5-1-2021

The role and challenges of cluster randomised trials for global health

Louis Dron

Monica Taljaard

Yin Bun Cheung

Rebecca Grais

Nathan Ford

See next page for additional authors

Authors

Louis Dron, Monica Taljaard, Yin Bun Cheung, Rebecca Grais, Nathan Ford, Kristian Thorlund, Fyezah Jehan, Etheldreda Nakimuli-Mpungu, Denis Xavier, Zulfiqar Ahmed Bhutta, Jay J H. Park, and Edward J. Mills

Clinical Trials in Global Health 3

The role and challenges of cluster randomised trials for global health

Louis Dron, Monica Taljaard, Yin Bun Cheung, Rebecca Grais, Nathan Ford, Kristian Thorlund, Fyezah Jahan, Etheldreda Nakimuli-Mpungu, Denis Xavier, Zulfiqar A Bhutta, Jay J H Park, Edward J Mills

Evaluating whether an intervention works when trialled in groups of individuals can pose complex challenges for clinical research. Cluster randomised controlled trials involve the random allocation of groups or clusters of individuals to receive an intervention, and they are commonly used in global health research. In this paper, we describe the potential reasons for the increasing popularity of cluster trials in low-income and middle-income countries. We also draw on key areas of global health research for an assessment of common trial planning practices, and we address their methodological shortcomings and pitfalls. Lastly, we discuss alternative approaches for population-level intervention trials that could be useful for research undertaken in low-income and middle-income countries for situations in which the use of cluster randomisation might not be appropriate.

Lancet Glob Health 2021; 9: e701–10

See **Comment** page e575

This is the third in a **Series** of four papers about clinical trials in global health

All papers in the **Series** are available at www.thelancet.com/series/clinical-trials-global-health

Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada (L Dron MSc, K Thorlund PhD, Prof E J Mills FRCP); **Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada** (M Taljaard PhD); **School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada** (M Taljaard); **Centre for Quantitative Medicine, Duke–NUS Medical School, Singapore** (Y B Cheung PhD); **Center for Child Health Research, Tampere University, Tampere, Finland** (Y B Cheung); **Epicentre, Paris, France** (R Grais PhD); **Centre for Infectious Disease Epidemiology and Research, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa** (N Ford FRCP); **Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan** (F Jahan MBBS); **Department of Psychiatry, Makerere University, Kampala, Uganda** (E Nakimuli-Mpungu PhD); **Department of Pharmacology and Division of Clinical Research, St John's Medical College, Bangalore, India** (Prof D Xavier MD); **Centre for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada** (Prof Z A Bhutta PhD); **Institute of Global Health and Development, and Centre of Excellence in Women and Child Health, Aga Khan University, Karachi, Pakistan**

Introduction

Use of robust randomised clinical trial (RCT) evaluation is crucial to determine which interventions would be useful for public health or clinical care.[1,2] Certain interventions are delivered at a population level or at a group level, and these interventions can result in changes to group behaviours, leading to large scale population-level effects.[3] The cluster RCT design (herein, cluster trial) is a specific trial design that is used to evaluate interventions delivered at a group level.[4] In cluster trials, whole groups of structured collections of individuals or health system service delivery platforms, such as facilities, are randomly assigned to receive interventions, and these groups are referred to as clusters. Examples of clusters include communities, health clinics, or schools. In contrast to the individual RCT, in which the group allocation of interventions is determined by randomisation of individual participants, cluster trials randomly assign interventions to a whole cluster of individuals. Interventions themselves can be administered at a cluster level (eg, mosquito egg traps[5]) or at an individual level (eg, vaccinations[6]). In cluster trials, outcomes can be measured at a cluster level[6] or at an individual level.[7]

There is an increasing popularity of the use of cluster trials in low-income and middle-income countries (LMICs).[8] There have been many successful high-profile trials that have used a cluster trial design, and many of the design features of cluster trials lend themselves well to priority areas of research set in LMICs. Despite the increasing use of this design, there can be methodological and interpretational challenges. Cluster trials are complex due to the interplay between the similarities of individual participants within a cluster and the differences between clusters.[9,10] Because outcomes of individuals within the same cluster are correlated, standard methods for design and analysis of individual RCTs do not suffice for cluster trials. Furthermore, because special considerations are required when designing and analysing cluster trials, there have been great efforts towards unifying and improving the standards of the design and analysis interpretation of cluster trials.[9,11] However, some evaluations have identified several design and interpretational challenges to cluster trials that often arise due to inadequate planning.[12–14]

In this third paper of the Series, we first discuss the attributes of cluster trials in the context of global health research, followed by specific challenges that are associated with planning and implementing such attributes in LMICs. We then draw on examples of

Key messages

- When interventions can be delivered at the individual level, there should be clear scientific justifications for cluster randomisation. Cluster trials are ideal for evaluating interventions that can only be delivered at the level of clusters or interventions that have shown to be effective under controlled conditions in individually randomised clinical trials that require further evaluation of whether they work at scale. In appropriate situations, cluster trials can be better tailored for implementation science than individually randomised controlled trials.
- During the trial planning stage, it is difficult to estimate clustering effects (eg, intracluster correlation) and other key assumptions required in traditional trial designs (eg, event rates and treatment effects), so it might be useful to plan for multi-stage approaches by use of interim data to estimate key trial parameters and reassess the sample size.
- When there are only a small number of clusters available for randomisation, the use of covariate constrained allocation methods might be useful, given that baseline imbalance between intervention and control groups might be difficult to avoid with simple randomisation.

(Prof Z A Bhutta); **Department of Experimental Medicine, University of British Columbia, Vancouver, BC, Canada** (J J H Park MSc); **School of Public Health, University of Rwanda, Kigali, Rwanda** (Prof E J Mills); **Cytel, Vancouver, BC, Canada** (Prof E J Mills)

Correspondence to: Prof Edward J Mills, Cytel, Vancouver, BC V5Z 1J5, Canada **millsej@mcmaster.ca**

Panel 1: Summary of the MORDOR trial: a cluster trial with planned interim analysis on mass distribution of azithromycin for child mortality in sub-Saharan Africa

MORDOR is a placebo-controlled cluster trial that has been undertaken in Malawi, Tanzania, and Niger, to evaluate the efficacy of biannual mass distribution of oral azithromycin on all-cause mortality for children aged 1–59 months.[7] Before this study, oral azithromycin had shown effectiveness against trachoma and other infectious diseases, such as malaria, diarrhoea, and pneumonia.[15–20] Because mass distribution of azithromycin cannot be tested in individually randomised clinical trials, due to risk of contamination between groups, a cluster randomised trial design was ideal to answer this research question.

