

Statistical power and validity of Ebola vaccine trials in Sierra Leone: a simulation study of trial design and analysis



Steven E Bellan, Juliet R C Pulliam, Carl A B Pearson, David Champredon, Spencer J Fox, Laura Skrip, Alison P Galvani, Manoj Gambhir, Ben A Lopman, Travis C Porco, Lauren Ancel Meyers, Jonathan Dushoff

Summary

Background Safe and effective vaccines could help to end the ongoing Ebola virus disease epidemic in parts of west Africa, and mitigate future outbreaks of the virus. We assess the statistical validity and power of randomised controlled trial (RCT) and stepped-wedge cluster trial (SWCT) designs in Sierra Leone, where the incidence of Ebola virus disease is spatiotemporally heterogeneous, and is decreasing rapidly.

Methods We projected district-level Ebola virus disease incidence for the next 6 months, using a stochastic model fitted to data from Sierra Leone. We then simulated RCT and SWCT designs in trial populations comprising geographically distinct clusters at high risk, taking into account realistic logistical constraints, and both individual-level and cluster-level variations in risk. We assessed false-positive rates and power for parametric and non-parametric analyses of simulated trial data, across a range of vaccine efficacies and trial start dates.

Findings For an SWCT, regional variation in Ebola virus disease incidence trends produced increased false-positive rates (up to 0.15 at $\alpha=0.05$) under standard statistical models, but not when analysed by a permutation test, whereas analyses of RCTs remained statistically valid under all models. With the assumption of a 6-month trial starting on Feb 18, 2015, we estimate the power to detect a 90% effective vaccine to be between 49% and 89% for an RCT, and between 6% and 26% for an SWCT, depending on the Ebola virus disease incidence within the trial population. We estimate that a 1-month delay in trial initiation will reduce the power of the RCT by 20% and that of the SWCT by 49%.

Interpretation Spatiotemporal variation in infection risk undermines the statistical power of the SWCT. This variation also undercuts the SWCT's expected ethical advantages over the RCT, because an RCT, but not an SWCT, can prioritise vaccination of high-risk clusters.

Funding US National Institutes of Health, US National Science Foundation, and Canadian Institutes of Health Research.

Introduction

At the peak of the devastating 2014–15 Ebola virus disease outbreak in parts of west Africa, international public health agencies and pharmaceutical companies committed themselves to assess the efficacy of several candidate vaccines. Even with the outbreak in apparent decline, vaccine trials could prove important to minimise future outbreaks of the disease. Various alternative vaccine trial designs were proposed, each striking a different balance between ethical, logistical, and statistical concerns. In February, 2015, an individually randomised controlled trial (RCT) was initiated in Liberia, and a ring vaccination trial began in Guinea in March, 2015.¹ A stepped-wedge cluster trial (SWCT) was originally proposed for Sierra Leone, but this design has been revised to a phased-rollout RCT, the implementation of which is imminent.² In an RCT, trial participants are randomly assigned at the individual level to a vaccine or control group. In an SWCT, all trial participants are vaccinated but in a random sequence of geographically distinct clusters of individuals. In the newly proposed ring vaccination trial design, contacts of incident Ebola virus disease cases are randomly assigned to be vaccinated either immediately (as in traditional ring vaccination strategies) or after some delay.³ Although the

SWCT is no longer planned, we assess the tradeoffs between this trial design and RCT designs in Sierra Leone to inform decisions during similarly challenging circumstances that might arise in future epidemics. We do not consider the ring vaccination trial, which was never proposed for Sierra Leone and which would need a different modelling framework to assess it than that used in our study.

When the risk of Ebola virus disease remains high, testing candidate vaccines with an RCT—in particular, assigning trial participants to a control group—can present an ethical problem. Phase 1–2 trial results suggest that candidate vaccines are safe and have a strong promise of efficacy. If the medical community believes that—in view of the high case-fatality rate of the disease—trial participants at substantial risk of infection are likely to fare better in the vaccinated group than in the control group, then an RCT might lack clinical equipoise.^{4,5} However, an uncontrolled trial would be susceptible to confounding bias, which would erode the reliability of resulting vaccine efficacy estimates.

