [54]. For example, low expectations of improvement among participants within the comparator group may lead them to search for and receive the experimental intervention[54]. Such deviations from the intended intervention that arise in the experimental context can cause bias within the estimated effects of both assignments to intervention and adhering to intervention [49].

Appropriate analyses

There are two basic types of analysis in cRCTs: intention-to-treat (ITT) and 'per-protocol' analysis principles. ITT analysis means analyzing the data as randomized initially without considering whatever happened after randomization, loss of follow-up, protocol deviation, or lack of outcome data due to the absence of the respondents during the data collection period [55]. Some authors may report a 'modified' intention-to-treat (mITT) analysis in which participants with missing outcome data are excluded. Such an analysis could also be biased; this is often addressed within the domain of 'bias due to missing outcome data. Note that the phrase 'modified' intention-to-treat is employed in several ways and should ask for the inclusion of participants who received a minimum of one dose of treatment; the use of the term refers to missing data instead of adherence to intervention [56]. For the effect of adhering to intervention, appropriate analysis approaches are described by Hernán and

Robins [57]. Instrumental variable approaches are often utilized in some circumstances to estimate the effect of intervention among participants who received the assigned intervention [49]. Thus, the public health researchers clearly write the appropriate analysis method in their protocol during the proposal development process to minimize this type of bias.

5. Minimizing bias due to missing outcome data

The potential explanations for missing outcome data based on the National Research Council 2010 comprise: participants withdraw from the study or can't be located ('loss to follow-up' or 'dropout'); participants do not attend a study visit at which outcomes should have been measured; participants attend a study visit but do not provide relevant data; records or data are unavailable or lost for other reasons; and participants cannot experience the result, for instance, because they are dead [49]. Some participants could also be excluded from the analysis for reasons aside from missing outcome data. For example, a simple 'per-protocol' analysis is limited to subjects who obtained the intended intervention and excludes those who deviated from the intervention protocol [58]. The possible bias created by such analyses or another exclusion of eligible study subjects for whom outcome data are inaccessible was addressed within the domain of 'bias originates due to deviations from intended interventions' [49]. Even when an analysis is described as ITT, it's going to exclude participants with missing outcome data due to different reasons mentioned above and be in danger of bias (such analyses could also be described as 'modified intention-to-treat (mITT) analyses') [56].

When do missing outcome data cause bias?

Whether missing outcome data causes bias in incomplete case analyses depends on whether the missingness mechanism is said to affect the truth value of the result. Equivalently, we will consider whether the measured (non-missing) outcomes differ systematically from the missing outcomes (the true values of participants with missing outcome data) [49,59]. Missing outcome data won't cause bias if missingness within the outcome is unrelated to its true value, within each intervention group, will cause bias if missingness within the outcome depends on both the intervention group and therefore the true value of the outcome, and often cause bias if missingness is said to its true value and, additionally, the effect of the experimental intervention differs from that of the comparator intervention [59,60].

When is the amount of missing outcome data sufficiently small to exclude bias?

It is tempting to categorize the risk of bias based on the proportion of participants who have missing outcome data. Unfortunately, there is no reasonable threshold for defining 'small enough' in terms of the proportion of missing outcome data [60]. On the other hand, when bias results from missing outcome data, the degree of bias will rise as a result of the increased number of missing outcome data [59]. It is customary to classify missing outcome data that is less than 5% as "small" (with associated implications for bias risk) and more than 20% as "large." [59,60].

Missing outcome data measurements may cause bias within the intervention effect size estimate, which means underestimating or overestimating the intervention effect. Besides, a significant number of missing outcome data will decrease the sample size, which in turn decreases the power of the study. During this time, the probability of making a type II error is highly likely, which likely increases the false negative result. To minimize this bias, public health researchers should be careful at the design stage to exert maximum efforts to decrease the missing outcome data of their trial by increasing follow-up to minimize the loss to follow-up or dropout, increasing participants to attend a study visit at which outcomes are measured, encouraging participants to provide relevant data, protecting data or records from loss, and conducting appropriate data analysis. After all these measures and efforts, if there are still significant missing outcome data, the researchers should conduct a comparability of loss to follow between two arms and post-hoc power analysis. If the loss of follow-up comparable between two groups and power is adequate to detect an association between the arms, the researchers will assure their data is not prone to bias due to missing outcome data.

6. Minimizing bias in measurement of the outcome

Errors or mistakes in the measurement of outcome variables can bias intervention effect size estimates [41,49,52,54]. These are often mentioned as measurement error (for continuous outcomes), misclassification (for dichotomous or categorical outcomes), or under-ascertainment or over-ascertainment (for events) [50,53,54]. Measurement errors could also be differential or non-differential about intervention assignment. Differential measurement errors are associated with intervention assignment [49]. Such procedures are systematically not similar between the comparator and experimental intervention arms and are less probable when outcome evaluators are blinded to intervention allocation. Non-differential measurement errors are unrelated to intervention assignment and aren't addressed closely [42,51]. The risk of bias in this domain depends on the subsequent five considerations. First, determine whether the tactic of measuring the result is acceptable. Outcome variables in cRCTs should be measured using proper outcome methods. For example, portable blood sugar machines employed by trial participants might not reliably measure below 3.1 mmol, resulting in an inability to detect differences in rates of severe hypoglycemia between an insulin intervention and placebo and an underrepresentation of the true incidence of this adverse effect. This measurement could be unsuitable for this outcome variable. Second, whether measurement or ascertainment of the result differs, or could differ, between intervention groups. The methods won't measure or ascertain outcomes; they should be equivalent across intervention groups. This is usually the case for pre-specified outcomes, but problems may arise with the passive collection of outcome data, as is usually the case for unexpected adverse effects. Third, who is the outcome assessor? The outcome assessor can be: the participant, when the outcome is a participant-reported outcome such as pain, quality of life, or self-completed questionnaire; the intervention supplier, when the outcome is the effect of a clinical examination, the incidence of a clinical result, or a therapeutic decision like a decision to provide a surgical treatment; an observer not directly involved in the intervention provided to the participant, such as an adjudication committee, or a health professional recording outcomes for inclusion in disease registries. Fourth, whether the result assessor is blinded to intervention assignment status during the data collection period will determine the level of bias introduced. Blinding of outcome assessors is usually possible even when blinding of participants and trial personnel during the trial isn't feasible. However, it's particularly difficult for participant-reported outcomes. Fifth, whether the assessment of outcome is probably going to be influenced by knowledge of the intervention received. For trials during which outcome assessors weren't blinded, the danger of bias will depend on whether the result assessment involves judgment, which depends on the sort of outcome [41,49,52,54].