The investigators of MORDOR prespecified plans for an interim analysis for stopping the trial early on the basis of superiority and futility. One interim analysis was implemented after mortality data became available for a third of the person-years in the study, with each country having contributed at least 6 months of follow-up for half or more of the clusters in each country. An α value of 0·001 was used as the decision rule for stopping the trial early for superiority, with an α value of 0·049 used to determine statistical significance at the final analysis. The decision rule for stopping the trial early for futility was prespecified as 20% or lower conditional power, to detect a 25% treatment effect. Because the decision rules for stopping the trial early were not met, this cluster trial did not stop after the interim analysis was done.

In the MORDOR trial, 1533 communities (clusters) were randomly assigned to receive either mass distribution of oral azithromycin or placebo on study completion, with 190 238 children enrolled (323 302 person-years monitored). This trial showed large mortality reduction benefits with mass distribution of azithromycin; the azithromycin group had a 13·5% (95% CI 6·7–19·8) lower child mortality risk compared with the placebo.[7]

Panel 2: Summary of the PASTAL trial: an adaptive multiarm, multi-stage cluster trial for HIV testing and linkage to care

PASTAL is a phase 2 adaptive multiarm, multi-stage cluster trial in Malawi that evaluated the effects on the uptake of HIV testing and subsequent HIV services by male partners of pregnant women accessing antenatal clinics.[21,22] This trial started with one control arm and five intervention arms. As the standard of care, the pregnant women within clusters randomised to the control group received a personalised invitation letter to a male-friendly HIV clinic that was addressed to their male partners, and this clinic would offer HIV testing, linkage to care (if shown to be HIV-positive), and pregnancy health education. The clusters randomly assigned to one intervention group would receive the standard of care and two self-test kits; those assigned to two other intervention groups would receive standard of care, two self-test kits, and a conditional financial incentive for the amount of either US$3 (one group) or $10 (in the other group) for male partners who self-tested and attended the HIV clinic. Another intervention group received standard of care, two self-test kits, and conditional lottery entry, with a 10% chance of winning $30; and the final intervention arm received standard of care, two self-test kits, and a telephone call reminder of their clinical appointments.

Stopping the trial early was only considered for futility reasons, and not for superiority. At the first stage, the outcomes reported in each intervention group were compared with the control group, and the intervention group was considered to be dropped from the trial if the pairwise comparison showed a p value that exceed 0·20. An independent data safety monitoring board used this p-value threshold and considerations of cost and safety to remove intervention groups after the interim analysis at the first stage. The data safety monitoring board stopped the interventional approach that included a conditional lottery entry after the interim analysis showed a p value of 0·211. The intervention that consisted of standard of care and self-test kits only was continued to the second stage, despite meeting the threshold for stopping, due to the local policy makers wanting to test this approach in the subsequent stage. In addition to stopping the trial early, sample size reassessment was used to re-calculate the sample size required for the second (final) stage of the trial. The empirical estimates from the interim analysis were used to re-calculate the sample size required under the constraints of 10% family-wise type I error rate and 80% statistical power.

At the end of the trial, there were 71 antenatal clinics (clusters) involved, with 2349 women randomly assigned (six clusters per arm in the first stage and seven clusters per arm in the second stage, after dropping the arm with the lottery). Compared with control clusters in which 17·4% of the male partners were tested for HIV, a higher proportion of male individuals met the primary endpoint in all intervention arms. For instance, among the intervention arm with conditional financial incentives of $3 or $10, which had the largest effect sizes, higher adjusted risk ratios of 3·01 (95% CI 1·63–5·57; for the $3 group) and 3·72 (1·85–7·48; for the $10 group) were found.

cluster trials implemented across key global health research areas, such as maternal, newborn, and child health (MNCH), malaria, and water, sanitation, and hygiene (WASH), to assess the current cluster trial planning practices in LMICs. Lastly, we describe examples of high-quality and innovative LMIC-based cluster trials, (panels 1–3) and we outline alternative approaches and supplementary methodologies to

cluster designs that can be used to improve their efficiency.

Why cluster trials are often used in research set in LMICs

Cluster trials can be useful in LMIC settings for several reasons. Cluster randomisation can potentially reduce treatment contamination between intervention and control groups.[4,10,27] For instance, clinical trials of nutritional interventions that are set in small, rural communities can appropriately and effectively use this design, since masking of nutritional interventions at the individual level might not be possible and treatment contamination could occur by food sharing between community members randomly assigned to different groups. If participants were individually assigned to the control group, they could potentially receive food supplements intended for the intervention group, resulting in a partial effect of the intervention in the control group, with or without a lesser effect in the intervention group. In such a scenario, an underlying true effect of the nutritional intervention is less likely to be detected or might be blunted. As a result, several cluster trials have had notable successes at improving health outcomes that would otherwise be difficult to assess with individual RCTs. These (non-exhaustively) include Ebola vaccination,[28] hypertension management,[29] and overall childhood mortality.[7] In these cases, and in many more, careful planning and consideration of the research question and mechanistic principles of treatment has led to improved understanding across a multitude of diseases, particularly in resource-constrained settings.

In such a scenario, unless the sample size calculation accounts for contamination, the effect of the nutritional intervention is unlikely to be detected with the collected data. Increasing the degree of treatment contamination can reduce the statistical power of a given two-arm individual RCT (figure). Similar findings of the substantial risk of false-negative results have been shown in the literature.[27,30]

By contrast, if the intervention can be delivered at the level of the individual, it might be preferable to conduct an individual RCT with a larger sample size, rather than by cluster randomisation.[31] The delivery of an intervention at an individual level is particularly appropriate in circumstances in which there is only a modest risk of contamination, there is a large intracluster correlation (ICC), or the cluster sizes are large.[27] Sample size requirements to reach 80% statistical power with varying degrees of treatment contamination for a two-arm individual RCT and for a two-arm cluster trial without any contamination are shown in the figure and appendix (pp 2–3). The sample size required for an individual RCT is smaller than that for a cluster trial with a modest degree of treatment contamination. For instance, when 15% relative risk reduction is expected at a baseline event rate of 20%, an individual RCT without contamination requires a sample size of 2069 per trial arm, compared with 2855 in a cluster trial. The individual RCT is more efficient than a cluster trial, as long as treatment contamination is below 18%. The contamination threshold can increase with larger cluster sizes or ICCs, or both, when treatment effects are larger.