In October, 2014, when the incidence of Ebola virus disease was still rising in west Africa, the US Centers for Disease Control and Prevention (CDC) proposed an SWCT in view of these concerns. In theory, SWCTs

Lancet Infect Dis 2015;
15: 703–10

Published Online
April 15, 2015
[http://dx.doi.org/10.1016/S1473-3099\(15\)70139-8](http://dx.doi.org/10.1016/S1473-3099(15)70139-8)

See [Comment](#) page 624

Center for Computational Biology and Bioinformatics (S E Bellan PhD) and Department of Integrative Biology (S J Fox BS, Prof L A Meyers PhD), The University of Texas at Austin, Austin, TX, USA; Department of Biology (J R C Pulliam PhD) and Emerging Pathogens Institute (J R C Pulliam, C A B Pearson PhD), University of Florida, Gainesville, FL, USA; School of Computational Science and Engineering (D Champredon MSc) and Department of Biology (J Dushoff PhD), McMaster University, Hamilton, ON, Canada; Center for Infectious Disease Modeling and Analysis, Yale School of Public Health (L Skrip MPH, Prof A P Galvani PhD) and Department of Ecology and Evolution (Prof A P Galvani), Yale University, New Haven, CT, USA; Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia (M Gambhir PhD); Division of Preparedness and Emerging Infections, National Center for Emerging and Zoonotic Infectious Diseases (M Gambhir) and Division of Viral Diseases, National Center for Immunization and Respiratory Diseases (B A Lopman PhD), Centers for Disease Control and Prevention, Atlanta, GA, USA; IHRC Inc, Atlanta, GA, USA (M Gambhir); Rollins School of Public Health, Emory University, Atlanta, GA, USA (B A Lopman); Francis I Proctor Foundation (Prof T C Porco PhD), Department of Ophthalmology (Prof T C Porco), and Department of Epidemiology & Biostatistics (Prof T C Porco),

University of California, San Francisco, CA, USA; and The Santa Fe Institute, Santa Fe, NM, USA (Prof L A Meyers)

Correspondence to: Dr Steven E Bellan, Center for Computational Biology and Bioinformatics, College of Natural Sciences, The University of Texas at Austin, 1 University Station C4500, Austin, TX 78712, USA
steve.bellan@gmail.com

Research in context

Evidence before this study

We searched PubMed for articles published from Jan 1, 1990, to March 3, 2015, using the following search terms: "(Ebola vaccine trial) OR (stepped wedge AND [statistic* OR power analysis OR Ebola OR infectious disease OR HIV OR dengue OR emerging OR outbreak OR epidemic OR pathogen])", excluding studies of non-infectious diseases, and we also identified relevant citations within selected articles. We did not identify any studies assessing Ebola vaccine efficacy trial design. We found methodological papers describing power analyses for stepped-wedge cluster trials (SWCTs) and randomised controlled trials (RCTs), and studies assessing trial design for evaluation of candidate vaccine efficacy for other infectious diseases. Of these reports, only two studies discussed the potential effects of spatiotemporal variation in infection risk on trial design; however, they did not assess the resulting effects on trial validity or power.

Added value of this study

In this study, we provide an assessment of Ebola candidate vaccine efficacy trial design and a comparison of RCT and SWCT designs in the context of spatiotemporally variable infection risk and realistic logistical constraints. We show that spatiotemporal variation in infection risk invalidates traditional

statistical analyses of SWCT designs and develop a permutation test that allows valid analysis. Our findings show that, under identical logistical constraints and within the present epidemiological context of Sierra Leone, an RCT has three-to-ten-times greater power than an SWCT to detect an effective vaccine, largely because of an RCT's ability to prioritise high-risk clusters for earlier enrolment. Finally, we argue that the SWCT loses its ethical advantages over the RCT if predictable heterogeneity in risk exists between clusters such that rollout of vaccination first to high-risk clusters is anticipated to be more effective than is vaccination of clusters in random order.

Implications of all the available evidence

An SWCT design to assess candidate Ebola vaccines in Sierra Leone will have little power to detect efficacy and might not have its anticipated ethical advantages over other randomised trial designs. An RCT, by contrast, might have sufficient power to detect efficacy but must start soon to avoid substantial reductions in power as the epidemic declines. More generally, we note that researchers should be cautious when using SWCT or other crossover designs to assess interventions in the context of emerging infectious disease epidemics, or in other contexts with spatiotemporally variable infection risk.

allow comparison between randomised treatment assignments without delaying vaccination to any participants.^{6,7} If practical constraints, such as low availability of vaccines or trained personnel, restrict the delivery of vaccines, then vaccination of groups as quickly as possible while randomising their order could avoid the ethical dilemma of withholding vaccines while allowing not-yet-vaccinated participants to serve as an unbiased control population. However, two plausible scenarios exist under which an SWCT loses its ethical advantage over an RCT: first, if vaccine delivery is delayed deliberately to improve trial power, and second, if predictable heterogeneity in risk exists between clusters such that prioritising vaccination of high-risk clusters is anticipated to be more effective than is vaccinating clusters in a random order.