7. Minimizing bias in selection of the reported result

This domain addresses bias resulting from the reported result being chosen (by the trial authors) among several intervention effect estimates based on its direction, magnitude, or statistical significance. [61]. The distinction between an outcome domain, which is typically a state or endpoint of interest regardless of how it is measured, and an outcome analysis, which is typically a chosen result obtained by analyzing one or more outcome measurements, must be made in order to address the risk of bias [42,51]. The bias in the reported result selection typically stems from the necessity for the results to align with the interests of the researcher or to hold sufficient significance for the publication process [62]. It can arise for both good and bad reasons, though the motivations or interests may differ. For example, in studies comparing a treatment intervention to a placebo, researchers who have a predetermined or conferred interest in presenting that the trial intervention is beneficial and harmless may be selective in reporting efficacy estimates that are statistically significant and positive to the experimental intervention, alongside harm estimates that are not significantly different between groups [62,63]. Other researchers, on the other hand, may selectively report harm estimates that are statistically significant and adverse to the treatment intervention if they believe that publicizing the presence of harm will improve their chances of publishing in a high-impact journal [62]. This domain considers whether the study was analyzed in accordance with a pre-specified protocol plan that was completed prior to the availability of non-blinded outcome data for analysis. We strongly encourage review authors to aim to retrieve the trial's pre-specified analysis intentions. This allows for the identification of any post-hoc outcome measures or analyses that are

omitted or added to the results report [49]. If the study protocol and full statistical analysis plan are not publicly available, review authors should ideally request them from the study authors. Furthermore, if outcome measures and analyses mentioned in a written piece, protocol, or trial registration record are not reported, study authors may be asked to clarify whether those outcome measures were analyzed and, if so, to provide the information. Trial protocols should specify how unexpected adverse outcomes (potentially indicating unanticipated harms) will be collected and analyzed [63]. However, for some trials, the analysis intentions won't be readily available. It is still possible to assess the danger of bias in the selection of the reported result. For example, outcome measures and analyses listed within the methods section of a piece of writing are often compared with those reported. Besides, outcome measure variables and analyses must be compared across diverse articles about the trial. The selective reporting of a particular outcome variable (according to the findings) from among estimates for several outcomes measured within an outcome domain [62]. In 2004, the International Committee of Medical Journal Editors suggested different mechanisms to minimize this problem, like developing a protocol registration framework and database, designing reporting guidelines, and stating the roles and responsibilities of supervisors and sponsor organizations. Despite these efforts, awareness and practice of aligning with these suggestions are poor in public health research that utilizes cRCTs [36].

Trial registration

In 2004, the International Committee of Medical Journal Editors declared that all experimental studies starting after July 1, 2005, must be registered before being considered for publication [64]. However, in public health research, trial protocol registration may occur late or not at all [65]. Health sciences journals have been slow to adopt policies requiring mandatory medical trial registration as a prerequisite for publication [66].

Reporting procedure of CRCTs

The 2010 CONSORT Statement is the most evidence-based and compact set of references for reporting RCT results [67]. The 2010 CONSORT checklist consists of 25 items, including multiple sub-items that address the most prevalent types of randomized controlled trials (RCTs): individually randomized, two-group, and parallel designs [19,68]. The CONSORT extensions checklist has been published and made available for other types of RCT study designs. The 2010 CONSORT Statement, for example, extended to cluster randomized trials [36] as well as the 2010 CONSORT Statement on nonpharmacologic treatment interventions [69].

The increased use of cRCTs necessitates the need for reliable, consistent, high-quality, and valid reporting. The first CONSORT Statement extension to cRCTs was published and released in 2004 [36], and it was updated and reorganized in 2010 [70]. These should serve as useful guides for CRCT readers, designers, and reviewers. A separate, currently published and available review of 300 cRCT samples investigated the impact of the 2010 CONSORT Statement extension on overall method quality and inferred or concluded that compliance with published and currently available reporting guidelines, recommendations, and quality remains extremely low [71]. Similarly, the same review of cRCT reporting guideline adherence have been carried out and come up with equivalent inferences from the above review [72].

Considerations during the sample size calculation

Accounting for the effect of clustering

Researchers that design to use cRCTs to evaluate intervention effects on outcomes will use cluster randomization techniques to allocate study subjects to the comparator and intervention arms. Due to this design, there are two sources of variation in the measurements. The variability of study subjects within a cluster and between clusters. These two sources of variation combine to yield a rise in the variance as compared to the circumstances of individual randomization and must be measured and taken into account during the sample size calculation. Since the sample size required is directly

related to the variance, the sample size was calculated for an individual-level randomized trial need to be multiplied by the variance inflation factor (VIF) [73]. Failure to calculate the VIF and account for it will lead to a decrease in study statistical power, which will also decrease the capability of detecting the actual effect of the intervention or make it more susceptible to type II error [74]. A review conducted to evaluate the methodology and quality of sample size computations in cRCTs among a sample of three hundred cRCTs published reported that merely 166 (55.3%) presented a clear and simple sample size computation formula, of which merely 102 (34%) measured and accounted properly for clustering effects [75]. In response to this low practice of accounting for the effects of clustering, this article provides detailed information on assumptions, appropriate software for sample size calculation, and pertaining sample sizes for individual and cluster-level randomized trials in public health research.