***Panel 3*: The THRio study (NCT00107887): a stepped-wedge cluster trial for tuberculosis screening and preventive therapy**

The tuberculosis/HIV in Rio de Janeiro (THRio) study was a stepped-wedge cluster trial that aimed to evaluate the role of staff training in tuberculosis screening, tuberculosis skin tests, and use of isoniazid preventive therapy versus standard of care in HIV clinics in Rio de Janeiro, Brazil.[23–26] Of the 29 HIV clinics (clusters) enrolled in this stepped-wedge trial, two (7%) clinics were randomly assigned to the intervention group every 2 months between September, 2005, and August, 2009. A coefficient of variation of 0·2 for tuberculosis incidence rates was used to calculate the sample size, assuming a control event rate of 3·65 events per 100 person-years and a 40% reduction in treatment effect size (ie, of 2·20 events per 100 person-years in the intervention group) to reach a 5% type I error rate and 80% statistical power. The primary analysis used an intention-to-treat analysis, including all patients who were eligible to receive the tuberculosis skin test or isoniazid preventive therapy (n=12 816 patients).

Compared with the control, the intervention showed an adjusted hazard ratio of 0·73 (95% CI 0·54–0·99) for tuberculosis incidence and an adjusted hazard ratio of 0·69 (0·57–0·83) for the composite outcome of incidence of and death from tuberculosis (with adjustments made for age, sex, use of highly active antiretroviral therapy at baseline, and time-varying CD4 concentration). Additionally, the rate of tuberculosis skin testing showed an improvement from 19 events per 100 person-years observed during the control period to 59 events per 100 person-years in the intervention period, and the rate of isoniazid preventive therapy improved from 36 events per 100 person-years to 144 events per 100 person-years between the control and intervention periods.

External validity

Another reason for the popularity of cluster trials over individual RCTs is the perceived improved external validity (or generalisability).[4,32] This perception of improved external validity is particularly prevalent when interventions engage large cluster groups, such as health districts. Global health research often aims to be pragmatic, to answer whether interventions can work under real-world conditions, and to be applicable to a diverse population.[33,34] Many health challenges in LMIC contexts are well addressed by these aims. Cluster trials provide a unique approach in infectious disease interventional research, facilitating assessment of population-level efficacy in the presence of herd effects. Scaling of interventions is also reported as a benefit of cluster trials. For instance, trials of interventions administered at the community level more closely resemble how a scaled intervention would be distributed logistically. Further, cluster trials frequently cover the entire community to provide a better estimate of group benefit. A successful example of a community-based scale-up interventional trial is the Control of Blood

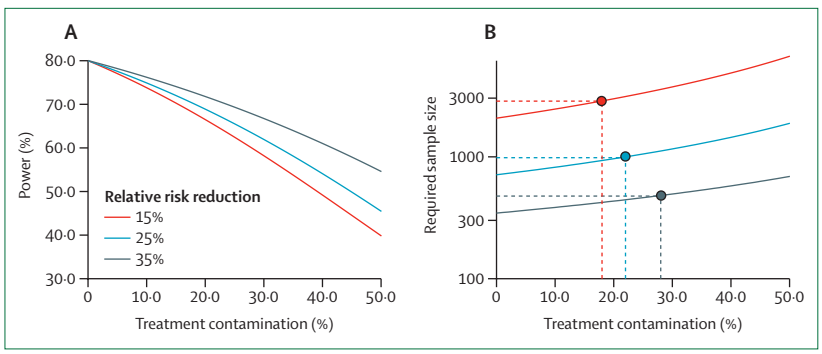

Figure: **Effects of treatment contamination on statistical power of a two-arm individually randomised clinical trial assessing a dichotomous outcome and on the sample size required for 80% power**
(A) Effect of treatment contamination on power of individual randomised clinical trials. (B) Sample size per trial arm required for 80% power in an individual randomised clinical trial, with contamination (solid lines) and the degree of contamination in an individual randomised clinical trial that requires a sample size identical to that of a cluster trial without contamination (dashed lines) shown. Detailed assumptions are provided in the appendix (pp 2–3).

See Online for appendix

Pressure and Risk Attenuation study, set in Bangladesh, Pakistan, and Sri Lanka, which was a multi-part intervention delivered by community health workers to manage hypertension. A strength of this study was that the programme was focused on outreach by community health workers rather than on guiding communities to treatment centres and other settings outside of their communities.[29]

Many individual RCTs are carried out with specific types of patients under controlled conditions, with strict inclusion and exclusion criteria (mechanistic individual RCTs) for improved internal validity, but often this internal validity is achieved at the expense of external validity (explanatory trials). This type of trial design has been criticised for not reflecting real-world conditions and having unrealistic clinical populations, leading to poor external validity.[35–37] As such, evidence from cluster trials is appealing to policy makers because their implementation can more accurately reflect the real-world roll-out of novel interventional strategies. However, related approaches, such as pragmatic individual RCTs, can be carried out with diverse groups of participants to evaluate the effectiveness of interventions in real-life routine settings without the use of cluster trial designs.[36]

Acceptability, adherence, and ethical considerations

Cluster trials are argued to have higher acceptability by the communities receiving the interventions and improved adherence to the allocated intervention compared with individual RCTs.[4,10] When treatment group masking is not possible, individual randomisation can create resentment among individuals randomly assigned to the control group, particularly if the given trial aims to evaluate the role of a desired intervention, such as cash transfers.[4] Cluster trials could be more appropriate to evaluate these types of interventions because it might be unfair and unethical for certain individuals to receive financial aid while their peers knowingly do not.[4,38]

Cluster trials are conceptually able to improve acceptability and adherence to an assigned intervention, but these qualities are rarely assessed and reported in cluster trial publications.[32,39] Improved reporting on such topics could strengthen these discussions. Some shortcomings have been discussed with regard to assessments of cluster trial acceptability from previous work. In global health research, acceptability at the community level is often assessed by village leaders, who might not accurately speak on behalf of the individuals.[38] This approach could affect the adherence of individual members of the cluster, although it is important to note that previous reviews of adherence in cluster trials have indicated poor reporting of this characteristic, complicating the interpretation of results.[39]

Practical advantages

Other reasons for the use of cluster randomisation include practical advantages, such as administrative convenience, lower implementation costs, and improved acceptability and compliance compared with individual RCTs.[4,39,40]

Trial planning challenges for cluster trials in global health

A specific consideration in cluster trials is related to the interplay between individuals and clusters. In cluster trials, the responses of individuals within each cluster are often correlated with respect to environmental, socioeconomic, and other specified or non-specified prognostic factors.[40] This correlation leads to an increase of within-cluster correlation and between-cluster variability in the health outcomes concerned.[4] In turn, the overall variability of the observed treatment effects is inflated, resulting in statistical inefficiency that might require a much larger sample size compared with individual RCTs.[41] Cluster trials that fail to account for the effects of clustering in the analysis will result in inflated type I error rates.[10,14] As such, it is vital to emphasise the importance of including these characteristics in any statistical analysis plan and subsequent publication. There is often a misunderstanding among investigators and funders that cluster trials facilitate enrolment of a larger number of participants (more than 1000) than individual RCTs and that these large sample sizes therefore might be more representative or provide better precision.[42,43]