Scenario one: What do the researchers do to calculate the sample size if there is a lack of previous seminar studies on the research topic?

In this case, this article suggests the researchers use the rule of thumb or conservative approach to estimate the minimum required sample size. Suppose that a researcher wants to test the effects of egg-based, locally prepared food to improve moderate acute malnutrition (MAM) in a certain area, but there have been no previous similar studies. Therefore, by the rule of thumb or conservative approach, take 50% of children as MAM. Consider this proportion as a proportion of children who have MAM without any intervention (control group). Also assume that your intervention will decrease MAM among the intervention group by 15%, so that the proportion of children in the intervention group will be 35%. Now you have at least basic information about the proportion and fixed power of your study at 80%, level of confidence at 95%, and comparator-to-experimental group ratio at 1. Now you have all the basic information you need to calculate the sample size for an individual-level randomized trial (IRT). This article uses OpenEpi version 3.01 to calculate sample size, but you can use any sample size calculation tool, such as Epi Info, G-Power, WHO EVM calculator, PharmaSchool sample size calculator, etc. Based on the above assumption, the estimated sample size for IRT was 368 for both arms (184 in the intervention group and 184 in the comparator group). This sample size needs to be adjusted for loss to follow-up by adding 15% (you can add loss to follow-up from 10–30% based on the nature of your study topic, liability to loss to follow-up, and sensitivity of your research topic). The sample size adjusted for loss to follow-up was 423. We will calculate VIF to increase sample size (N) to account for the effect of clustering. It is calculated by using the following formula:

- $VIF = 1 + (m-1) \times ICC$
- m = average number of study subjects per cluster
- ICC = intra-cluster correlation

From where do we obtain the ICC to calculate the VIF? We obtain ICC from previous studies reports, by conducting pilot studies, and by the conservation approach. This article receives 0.05 by rule of thumb or conservative approach based on the recommendation of Donner and Allan from ranges of ICC values (0.01–0.05).[73].

How do you calculate the number of clusters needed for your trial?

Minimal cluster number required = the calculated effective sample size based on IRT multiplied by ICC = $423 \times 0.05 = 21.15$. You can increase the number of clusters as you like if you want to increase cluster sufficiency and ensure adequate power of study to detect the intervention effect after you determine the minimum needed clusters for your study. This article considers 30 clusters (15 in the intervention group and 15 in the comparator group).

How do you calculate the average number of study subjects per cluster needed for your trial?

If you assume an equal number of study subjects in each cluster, you can easily calculate it by dividing the calculated effective sample size by the determined number of clusters in your trial: $423/30 = 14.1$. Now we have all the information needed to calculate the VIF and insert the needed information into the formula mentioned above. $VIF = 1 + (m-1) \times ICC = 1 + (14.1-1) \times 0.05 = 1.655$. The final estimated sample size after accounting for the effect of clustering was 700 for both groups (350 in the intervention group and 350 in the comparator group).

Scenario two: What do the researchers do to calculate the sample size if there is a previous national survey report or small pocket cross-sectional seminar studies on the research topic?

Let us assume a proportion of MAM of 47% in a previous national survey report or small-pocket cross-sectional seminar study. In this case, this article suggests the researchers take 47% of children as MAM. Consider this proportion as a proportion of children who have MAM without any intervention (control group). Also assume that your intervention will decrease MAM among the intervention group by 15%, so that the proportion of children in the intervention group will be 32%. Now you have at least basic information about the proportion and fixed power of your study at 80%, level of confidence at 95%, and comparator-to-experimental group ratio at 1. Now you have all the basic information you need to calculate the sample size for an individual-level randomized trial (IRT). Then, follow the same aforementioned sample size calculation procedure to calculate the minimum required sample size in this scenario.

Scenario three: What do the researchers do to calculate the sample size if there is a previous seminar quasi-experimental study on the research topic?

Let us assume a proportion of MAM before intervention was 45% and a proportion of MAM after intervention was 30% in a previous seminar quasi-experimental study. In this case, this article suggests the researchers take 45% of children as a proportion of MAM in the comparator group. Also assume that 30% of children are MAMs in the intervention group due to a lack of previous cRCTs on a similar topic. Now you have at least basic information about the proportion and fixed power of your study at 80%, level of confidence at 95%, and comparator-to-experimental group ratio at 1. Now you have all the basic information you need to calculate the sample size for an individual-level randomized trial (IRT). Then, follow the same aforementioned sample size calculation procedure to calculate the minimum required sample size in this scenario.

Last scenario: What do the researchers do to calculate the sample size if there is a previous cRCT on the research topic?

This article suggests researchers follow the same aforementioned sample size calculation procedure to calculate the minimum required sample size in this scenario by taking the required information from the previous cRCTs.

Considerations during the data analysis

The cRCTs should consider robust methods of data analysis to provide a precise result by accounting for both the effects of confounders and clustering. A review conducted to evaluate the methodology quality of cRCTs among a sample of three hundred cRCTs published reported that merely 96 (32%) properly accounted for clustering and the effects of confounders during data analysis [75]. Thus, in this section, this article discusses how to account for the effects of clustering and confounders during data analysis.

The type of data analysis is most frequently determined by outcome variables. There are three types of variables, such as categorical, numeric discrete, and numeric continuous, in cRCTs outcomes. A chi-square test is used for categorical outcomes to test the effect of intervention between two groups in non-adjusted analysis, while an independent t-test and an ANOVA are used for numeric outcome data. The linear mixed models (LMM), generalized linear mixed models (GLMM), and generalized estimation equations (GEE) are used to adjust for the effects of confounders and clustering in cRCTs. Even though LMM is more powerful than GLMM and GEE, it may still be ineffective if the outcome data is not normally distributed and the cluster size fluctuates. When any of these assumptions are violated, GLMM and GEE must be applied.

Table 1. Summary of the model used to analyze cRCTs outcome data.