The effects of clustering on the sample size can be expressed through the design effect, a statistical measure of relative inflated variance due to cluster randomisation.[4] For instance, a cluster trial with a design effect of 3 would require triple the sample size of an individual RCT. Calculating the design effect requires an estimated value of the ICC or coefficient of variation. ICC measures the similarity of the individual responses from within the cluster, compared with the responses from different clusters.[4,44,45] As the ICC value decreases, so does the required sample size for statistical power.[45,46] The

coefficient of variation (a standardised measure of the between-cluster variance) can also be used to express the effects of clustering to calculate the sample size of cluster trials, and is directly related to the sample size.[4]

For both cluster trials and individual RCTs, investigators must prespecify characteristics such as effect sizes and loss-to-follow-up rates. Prespecification of trial parameters is an added challenge in cluster trials because entire clusters might be lost to follow-up, and it might be difficult to make such assumptions at the start of the study. Availability of a sampling framework (the existing number of clusters to choose from) for randomisation and the size of the cluster all add to the complexity of sample size calculation when compared with individually randomised trials.[47] As such, pretrial activities assessing the true number of available clusters, running scenario analyses, interim evaluations and, where appropriate, clinical trial simulation can assist in developing contingency plans for these challenges.

Number of clusters randomised within the trial

Knowledge of the number of clusters available for randomisation is important for trial planning for a multitude of reasons. Randomisation might help to ensure that there is balance in observable and non-observable prognostic factors between groups being compared, but balance in prognostic factors is only retained when there are a large number of clusters that are being randomised. There are often geographical and other practical constraints that limit the number of clusters that can be enrolled in a given clinical trial, particularly in LMICs. For instance, there might be only a small number of hospitals in a region of interest, or there might only be enough funding to enrol a small number of clusters. The number of clusters should be adequate, to minimise the chance of imbalance; although, this number must also be considered in relation to the potential for overpowering the trial.[48,49] Notably, although increased cluster numbers might generally improve balance, specific plausible variables of interest (eg, socioeconomic status and development parameters) must also be adequately accounted for with increased cluster numbers. As such, careful evaluation of cluster sizes and local restrictions (eg, the total number of schools or communities available) is essential for effective trial planning.

Confounding resulting from imbalance can result in inflated risks of false-positive and false-negative findings. Although statistical techniques can adjust for baseline covariates in cases of imbalance, these techniques still require the availability of a relatively large number of individual patient-level covariates, which might not be available. The use of a large number of clusters is still preferred, such that randomisation can ensure balance in terms of both observable and non-observable factors between groups being compared.[50] Finally, it is important to recognise that power in a cluster trial depends to a greater extent on the number of clusters than the cluster sizes: a hypothetical cluster trial with hundreds of thousands of participants but only a few clusters might never reach 80% statistical power, owing to the diminishing returns that increasing participant numbers relative to increasing cluster numbers provides for cluster trials.[48]

Cluster sizes and their variability

Cluster sizes and their variability are important limiting factors for cluster trials. The cluster size (ie, the number of participants per cluster) required to reach sufficient statistical power depends on the number of clusters available for randomisation. When the number of clusters increases, fewer participants per cluster are ideal, and when only a small number of clusters are available, a larger number of participants per cluster is required.[4,46] Although statistical power might be improved by increasing the total sample size, increasing the sample size can have diminishing returns on statistical power, especially when the ICC is large.[49] Increasing the sample size becomes a particular challenge when communicating the results of cluster trials. There is often a temptation to relay the significance of a trial in relation to the total sample size of the trial. In cluster trials, substantial numbers of patients are often recruited, but each individual's respective contribution to the statistical power of the trial is unclear without reference to the total cluster size and associated variability.

In addition to cluster sizes, it is important to consider the variability in sizes for trial planning because cluster size variability can negatively affect the statistical power.[51,52] Compared with cluster trials with equal cluster sizes, cluster trials with highly unequal cluster sizes require larger sample sizes to reach the same statistical power.[46] In fact, the required sample size increases as the cluster size variability increases, because smaller clusters will have less precise treatment estimates than those from larger clusters.[46] Cluster size variability might cause imbalances between intervention and control groups that could be difficult to adjust for with the use of statistical analyses.[51] In instances for which allocation concealment is not possible, cluster size variability and baseline imbalance might result from selection or attrition biases, or both.[53]

Cluster size variability is common in community trials where, based on eligibility, there might be an under-recruitment of individuals in some of the clusters or an over-recruitment in the other clusters, or both.[51] If recruitment is delayed due to under-recruitment in some regions, there might be a perception that recruiting more participants from other clusters will help to balance the trial. However, this alteration might exaggerate the cluster size variability or lower statistical power further and introduce biases, affecting both the internal validity and generalisability of the trial findings.[51] The overall effect of this variability on subsequent statistical power and operational characteristics can be variable,[54] and it is

frequently challenging to formally assess the effects of variable cluster sizes, owing to poor reporting of key trial characteristics within the literature.[9]

Anticipating the clustering effects

Anticipating the effects of clustering can be challenging. There is often no accurate estimate of ICC, and as a result, the ICC is often ignored or a nominal value (typically 0·05) is used for sample size calculations of cluster trials.[11] Even with accurate assumptions about the underlying treatment effects and baseline risks, which is also required for individually randomised trials, a cluster trial might be underpowered with an underestimated ICC assumption or, on rare occasions in which the ICC is overestimated, the trial can end up recruiting excessive numbers of clusters and sample sizes. To facilitate future research activities and improve the interpretation of cluster trials, reporting of observed ICC values is crucial. Reporting of ICC is currently recommended in the Consolidated Standards of Reporting Trials extension for cluster trials (item 17a), but this recommendation is limited to the primary outcome only, and ICC is specific to each outcome, creating challenges with interpreting secondary outcomes.[9]

Because a large number of clusters are usually required to achieve balance between intervention and control arms, cluster trials in global health research often involve multiple geographical regions across several countries. When carrying out a cluster trial across multiple regions, there are frequently substantial disparities in disease burden between geographical settings. These disparities can contribute to high variability between different clusters and thus inflate the ICC. For instance, recent work has shown substantial subnational disparities in the prevalence of childhood stunting in several African countries, with national stunting prevalence varying by more than five times between nearby countries.[55]

There can also be substantial disparities in disease burden within a single country, making the implementation of single-nation cluster trials difficult in global health research. For example, in Rwanda, there are disparities in the prevalence of childhood stunting between different districts across Rwanda (appendix p 4).[56] Although the stunting prevalence in Kigali (the capital city) is low (22%), the prevalence of stunting is much higher in four nearby provinces (ie, the western, northern, southern, and eastern provinces). Geographical disparities in disease burden can pose challenges when trying to estimate the clustering effects as part of trial planning of cluster trials in global health research. Another example of the influence of geography on trial outcomes is cord cleansing with chlorhexidine, which in settings of a high incidence of home-birth deliveries (around 75%) showed marked improvements in neonatal mortality.[57] When repeated in settings of a high incidence of facility-birth deliveries (around 65%), minimal improvements were noted, emphasising the importance of cluster selection.[58]

Maintaining internal validity

As briefly mentioned, cluster trials are often preferred in global health research for their pragmatism and generalisability, but at a compromise of internal validity.[4,42] Balancing for external validity and internal validity needs careful consideration. Although selection bias is possible in individual RCTs, such biases can be more problematic for cluster trials because attrition bias can occur at both the individual and cluster levels.[42] For instance, for a hospital-based cluster trial, doctors acting as recruiters for the hospital might be less motivated to recruit patients if their hospital was randomly assigned to the control group than doctors in hospitals allocated to the intervention group.

There are several strategies to minimise selection bias for cluster trials, covering both individual-level and cluster-level selection biases.[42,43,59] Bias from study recruiters can be minimised by use of an independent recruiter who is masked to group allocation. Selection bias at the individual level can be minimised by identifying and obtaining consent from participants before cluster randomisation,[60] or by offering delayed assignment to the other group (eg, at the end of the study).

Ethical challenges

In addition to the aforementioned statistical issues, cluster trials can raise difficult ethical questions for global health research. A useful resource for researchers interested in these challenges is the Ottawa Statement on Ethical Design and Conduct of Cluster Randomised Trials.[61] In essence, cluster trials share many ethical challenges with individual RCTs, although there are several unique design features which warrant a more detailed review.

In individual RCTs, obtaining informed consent from each individual study participant is usually required; however, in cluster trials, informed consent might only be obtained from selected members of the community if the interventions can only be administered at the community level.[38] The role of such community gatekeepers in cluster trials is a contentious topic in medical ethics; for further discussion, readers are encouraged to review Gallo and colleagues' work reviewing this specific issue.[62]

For interventions that can be evaluated by use of individual RCTs, it is important that these interventions have already shown their clinical utility and pose minimal risks to the population before being considered for cluster trials, given that these designs often result in large numbers of participants. Additionally, it is important to consider the potential ethical issues of withholding a given intervention from certain communities and not others, because this can contribute to the perception of health inequity among trial participants and between communities.[38] As data accumulate in a cluster trial, or if policy changes occur during the implementation of the trial, there might be a strong obligation to modify or even stop the study if the study intervention is unsafe or ineffective, or both.[38]

	All (n=80)	MNCH (n=23)	Malaria (n=32)	WASH (n=25)
Proportion of trials with binary or count outcome as the primary outcome	62 (78%)	10 (43%)	32 (100%)	20 (80%)
Binary or count baseline effect was >50% smaller than predicted	11/45 (24%)	1/9 (11%)	7/24 (29%)	3/12 (25%)
Binary or count treatment effect was meaningfully different to predicted value*	27/53 (51%)	7/10 (70%)	14/29 (48%)	6/14 (43%)
Intracluster correlation or coefficient of variation was meaningfully different to predicted value†	9/13 (69%)	4/5 (80%)	5/6 (83%)	0/2 (0%)

Data are n (%) or n/N (%). MNCH=maternal, newborn, and child health. WASH=water, sanitation, and hygiene. *Observed effect of <25% relative risk reduction or <2% absolute differences for binary outcomes. †A difference of >50% relative to the originally planned value, regardless of direction of effect. The substantially smaller denominator value is because of the small (16%) number of trials that reported both a planned and an observed intracluster correlation or coefficient of variation.

Table: **Assessment of sample size calculation practices for cluster trials in global health: baseline event rate, treatment effect size, and intracluster correlation assumptions across MNCH, malaria, and WASH trials**

Assessment of cluster trial planning practices in global health

To provide some examples of cluster trial planning practices in global health research set in LMICs, we did a focused literature review of LMIC-based cluster trials that have been published between Jan 1, 2010, and Nov 1, 2018 for three key areas of global health research: MNCH, WASH, and malaria. We evaluated the cluster trial planning practices in global health research set in LMICs by assessing the reported assumptions used to derive sample size estimations among the 80 eligible cluster trials in these selected disease topics (table). The search terms, eligibility criteria, and details of our analytical approaches for this assessment are provided in the appendix (pp 5–27).

Although not unique to cluster trials or global health research, cluster trial sample size calculations are often done with higher baseline event rates, overly optimistic treatment effect sizes, and lower clustering effects than those that are actually observed later. Among the cluster trials on the topics of MNCH, WASH, and malaria, we identified that a large proportion of these trials used higher rates of baseline event rates and anticipated treatment effects for sample size calculations than those that were later actually observed. Of the 80 included trials, 62 (78%) of the trials were undertaken with binary or count data primary outcomes. Of the 80 trials assessed, 45 (56%) reported data on assumed baseline risks. Where presented, 11 (24%) of 45 trials reported a baseline risk or rate at least 50% smaller than anticipated. The proportion of overestimates varied across the included topics, with MNCH representing the lowest proportion (11%), and malaria (29%) and WASH (25%) trials representing similar proportions. Three (7%) of the 45 trials had overestimated the baseline risk or rate during planning by at least 50%.

In examining treatment effect sizes, we observed that 27 (51%) of 53 of the included trials reported overestimation of treatment effects, in which the observed effect size was at least 50% smaller than the assumed effect size for sample size calculations. Higher proportions of overestimation (ie, an effect size less than 50% of the assumed effect size) were noted in MNCH trials (70%) relative to malaria (48%) and WASH (43%) trials. It was more difficult to assess the cluster trial planning practices in terms of clustering effects because the ICC was often inadequately reported. Of 80 cluster trials identified, only 13 trials (16%) reported on both planned and observed ICC. Of the 13 trials that reported both planned and observed ICCs, 9 (69%) noted a difference of 50% or greater than assumed values. Overall, 8 (18%) of 45 trials showed an overestimation of baseline effects at the same time as an overestimation of treatment effects. Similarly, 12 (27%) of 45 trials reported overestimations of treatment effects and either ICC or baseline effects.