S.no	Type of outcome data	Appropriate model
1	Numeric continuous (assumption fulfilled)	Multi-level mixed effects linear regression
2	Categorical (if outcome of interest is less than 20%)	Multi-level mixed effects logistic regression

3	Categorical (if outcome of interest is greater than 20%)	Multi-level mixed effects modified Poisson regression with robust standard error Multi-level mixed effects log-binomial regression
4	Count (equi-dispersed)	Multi-level mixed effects Poisson regression
5	Count (over dispersed)	Multi-level mixed effects negative binomial regression
6	Count (under dispersed)	Multi-level mixed effects quasi-Poisson regression
7	Count (excess zero)	Multi-level mixed effects zero-inflated regression

Practical application

Most of the time, the baseline individual and cluster-level covariates are expected to be balanced or comparable between the two arms in carefully designed cRCTs. However, this doesn't work for some time, so an imbalance occurs between the two arms. Thus, this review article suggests researchers should check the baseline comparability between two arms and adjust any significant imbalanced covariates between two arms using multivariate analysis. Similarly, we should be able to adjust for the effect of clustering using an appropriate multi-level regression model. In the following section, this article demonstrates the practical aspects using a data set from previously published papers elsewhere [76]. Table 2 demonstrates covariates that are significantly imbalanced between two arms in the sample data set. Let us see what happens when we ignore to account for the effects of confounders and clustering during data analysis.

Table 2. Imbalanced covariates between the two arms (N = 1,070).

Variables	Intervention group	Control group	Total	P- value
Women's occupation status				0.001
Housewife	386 (71.5)	409 (77.2)	795 (74.3)	
Farmer	12 (2.2)	37 (7.0)	49 (4.6)	
Government employee	71 (13.1)	41 (7.7)	112 (10.5)	
Merchant	71 (13.1)	43 (8.1)	114 (10.7)	
Husband occupation status				0.001
Government employee	77 (14.3)	40 (7.5)	117 (10.9)	
Merchant	299 (55.4)	247 (46.6)	546 (51.0)	
Farmer	164 (30.4)	243 (45.8)	407 (38.0)	
Use of mass media				0.001
No	234 (43.3)	299 (56.4)	533 (49.8)	
Yes	306 (56.7)	231 (43.6)	537 (50.2)	
Wealth quintile				0.001
Lowest	131 (24.3)	82 (15.5)	213 (19.9)	
Second	77 (14.3)	138 (26.0)	215 (20.1)	

Middle	88 (16.3)	126 (23.8)	214 (20.0)
Fourth	113 (20.9)	101 (19.1)	214 (20.0)
Highest	131 (24.3)	83 (15.7)	214 (20.0)

Note: The p-value in this table is based on the chi-square test.

The effect of intervention on outcome is highly significant in unadjusted analysis ($X^2 = 19.51$, $p = 0.01$). Nevertheless, after controlling for confounders and clustering, the effect of intervention was insignificantly different between the two arms (adjusted risk ratio: 1.16; 99% CI: 0.88–1.49) (Table 3). This finding indicates that several studies are reporting wrong results that, cumulatively, after meta-analysis, will lead to country policy change. Thus, researchers in the public health discipline should carefully analyze their data to avoid drawing wrong results and conclusions.

Table 3. Confounders and clustering adjusted analysis (N = 1,070).

Variables	Postnatal care utilization		CRR (99% CI)	ARR (99% CI)
	Yes	No		
Group				
Interventional	353 (65.4)	187 (34.6)	1.26 (1.04, 1.54)	1.16 (0.88, 1.49)
Comparator	276 (52.1)	254 (47.9)	1	1

Variables adjusted in the models were women's and husband occupation, mass media use, wealth index, history of the previous stillbirth, husband's attitude about maternal health service use, woman's decision-making power, women's score on attitude towards postnatal care, women's perceived quality of postnatal care score, birth preparedness plan, women's knowledge of obstetric danger signs and community-level women's literacy.

This article reported a 99% confidence interval (CI) instead of a 95% CI. What do you think the reason is?

Most of the time, researchers conduct research to assess the effect of a single health intervention or treatment on multiple health outcomes. During this time, a predetermined level of significance should be adjusted to decrease the probability of committing a type I error. There are different formulas suggested by different statisticians to adjust the predetermined level of significance to avoid a multiple outcome comparison effect on a single intervention. The statistical significance was fixed at a p-value of < 0.05 to reject the null hypothesis for this article. To adjust for type I error inflation, which can result from the influence of multiple comparisons or testing problems in a single intervention, this statistical significance limit was adjusted. This article makes adjustments using the Bonferroni correction formula. By dividing the predefined significance level by the total number of statistical tests compared for a single intervention, the adjusted significance level was computed. In this article, the adjusted level of significance is $0.05/5 = 0.01$ because this article compared a single intervention effect on five outcome variables. When the p-value was less than 0.01, the association was accepted as statistically significant [77,78]. Thus, adjusted risk ratios with 99% CIs that did not contain 1 were used to declare a statistically significant association between the intervention and outcome variables. However, a review conducted to evaluate the methodology quality of cRCTs reported that merely a few trials properly calculated and accounted for the effects of multiple comparison problems during data analysis [75]. Thus, this article strongly suggests researchers calculate the multiple comparisons effect because it inflates type I error, which frequently lead to wrong conclusions.

Model performance

In this section, this article demonstrates the performance of the difference model to estimate the effect size of the intervention outcome. As aforementioned above in this article, the study outcome variable is determined by choosing an appropriate statistical model to calculate the precise and unbiased effect size estimators. However, this is not merely a reason to choose an appropriate model; there are a number of determinants and factors. This article demonstrates an adjusted effect size estimate from three different models with a binary outcome variable.

Table 4. Multi-level mixed effects logistic regression.