There are many possible reasons for the high frequency of erroneous assumptions used for cluster trial planning in terms of baseline event rates, treatment effects, or clustering effects. Erroneous assumptions used, particularly for treatment effect sizes, might be due to optimism held by the trial investigators.

Although we report on the widespread mismatch in anticipated baseline event rates and treatment effects relative to those that were later actually observed in the global health cluster trials, this disparity was not frequently discussed in the publications. One publication noted that their substantially lower observed base event rate for malaria prevalence was almost certainly due to a drought that coincided with the initiation of their trial.[63] Another malaria trial noted that stopping of one outcome measurement of parasite detection was due to the low event rate of parasite carriage in enrolled clusters,[64] whereas two other trials noted that their planned sample size was increased to retain appropriate statistical power.[65,66]

Alternative approaches and designs for cluster trials in global health

Interim evaluations

Cluster trials usually have a single-stage approach with a fixed trial design, in which all clusters are enrolled and randomised simultaneously and then observed for a fixed duration. Instead of using a single-stage approach, it might be beneficial to use a multi-stage approach, in which interim data are used to estimate the key trial parameters and calculate the sample size again (sample

size reassessment).[67] To determine the timing of the interim analysis, a sufficient number of observations from the enrolled clusters are required for a reliable estimate of the ICC at the interim analysis.[68]

For instance, if a sample size reassessment shows that the new requirements in sample size or in the number of clusters exceed what is realistically feasible, a decision to stop the trial might be considered. The use of a multistage approach with preplanned interim evaluation plans can represent an appealing mechanism to minimise the risks of false-negative findings for cluster trials in which uncertainty exists over the planned trial parameters.[69] High-profile cluster trials have already integrated such methodologies, such as early stopping rules for efficacy and futility,[7] as well as preplanned rules for stopping of treatment arms in isolation for multiarm trials.[70]

Statistical methods for imbalance

It is often assumed that randomisation can remove selection bias and produce groups that are comparable in terms of both measurable and unmeasurable factors. However, in cluster trials, this assumption might not hold true, especially when there is a low number of clusters available. The potential risk of baseline imbalance between the intervention and control groups can be substantial when there is a small number of clusters available.[40] If data are available for important cluster-level or individual-level covariates, covariate-constrained randomisation could be a useful method to prevent imbalance in studies for which there is only a small number of clusters.[71]

In covariate-constrained randomisation, investigators determine criteria for balance across key baseline covariates of interest. All possible configurations of allocations of participating clusters can then be generated to subsequently be narrowed down to a smaller list of potential allocations by use of the prespecified criteria for balance. Then, the actual allocation would be chosen randomly from this narrower list to maintain an appropriate level of randomness in this selection process. When using covariate-constrained randomisation, it is important to prespecify the key covariates and the criteria for balance before generating all possible allocation schemes. It is also important not to apply overconstraint through strict balancing requirements because this scenario might result in very few possible allocations and, thus, not represent true randomisation.[71] Covariates used in the constrained randomisation should be adjusted for in the statistical analyses, to avoid inflation of type I error rates.[72]

Stepped-wedge designs

A stepped-wedge design is a type of cluster trial design in which all clusters start in the control condition but gradually cross over to the intervention condition, in randomised sequence, until every cluster included in the trial has received the intervention of interest.[73] This design differs from the conventional parallel design that is most often used in cluster trials because, in parallel design, clusters are randomly assigned to intervention or control groups; not all clusters receive the intervention of interest before the cluster trial with a parallel design is completed.[73]

Stepped-wedge designs can help to minimise some of the drawbacks of parallel arm designs under select circumstances. For instance, stepped-wedge designs might be useful when there is only a small number of clusters available. Stepped-wedge designs can often reach the same statistical power as the parallel cluster trial design but with fewer clusters, particularly when ICC values are greater than 0·03.[73,74] The reason stepped-wedge designs might require fewer clusters is because each cluster is exposed to both control and intervention conditions; hence, the estimation of treatment effects in a stepped-wedge trial can benefit from the use of within-cluster comparisons. For example, using the results of a previous cluster trial,[75] Hooper and Bourke[74] have shown that a stepped-wedge design consisting of two, three, or four regular intervals (the so-called steps), in which a group of clusters are randomised to cross from the control to the intervention, reduced the required number of clusters to reach appropriate statistical power by 23%, 54%, and 66%, respectively. Another advantage of stepped-wedge designs is less sensitivity of statistical power to the ICC value. Although the statistical power is reduced in both parallel and stepped-wedge trial designs when an underestimated ICC value is used for sample size calculation, the statistical power for the stepped-wedge design is less affected than that of the parallel design.[73] However, stepped-wedge cluster trials increase statistical complexities compared with parallel cluster trials in that correlation coefficients are more complex: not only do correlations within the same cluster need to be accounted for, correlations in repeated measures from the same cluster over time also need to be factored in.

Although stepped-wedge trials offer promise when high numbers of individuals per cluster are available, these trials do have drawbacks. Even though increased steps can provide improved statistical properties, this approach also requires more measurements (and, as such, a potentially longer trial duration) per cluster compared with parallel designs.[76] As such, there is an important need during trial planning to balance the number of steps and associated sample size reduction against increased operational costs and trial duration with increased numbers of observations, and whether there are any associated ethical implications with this. The associated operational complexity of coordinated calendar timing for stepped-wedge designs must also not be underestimated. Potentially as a consequence of this underestimation, authors have previously noted that approximately half of stepped-wedge trials failed to reach their planned sample size and implementation schedules.[77]

Conclusion

Evaluating whether interventions have an effect at the population level is challenging. The most commonly applied method for evaluating population-level effects—ie, the cluster trial—has complications that can be difficult to overcome. When carried out successfully, these trials represent a unique and powerful investigational tool, but careful considerations should be made ahead of their implementation. Cluster trials are typically much larger than individual RCTs, so these trials require considerably larger financial resources, posing challenges for LMICs and research funding bodies that are already stretched. More rigorous pretrial estimation processes, alongside integration of novel trial methodologies, might reduce the prevalence of these problems and minimise the number of participants recruited, the money spent, and the time lost on trials that do not have the necessary characteristics to draw meaningful conclusions.

Contributors

LD, JJHP, and EJM conceptualised the paper. LD, RG, NF, KT, JJHP, ZAB, and EJM acquired and analysed the data. All authors interpreted the data. LD, JJHP and EJM drafted the paper. All authors critically revised the paper for important intellectual content. EJM obtained the funding, and JJHP and EJM provided administrative, technical, and material support and supervised the study.