Variables	Antenatal care use		COR (95% CI)	AOR (95% CI)
	Yes	No		
Intervention group				
Comparator	145 (17.6)	679 (82.4)	Ref	Ref
Interventional	149 (48.7)	157 (51.3)	11.58 (2.45, 54.68)	14.32 (3.75, 54.68)

Variables adjusted in the models were women’s and husband occupation, mass media use, wealth index, history of the previous stillbirth, husband’s attitude about maternal health service use, woman’s decision-making power, women’s score on attitude towards antenatal care, women’s perceived quality of antenatal care score, birth preparedness plan, women's knowledge of obstetric danger signs and community-level women’s literacy.

Table 5. Multi-level mixed effects modified Poisson regression with robust standard error.

Variables	Antenatal care use		CRR (99% CI)	ARR (99% CI)
	Yes	No		
Intervention group				
Comparator	145 (17.6)	679 (82.4)	Ref	Ref
Interventional	149 (48.7)	157 (51.3)	11.58 (2.45, 54.68)	3.51 (2.16, 5.77)

Variables adjusted in the models were women’s and husband occupation, mass media use, wealth index, history of the previous stillbirth, husband’s attitude about maternal health service use, woman’s decision-making power, women’s score on attitude towards antenatal care, women’s perceived quality of antenatal care score, birth preparedness plan, women's knowledge of obstetric danger signs and community-level women’s literacy.

Table 6. Multi-level mixed effects log-binomial regression.

Variables	Antenatal care use		CRR (99% CI)	ARR (99% CI)
	Yes	No		
Intervention group				
Comparator	145 (17.6)	679 (82.4)	Ref	Ref
Interventional	149 (48.7)	157 (51.3)	11.58 (2.45, 54.68)	4.79 (2.69, 5.53)

Variables adjusted in the models were women’s and husband occupation, mass media use, wealth index, history of the previous stillbirth, husband’s attitude about maternal health service use, woman’s decision-making power, women’s score on attitude towards antenatal care, women’s perceived quality of antenatal care score, birth preparedness plan, women's knowledge of obstetric danger signs and community-level women’s literacy.

Which model do you select to report your findings?

This article rejects multi-level mixed effects logistic regression analysis for the following reasons: The prevalence of outcomes is greater than 20% (in this case, the utilization of antenatal care was 52%). If the prevalence of outcome is greater than 20%, multi-level mixed effects logistic regression is not an appropriate model due to its overestimated effect size, lower level of confounder control, and adjusted odds ratio, which is not easy to understand and interpret for non-epidemiologists or non-professionals [79–81].

This article rejects multi-level mixed effects log-binomial regression analysis for the following reasons: This model has a higher value of Akaike's information criteria (AIC), Bayesian information criteria (BIC), and log-likelihood as compared to multilevel mixed effects modified Poisson regression with robust variance. The best-fitting model is selected based on low AIC, BIC, and log-likelihood values during model performance analysis [82].

Because of its comparatively small AIC and BIC values, the multilevel mixed-effects modified Poisson regression with robust standard error best fits the model for the data in this article. Furthermore, compared to the log-binomial regression model, modified Poisson regression with robust standard error yields less biased estimates of the ARR for the medium sample size and highly common outcomes of interest [81]. This information led to the use of multilevel mixed effects modified Poisson regression in this article, and the results are interpreted as follows. Intervention has significantly improved antenatal care utilization (ARR: 3.51; 99% CI: 2.16-5.77) after accounted for effect of confounders and clustering.

Intervention effect modification assessment

The most frequently neglected analysis of cRCTs research in public health is intervention effect modification assessment. This article demonstrates how to analyze and report effect modification results. This article entered the interaction terms in the final Poisson regression with a robust standard error model for women's occupation and intervention status, husband's occupation and intervention status, mass media use and intervention status, wealth index and intervention status, and model family training and intervention status to see if women's occupation, husband's occupation, mass media use, wealth index, and model family training modifies the effect of intervention. It is implied that there was no significant effect modification because none of the interaction terms were statistically significant in the final model. However, the urban residence modified the intervention effect by 1.2, and this article effect size estimate increased from ARR = 3.51 to ARR = 4.71.

Influence of missing outcome data

This article highlighted the measures and efforts to minimize missing outcome data at the design stage. Despite these efforts and measures, the missing outcome data might happen in many trials due to different reasons. This article will demonstrate methods for handling missing outcome data because the data used in this article also contains missing outcome data. First, this article determined the proportion of missing outcomes data and found only 4.8% lost to follow-up, which is small or produces a low risk of bias as per the suggestion of a risk of bias assessment tool for cRCTs. Second, this article checked the comparability of loss to follow-up between two arms, and in both groups, the percentage of subjects lost to follow-up was similar (5.87% in the comparator group and 4.98% in the intervention group). Third, a post-hoc analysis of power in this article revealed that statistical power was 100%, meaning it's adequate to determine the intervention's effects. The public health researchers should be careful to assess the effect of missing outcome data in cRCTs.

Random-effect model results

This article calculated ICC by utilizing the intercept-only multi-level binary logistic model; the ICC value revealed that 29.65% of the variability in using ANC was explained by membership in clusters. Also, the median prevalence ratio value was 2.5, which means the utilization of antenatal

care varies 2.5 times when we select two areas at low and high risk of antennal care use in study settings.

Conclusion

Some interventions or treatments are appropriate to provide at group levels. This is termed cluster assignment or cluster randomization and is the most common practice in health services and public health interventions. The observations or measurements within a cluster tend to be more similar than measurements chosen completely at random. This disobeys the assumption of individuality or independence, which is at the core of common techniques of hypothesis testing and statistical estimation. Unable to measure and account for the dependence among individual measurements and the group to which they are included can have thoughtful effects on the design and analysis of such trials. Furthermore, sample size estimates will be too small, CIs too narrow, and p-values too small, which frequently lead to inaccurate results and wrong conclusions. Thus, the objective of this article review was to provide a fundamental concept of how to design and analyze efficient cRCTs in public health research, particularly in LMIC, to minimize all aforementioned problems. It provided detailed directions and examples of techniques for designing and analyzing clustered data and sample size calculations while planning trials. This article also provided a basic concept of how to select the best-fit model for a particular data set analysis in public health research.

Author Contributions: AY: Conceptualized, ensured data curation, did the formal analysis, and wrote the manuscript.

Acknowledgments: My greatest thanks go to Netsanet Kibru for her big support during the preparation of this article by giving her ideas.

Conflicts of Interest: Not applicable.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: All data generated or analyzed during this study are included in this published article and its supplementary information files.

List of abbreviations and acronyms

AIC	Akaike’s information criteria
AOR	Adjusted odds ratio
ARR	Adjusted risk ratio
BIC	Bayesian information criteria
COR	Crude odds ratio
CRR	Crude risk ratio
CSA	Central statistics agency
CIs	Confidence intervals
CONSORT	Consolidated standards of reporting trial
cRCTs	Cluster randomized controlled trials
GEE	Generalized estimation equations
GLMM	Generalized linear mixed models
ICC	Intra-cluster correlation
IRT	Individual randomized trial
ITT	Intention-to-treat
LMIC	Low and middle income countries
LMM	Linear mixed models
MAM	Moderate actuate malnutrition
mITT	Modified intention-to-treat
RCTs	Randomized controlled trials
SDT	Step-wedge design
VIF	Variance inflation factor
WHO	World Health Organization

References

1. Claybaugh, Z., *Research Guides: Organizing Academic Research Papers: Types of Research Designs*. library.sacredheart.edu. Retrieved 2020-10, 2020. **28**.
2. Parab, S. and S. Bhalerao, *Study designs*. International journal of Ayurveda research, 2010. **1**(2): p. 128.
3. Belbasis, L. and V. Bellou, *Introduction to epidemiological studies*. Genetic epidemiology: methods and protocols, 2018: p. 1-6.
4. Ranganathan, P. and R. Aggarwal, *Study designs: Part 1—An overview and classification*. Perspectives in clinical research, 2018. **9**(4): p. 184.
5. Checkoway, H., N. Pearce, and D. Kriebel, *Selecting appropriate study designs to address specific research questions in occupational epidemiology*. Occupational and environmental medicine, 2007. **64**(9): p. 633-638.
6. Omair, A., *Selecting the appropriate study design for your research: Descriptive study designs*. Journal of health specialties, 2015. **3**(3): p. 153.
7. Howick, J., *Introduction to study design*. URL: <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-study-design-april-20131.pdf> [accessed 2017-04-12][WebCite Cache ID 6pfDD1ddH], 2002.
8. Chalmers, T.C., et al., *A method for assessing the quality of a randomized control trial*. Controlled clinical trials, 1981. **2**(1): p. 31-49.
9. Littlejohns, P., et al., *National Institute for Health and Care Excellence, social values and healthcare priority setting*. Journal of the Royal Society of Medicine, 2019. **112**(5): p. 173-179.
10. Hannan, E.L., *Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations*. JACC: Cardiovascular Interventions, 2008. **1**(3): p. 211-217.
11. Bland, J.M., *Cluster randomised trials in the medical literature: two bibliometric surveys*. BMC medical research methodology, 2004. **4**: p. 1-6.
12. Wears, R.L., *Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data*. Academic emergency medicine, 2002. **9**(4): p. 330-341.
13. Jones, B.G., et al., *Bayesian statistics in the design and analysis of cluster randomised controlled trials and their reporting quality: a methodological systematic review*. Systematic Reviews, 2021. **10**(1): p. 1-14.
14. Murray, D.M., S.P. Varnell, and J.L. Blitstein, *Design and analysis of group-randomized trials: a review of recent methodological developments*. American journal of public health, 2004. **94**(3): p. 423-432.
15. Murray, D.M., *Design and analysis of group-randomized trials*. Vol. 29. 1998: Monographs in Epidemiology and.
16. Giuffrida, M.A., K.A. Agnello, and D.C. Brown, *Blinding terminology used in reports of randomized controlled trials involving dogs and cats*. Journal of the American Veterinary Medical Association, 2012. **241**(9): p. 1221-1226.
17. Frias, J., et al., *Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial*. Journal of medical Internet research, 2017. **19**(7): p. e246.
18. Moberg, J. and M. Kramer, *A brief history of the cluster randomised trial design*. Journal of the Royal Society of Medicine, 2015. **108**(5): p. 192-198.
19. Schulz KF, A.D., (Moher D; for the CONSORT Group), , "CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials". Br Med J. 340: c332. Doi:10.1136/bmj.c332. 2010.
20. Hayes R, M.L., *Cluster randomised trials*. Chapman and Hall/CRC Press , Boca Raton , FL , 2009. 2009.
21. Centre for Evidence-Based Medicine, *Study Designs*. 2016. [Last accessed on 2021 June 22]. Available online from: <https://www.cebm.net/2014/04/study-designs/>. 2016.
22. Vonesh, E.F.C., Vernon G, , "Crossover Experiments". Linear and Nonlinear Models for the Analysis of Repeated Measurements. London: Chapman and Hall. pp. 111–202. 1997.
23. Jones, B.K., Michael G, , *Design and Analysis of Cross-Over Trials (Second ed.)*. London: Chapman and Hall. 2003.
24. Gambia, T., *Study h. The gambia hepatitis intervention study. The gambia hepatitis study group*. Cancer Res, 1987. **47**: p. 5782-7.
25. Mdege, N.D., et al., *Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation*. Journal of clinical epidemiology, 2011. **64**(9): p. 936-948.
26. Hussey, M.A. and J.P. Hughes, *Design and analysis of stepped wedge cluster randomized trials*. Contemporary clinical trials, 2007. **28**(2): p. 182-191.
27. Mulfinger, N., et al., *Cluster-randomised trial evaluating a complex intervention to improve mental health and well-being of employees working in hospital—a protocol for the SEEGEN trial*. BMC public health, 2019. **19**: p. 1-16.
28. Brown, C.A. and R.J. Lilford, *The stepped wedge trial design: a systematic review*. BMC medical research methodology, 2006. **6**(1): p. 1-9.