Declaration of interests

DX reports grants from AstraZeneca, grants from Boehringer Ingelheim, grants from Bristol Myers Squibb, grants from Coca-Cola India, speaker fees from Eli Lilly, speaker fees from the Indian Council of Medical Research, grants from Pfizer, speaker fees from Sanofi, grants from the UK Medical Research Council, and grants from the Wellcome Trust, outside of the submitted work. All other authors declare no competing interests.

Acknowledgments

This Series was funded by the Bill & Melinda Gates Foundation. No authors received funding, and all authors volunteered their time. The funding source had no role in the design and implementation of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the paper; and decision to submit the paper for publication. The funding provided by the Gates Foundation was used for administrative support to coordinate submission efforts. The contents of the paper are the sole responsibility of the authors and might not necessarily represent the official views of the Gates Foundation or other agencies that might have supported the primary data studies used in the present study. The corresponding author (EJM) had the final responsibility for the decision to submit for publication.

References

1 Cockcroft A. Randomised controlled trials and changing public health practice. *BMC Public Health* 2017; **17** (suppl 1)**:** 409.

2 Ellenberg SS, Keusch GT, Babiker AG, et al. Rigorous clinical trial design in public health emergencies is essential. *Clin Infect Dis* 2018; **66:** 1467–69.

3 Hawe P, Potvin L. What is population health intervention research? *Can J Public Health* 2009; **100:** suppl I8–14.

4 Hayes RJ, Moulton LH. Cluster randomised trials. New York, NY: Chapman and Hall/CRC, 2009.

5 Garcia KKS, Versiani HS, Araújo TO, et al. Measuring mosquito control: adult-mosquito catches vs egg-trap data as endpoints of a cluster-randomized controlled trial of mosquito-disseminated pyriproxyfen. *Parasit Vectors* 2020; **13:** 352.

6 Nowalk MP, Lin CJ, Hannibal K, et al. Increasing childhood influenza vaccination: a cluster randomized trial. *Am J Prev Med* 2014; **47:** 435–43.

7 Keenan JD, Bailey RL, West SK, et al. Azithromycin to reduce childhood mortality in sub-Saharan Africa. *N Engl J Med* 2018; **378:** 1583–92.

8 Lorenz E, Köpke S, Pfaff H, Blettner M. Cluster-randomized studies. *Dtsch Arztebl Int* 2018; **115:** 163–68.

9 Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; **345:** e5661.

10 Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health* 2004; **94:** 416–22.

11 Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004; **328:** 702–08.

12 Rutterford C, Taljaard M, Dixon S, Copas A, Eldridge S. Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review. *J Clin Epidemiol* 2015; **68:** 716–23.

13 Siebenhofer A, Paulitsch MA, Pregartner G, et al. Cluster-randomized controlled trials evaluating complex interventions in general practices are mostly ineffective: a systematic review. *J Clin Epidemiol* 2018; **94:** 85–96.

14 Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. Appropriate statistical methods were infrequently used in cluster-randomized crossover trials. *J Clin Epidemiol* 2016; **74:** 40–50.

15 Chidambaram JD, Alemayehu W, Melese M, et al. Effect of a single mass antibiotic distribution on the prevalence of infectious trachoma. *JAMA* 2006; **295:** 1142–46.

16 Whitty CJ, Glasgow KW, Sadiq ST, Mabey DC, Bailey R. Impact of community-based mass treatment for trachoma with oral azithromycin on general morbidity in Gambian children. *Pediatr Infect Dis J* 1999; **18:** 955–58.

17 Fry AM, Jha HC, Lietman TM, et al. Adverse and beneficial secondary effects of mass treatment with azithromycin to eliminate blindness due to trachoma in Nepal. *Clin Infect Dis* 2002; **35:** 395–402.

18 Coles CL, Levens J, Seidman JC, Mkocha H, Munoz B, West S. Mass distribution of azithromycin for trachoma control is associated with short-term reduction in risk of acute lower respiratory infection in young children. *Pediatr Infect Dis J* 2012; **31:** 341–46.

19 Coles CL, Seidman JC, Levens J, Mkocha H, Munoz B, West S. Association of mass treatment with azithromycin in trachoma-endemic communities with short-term reduced risk of diarrhea in young children. *Am J Trop Med Hyg* 2011; **85:** 691–96.

20 Gaynor BD, Amza A, Kadri B, et al. Impact of mass azithromycin distribution on malaria parasitemia during the low-transmission season in Niger: a cluster-randomized trial. *Am J Trop Med Hyg* 2014; **90:** 846–51.

21 Choko AT, Corbett EL, Stallard N, et al. HIV self-testing alone or with additional interventions, including financial incentives, and linkage to care or prevention among male partners of antenatal care clinic attendees in Malawi: an adaptive multi-arm, multi-stage cluster randomised trial. *PLoS Med* 2019; **16:** e1002719.

22 Choko AT, Fielding K, Stallard N, et al. Investigating interventions to increase uptake of HIV testing and linkage into care or prevention for male partners of pregnant women in antenatal clinics in Blantyre, Malawi: study protocol for a cluster randomised trial. *Trials* 2017; **18:** 349.

23 Azadi M, Bishai DM, Dowdy DW, et al. Cost-effectiveness of tuberculosis screening and isoniazid treatment in the TB/HIV in Rio (THRio) study. *Int J Tuberc Lung Dis* 2014; **18:** 1443–48.

24 Durovni B, Cavalcante SC, Saraceni V, et al. The implementation of isoniazid preventive therapy in HIV clinics: the experience from the TB/HIV in Rio (THRio) study. *AIDS* 2010; **24** (suppl 5)**:** S49–56.

25 Golub JE, Cohn S, Saraceni V, et al. Long-term protection from isoniazid preventive therapy for tuberculosis in HIV-infected patients in a medium-burden tuberculosis setting: the TB/HIV in Rio (THRio) study. *Clin Infect Dis* 2015; **60:** 639–45.

26 Moulton LH, Golub JE, Durovni B, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention. *Clin Trials* 2007; **4:** 190–99.

27 Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001; **322:** 355–57.

28 Henao-Restrepo AM, Camacho A, Longini IM, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). *Lancet* 2017; **389:** 505–18.

29 Jafar TH, Gandhi M, de Silva HA, et al. A community-based intervention for managing hypertension in rural South Asia. *N Engl J Med* 2020; **382:** 717–26.

30 Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. *BMJ* 2010; **340:** c2073.

31 Moerbeek M. Randomization of clusters versus randomization of persons within clusters: which is preferable? *Am Stat* 2005; **59:** 173–79.

32 Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ* 2008; **336:** 876–80.

33 Rosala-Hallas A, Bhangu A, Blazeby J, et al. Global health trials methodological research agenda: results from a priority setting exercise. *Trials* 2018; **19:** 48.