29. Woertman, W., et al., *Stepped wedge designs could reduce the required sample size in cluster randomized trials*. Journal of clinical epidemiology, 2013. **66**(7): p. 752-758.
30. Hemming, K., et al., *The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting*. Bmj, 2015. **350**.
31. Juszczak, E., et al., *Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement*. Jama, 2019. **321**(16): p. 1610-1620.
32. Klar, N., *A Review of: "Cluster Randomised Trials, by RJ Hayes and LH Moulton"*. Boca Raton, FL: Chapman & Hall/CRC, 2009, ISBN 978-1-58488-816-1, xxii+ 315 pp., \$89.95. 2009, Taylor & Francis.
33. Edwardson, C.L., et al., *Effectiveness of an intervention for reducing sitting time and improving health in office workers: three arm cluster randomised controlled trial*. bmj, 2022. **378**.
34. Peters, T., et al., *Comparison of methods for analysing cluster randomized trials: an example involving a factorial design*. International journal of epidemiology, 2003. **32**(5): p. 840-846.
35. Edwards, S.J., et al., *Ethical issues in the design and conduct of cluster randomised controlled trials*. Bmj, 1999. **318**(7195): p. 1407-1409.
36. MK, C., *CONSORT statement: extension to cluster randomised trials*. BMJ, 2004. **328**: p. 48-64.
37. Eldridge, S. and S. Kerry, *A practical guide to cluster randomised trials in health services research*. Vol. 120. 2012: John Wiley & Sons.
38. Stel VS, J.K., Zoccali C, Wanner C, Dekker FW,, *The randomized clinical trial: an unbeatable standard in clinical research?* Kidney Int. 2007 Sep;72(5):539-42. doi: 10.1038/sj.ki.5002354. Epub 2007 Jun 27. PMID: 17597704. 2007.
39. Stel VS, Z.C., Dekker FW, Jager KJ,, *The randomized controlled trial*. Nephron Clin Pract. 2009;113(4):c337-42. doi: 10.1159/000237143. Epub 2009 Sep 11. PMID: 19752576. 2009.
40. JP., V., *When are observational studies as credible as randomized trials?* Lancet. 2004 May 22; 363(9422):1728-31. doi: 10.1016/S0140-6736(04)16261-2. PMID: 15158638. 2004.
41. Hemming K, E.S., Forbes G, Weijer C, Taljaard M. , *How to design efficient cluster randomized trials*. BMJ. 2017 Jul 14; 358:j3064. doi: 10.1136/bmj.j3064. PMID: 28710062; PMCID: PMC5508848. 2017.
42. Brierley G, B.S., Torgerson D, Watson J,, *Bias in recruitment to cluster randomized trials: a review of recent publications*. J Eval Clin Pract. 2012 Aug;18(4):878-86. doi: 10.1111/j.1365-2753.2011.01700.x. Epub 2011 Jun 20. PMID: 21689213. 2012.
43. Cumpston M, L.T., Page MJ, Chandler J, Welch VA, Higgins JP, Thomas J. , *Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Database Syst Rev. 2019 Oct 3;10:ED000142. doi: 10.1002/14651858.ED000142. PMID: 31643080. 2019.
44. Puffer S, T.D., Watson J. , *Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals*. BMJ. 2003 Oct 4;327(7418):785-9. doi: 10.1136/bmj.327.7418.785. PMID: 14525877; PMCID: PMC214092. 2003.
45. Diehr P, M.D., Koepsell T, Cheadle A,, *Breaking the matches in a paired t-test for community interventions when the number of pairs is small*. Stat Med. 1995 Jul 15;14(13):1491-504. doi: 10.1002/sim.4780141309. PMID: 7481187. 1995.
46. Ivers NM, H.I., Barnsley J, Grimshaw JM, Shah BR, Tu K, Upshur R, Zwarenstein M. , *Allocation techniques for balance at baseline in cluster randomized trials: a methodological review*. Trials. 2012 Aug 1;13:120. doi: 10.1186/1745-6215-13-120. PMID: 22853820; PMCID: PMC3503622. 2012.
47. LH., M., *Covariate-based constrained randomization of group-randomized trials*. Clin Trials. 2004;1(3):297-305. doi: 10.1191/1740774504cn024oa. PMID: 16279255. 2004.
48. Lorenz E, G.S., *Covariate-constrained randomization routine for achieving baseline balance in cluster-randomized trials*. The Stata Journal. 2017;17(2):503-10. 2017.
49. Higgins J, *Cochrane handbook for systematic reviews of interventions*. Version 5.1. 0 [updated March 2011]. The Cochrane Collaboration. www cochrane-handbook org. 2011. 2011.
50. Hahn S, P.S., Torgerson DJ, Watson J,, *Methodological bias in cluster randomised trials*. BMC Med Res Methodol. 2005 Mar 2;5:10. doi: 10.1186/1471-2288-5-10. PMID: 15743523; PMCID: PMC554774. 2005.
51. Eldridge S, K.S., Torgerson DJ,, *Bias in identifying and recruiting participants in cluster randomised trials: what can be done?* BMJ. 2009 Oct 9;339:b4006. doi: 10.1136/bmj.b4006. PMID: 19819928. 2009.
52. Campbell MK, E.D., Altman DG; CONSORT group,, *CONSORT statement: extension to cluster randomised trials*. BMJ. 2004 Mar 20;328(7441):702-8. doi: 10.1136/bmj.328.7441.702. PMID: 15031246; PMCID: PMC381234. . 2004.
53. Haahr MT, H.A., *Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors*. Clin Trials. 2006;3(4):360-5. doi: 10.1177/1740774506069153. PMID: 17060210. 2006.
54. Boutron I, E.C., Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, Ravaud P. , *Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review*. PLoS Med. 2006 Oct;3(10):e425. doi: 10.1371/journal.pmed.0030425. PMID: 17076559; PMCID: PMC1626553. 2006.