34 Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014; **5:** e45.

35 Williams BA. Perils of evidence-based medicine. *Perspect Biol Med* 2010; **53:** 106–20.

36 Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016; **375:** 454–63.

37 Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015; **16:** 495.

38 Weijer C, Grimshaw JM, Taljaard M, et al. Ethical issues posed by cluster randomized trials in health research. *Trials* 2011; **12:** 100.

39 Agbla SC, DiazOrdaz K. Reporting non-adherence in cluster randomised trials: a systematic review. *Clin Trials* 2018; **15:** 294–304.

40 Donner A, Klar N. Cluster randomization trials in epidemiology: theory and application. *J Stat Plan Inference* 1994; **42:** 37–56.

41 Peek N, Goud R, de Keizer N. Handling intra-cluster correlation when analyzing the effects of decision support on health care process measures. *Stud Health Technol Inform* 2013; **186:** 22–27.

42 Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? *BMJ* 2009; **339:** b4006.

43 Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med* 2009; **6:** e1000065.

44 Westgate PM. Intra-cluster correlation selection for cluster randomized trials. *Stat Med* 2016; **35:** 3272–84.

45 Donner A, Koval JJ. Design considerations in the estimation of intraclass correlation. *Ann Hum Genet* 1982; **46:** 271–77.

46 Ribeiro DC, Milosavljevic S, Abbott JH. Sample size estimation for cluster randomized controlled trials. *Musculoskelet Sci Pract* 2018; **34:** 108–11.

47 van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *J Clin Epidemiol* 2012; **65:** 1212–18.

48 Taljaard M, Teerenstra S, Ivers NM, Fergusson DA. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. *Clin Trials* 2016; **13:** 459–63.

49 Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. *BMJ* 2017; **358:** j3064.

50 Egbewale BE. Statistical issues in randomised controlled trials: a narrative synthesis. *Asian Pac J Trop Biomed* 2015; **5:** 354–59.

51 Carter B. Cluster size variability and imbalance in cluster randomized controlled trials. *Stat Med* 2010; **29:** 2984–93.

52 Lauer SA, Kleinman KP, Reich NG. The effect of cluster size variability on statistical power in cluster-randomized trials. *PLoS One* 2015; **10:** e0119074.

53 Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. *BMC Med Res Methodol* 2005; **5:** 10.

54 van Breukelen GJP, Candel MJJM. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *J Clin Epidemiol* 2012; **65:** 1212–18.

55 Osgood-Zimmerman A, Millear AI, Stubbs RW, et al. Mapping child growth failure in Africa between 2000 and 2015. *Nature* 2018; **555:** 41–47.

56 National Institute of Statistics of Rwanda. Rwanda Demographic and Health Survey 2014–15. 2016. https://dhsprogram.com/pubs/pdf/FR316/FR316.pdf (accessed Feb 8, 2021).

57 Mullany LC, Darmstadt GL, Khatry SK, et al. Topical applications of chlorhexidine to the umbilical cord for prevention of omphalitis and neonatal mortality in southern Nepal: a community-based, cluster-randomised trial. *Lancet* 2006; **367:** 910–18.

58 Semrau KEA, Herlihy J, Grogan C, et al. Effectiveness of 4% chlorhexidine umbilical cord care on neonatal mortality in Southern Province, Zambia (ZamCAT): a cluster-randomised controlled trial. *Lancet Glob Health* 2016; **4:** e827–36.

59 Yang R, Carter BL, Gums TH, Gryzlak BM, Xu Y, Levy BT. Selection bias and subject refusal in a cluster-randomized controlled trial. *BMC Med Res Methodol* 2017; **17:** 94.

60 Peri K, Kerse N, Robinson E, Parsons M, Parsons J, Latham N. Does functionally based activity make a difference to health status and mobility? A randomised controlled trial in residential care facilities (The Promoting Independent Living Study; PILS). *Age Ageing* 2008; **37:** 57–63.

61 Weijer C, Grimshaw JM, Eccles MP, et al. The Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials. *PLoS Med* 2012; **9:** e1001346.

62 Gallo A, Weijer C, White A, et al. What is the role and authority of gatekeepers in cluster randomized trials in health research? *Trials* 2012; **13:** 116.

63 Sangoro O, Turner E, Simfukwe E, Miller JE, Moore SJ. A cluster-randomized controlled trial to assess the effectiveness of using 15% DEET topical repellent with long-lasting insecticidal nets (LLINs) compared to a placebo lotion on malaria transmission. *Malar J* 2014; **13:** 324.

64 Shekalaghe SA, Drakeley C, van den Bosch S, et al. A cluster-randomized trial of mass drug administration with a gametocytocidal drug combination to interrupt malaria transmission in a low endemic area in Tanzania. *Malar J* 2011; **10:** 247.

65 Chandler CI, Webb EL, Maiteki-Sebuguzi C, et al. The impact of an intervention to introduce malaria rapid diagnostic tests on fever case management in a high transmission setting in Uganda: a mixed-methods cluster-randomized trial (PRIME). *PLoS One* 2017; **12:** e0170998.

66 Chen-Hussey V, Carneiro I, Keomanila H, et al. Can topical insect repellents reduce malaria? A cluster-randomised controlled trial of the insect repellent N,N-diethyl-m-toluamide (DEET) in Lao PDR. *PLoS One* 2013; **8:** e70664.

67 Lake S, Kammann E, Klar N, Betensky R. Sample size re-estimation in cluster randomization trials. *Stat Med* 2002; **21:** 1337–50.

68 Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *Int J Epidemiol* 2015; **44:** 1051–67.

69 Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ* 2018; **360:** k698.

70 Foy BD, Alout H, Seaman JA, et al. Efficacy and risk of harms of repeat ivermectin mass drug administrations for control of malaria (RIMDAMAL): a cluster-randomised trial. *Lancet* 2019; **393:** 1517–26.

71 Moulton LH. Covariate-based constrained randomization of group-randomized trials. *Clin Trials* 2004; **1:** 297–305.

72 Li F, Turner EL, Heagerty PJ, Murray DM, Vollmer WM, DeLong ER. An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med* 2017; **36:** 3791–806.

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

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

75 Murphy S, Moore G, Tapper K, et al. Free healthy breakfasts in primary schools: a cluster randomised controlled trial of a policy intervention in Wales, UK. *Public Health Nutr* 2011; **14:** 219–26.

76 Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol* 2013; **66:** 752–58.

77 Eichner FA, Groenwold RHH, Grobbee DE, Oude Rengerink K. Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size. *J Clin Epidemiol* 2019; **107:** 89–100.