55. Gravel J, O.L., Shapiro S., *The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say?* Clin Trials. 2007;4(4):350-6. doi: 10.1177/1740774507081223. PMID: 17848496. 2007.
56. Abraha I, M.A., *Modified intention to treat reporting in randomised controlled trials: systematic review.* BMJ. 2010 Jun 14;340:c2697. doi: 10.1136/bmj.c2697. PMID: 20547685; PMCID: PMC2885592. 2010.
57. Hernán MA, R.J., *Per-Protocol Analyses of Pragmatic Trials.* N Engl J Med. 2017 Oct 5;377(14):1391-1398. doi: 10.1056/NEJMs1605385. PMID: 28976864. 2017.
58. Hernán MA, H.-D.S., *Beyond the intention-to-treat in comparative effectiveness research.* Clin Trials. 2012 Feb; 9(1):48-55. doi: 10.1177/1740774511420743. Epub 2011 Sep 23. PMID: 21948059; PMCID: PMC3731071. 2012.
59. Higgins JPT, W.I., Wood AM., *Imputation methods for missing outcome data in meta-analysis of clinical trials.* Clinical Trials 2008; 5: 225-239. 2008.
60. Bell ML, F.M., Horton NJ, Hsu CH., *Handling missing data in RCTs; a review of the top medical journals.* BMC Med Res Methodol. 2014 Nov 19;14:118. doi: 10.1186/1471-2288-14-118. PMID: 25407057; PMCID: PMC4247714. 2014.
61. Kirkham JJ, D.K., Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR., *The impact of outcome reporting bias in randomized controlled trials on a cohort of systematic reviews.* BMJ. 2010 Feb 15; 340:c365. doi: 10.1136/bmj.c365. PMID: 20156912. 2010.
62. Page MJ, H.J., *Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study.* Syst Rev. 2016 Jul 8;5(1):108. doi: 10.1186/s13643-016-0289-2. PMID: 27392044; PMCID: PMC4938957. 2016.
63. Mansournia MA, H.J., Sterne JA, Hernán MA., *Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists.* Epidemiology. 2017 Jan;28(1):54-59. doi: 10.1097/EDE.0000000000000564. Erratum in: Epidemiology. 2018 Sep;29(5):e49. PMID: 27748683; PMCID: PMC5130591. 2017.
64. De Angelis C, *"Clinical trial registration: a statement from the International Committee of Medical Journal Editors".* The New England Journal of Medicine. 351 (12): 1250–1. Doi:10.1056/NEJMe048225. 2004.
65. Mathieu S, B.I., Moher D, Altman DG, Ravaud P., *"Comparison of registered and published primary outcomes in randomized controlled trials".* JAMA. 302 (9): 977–84. Doi:10.1001/jama.2009.1242. 2009.
66. Bhaumik, S., *"Editorial policies of MEDLINE indexed Indian journals on clinical trial registration".* Indian Pediatr. 50 (3): 339–40. doi:10.1007/s13312-013-0092-2. 2013.
67. Hollis S, C.F., *"What is meant by intention to treat analysis? Survey of published randomised controlled trials".* Br Med J. 319 (7211): 670–4. doi:10.1136/bmj.319.7211.670. 1999.
68. CONSORT Group, *"Welcome to the CONSORT statement Website".* Retrieved 2021-06-29.
69. Boutron I, M.D., Altman DG, Schulz K, Ravaud P *"Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: explanation and elaboration".* Annals of Internal Medicine. 148 (4): 295–309. Doi:10.7326/0003-4819-148-4-200802190-00008. 2008.
70. Equator Network, *Consort 2010 statement: extension to cluster randomised trials.* Available online from <https://www.equator-network.org/reporting-guidelines/consort-cluster/>. 2010.
71. Ivers, N., et al., *Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8.* Bmj, 2021. **343**.
72. Tokolahi, E., et al., *Quality and reporting of cluster randomized controlled trials evaluating occupational therapy interventions: a systematic review.* OTJR: occupation, participation and health, 2016. **36**(1): p. 14-24.
73. Donner, A., N. Birkett, and C. Buck, *Randomization by cluster: sample size requirements and analysis.* American journal of epidemiology, 1981. **114**(6): p. 906-914.
74. Higgins, J.P., et al., *Assessing risk of bias in a randomized trial.* Cochrane handbook for systematic reviews of interventions, 2019: p. 205-228.
75. Rutterford C, T.M., Dixon S, Copas A, Eldridge S., *Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review.* J Clin Epidemiol. 2015; 68(6):716–23. <https://doi.org/10.1016/j.jclinepi.2014.10.006>. 2015.
76. Yoseph, A., et al., *Effect of Community-Based Health Education Led by Women's Groups on Mothers' Knowledge of Obstetric Danger Signs and Birth Preparedness and Complication Readiness Practices in Southern Ethiopia: A Cluster Randomized Controlled Trial.* 2023.
77. Bland JM, A.D., *Multiple significance tests: the Bonferroni method.* BMJ. 1995 Jan 21;310(6973):170. doi: 10.1136/bmj.310.6973.170. . 1995.
78. Hsu JC, *Multiple comparisons: theory and methods.* London: Chapman & Hall: CRC Press, 1996. 1996.
79. Schmidt, C.O. and T. Kohlmann, *When to use the odds ratio or the relative risk?* Int J Public Health, 2008. **53**(3): p. 165-7.
80. Barros, A.J. and V.N. Hirakata, *Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio.* BMC Med Res Methodol, 2003. **3**: p. 21.

81. Petersen, M.R. and J.A. Deddens, *A comparison of two methods for estimating prevalence ratios*. BMC medical research methodology, 2008. **8**(1): p. 1-9.
82. Dziak, J.J., et al., *Sensitivity and specificity of information criteria*. Brief Bioinform, 2020. **21**(2): p. 553-565.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.