Although ethics and logistics govern the acceptability of trial designs, the size and structure of a trial determine its speed and success in assessing vaccine efficacy. Randomisation protects against confounding because, on average, random assignment of an intervention distributes known and unknown confounders equally between trial groups. However, randomisation alone does not ensure statistical validity. Another important concern is whether or not a study maintains its prespecified target false-positive (type I error) rate, usually set to 0.05, which is the probability that—in the absence of any effect—the study will, by chance alone, erroneously conclude that an effect is present. False-positive rates include spurious conclusions that the intervention decreases or increases

risk. Studies whose design produces a false-positive rate that is above this target value are invalid.⁸ Although other study characteristics can also invalidate a study, we assess validity with respect to the prespecified false-positive rate only. Inflation of the false-positive rate can arise from an inappropriate statistical model that overestimates the precision of the effect estimate. Importantly, this situation can occur even when estimates of an intervention effect remain unbiased, such as when the clustered nature of data is not properly accounted for.⁸

Finally, even valid trial designs might have insufficient statistical power to ascertain that a protective vaccine is indeed effective (a high type II error rate) and therefore waste valuable resources. With the assumption of identical trial populations, cluster-randomised designs (including SWCTs) typically have lower power than do individual-randomised designs (such as RCTs) because cluster-randomisation leaves similarities between individuals within groups, thus reducing the effective sample size.^{9,10}

In this Article, we compare the statistical validity and power for SWCT and RCT designs in Sierra Leone, where declining trends in the incidence of Ebola virus disease vary regionally.

Methods

In this study, we compare the false-positive rates and power of SWCT and RCT designs in four steps. First, we fit a stochastic exponential decay model to recent Ebola virus disease incidence trends in Sierra Leone and use the

model to project district-level incidence. Second, we simulate a trial population comprising several clusters, each of which is a geographically distinct high-risk subpopulation that has a temporally varying hazard rate based on our district-level incidence projections. Third, across a range of assumed vaccine efficacies and for 600 000 synthetic trial populations, we simulate both RCT and SWCT designs. Finally, we analyse the simulation data with parametric and non-parametric tests to estimate vaccine efficacy, assessing the false-positive rates and statistical power of trial designs and the corresponding analyses.

Projection of district-level incidence in Sierra Leone

Ebola virus infection risk is spatiotemporally heterogeneous—ie, both the present risk and the rate of decline vary regionally. To capture this variation, we used maximum likelihood to fit exponential decay functions to district-level incidence of Ebola disease in Sierra Leone,¹¹ from each district's peak incidence to the most recent data.¹² To project district-level incidence for the next 6 months, we sampled negative binomial random deviates around these decay curves that replicate the overdispersion in the recorded incidence data (figure 1, appendix p 20).

Simulation of trial populations

Each simulated trial population included 6000 individuals distributed into 20 clusters of 300 individuals, representing the structure of the trial originally planned

in Sierra Leone.¹³ Clusters represented high-risk subpopulations at distinct locations, such as personnel working in an Ebola treatment unit or a group of front-line caregivers within a district (eg, health-care workers, laboratory personnel, or burial team staff).¹⁴ We allowed for both cluster-level and individual-level variation in Ebola virus disease risk. Cluster-level variation and trends in infection risk were based on our district-level projections. Specifically, we assumed that each cluster lived in one of the districts in Sierra Leone, and we created cluster-level hazard trajectories by resampling district-level projections (figure 2A). We then assumed that, without effective vaccination, a proportion of the projected cases would occur in the trial population. Because an estimated 5·2% of Ebola virus disease cases in Sierra Leone occurred in health-care workers,¹⁵ we considered scenarios with this proportion set at 2·5%, 5%, 7·5%, and 10%. Individual-level variation in risk within clusters was modelled with a relative hazard ratio that was lognormally distributed around 1 with an SD of 1, to simulate biological or occupation-related differences in risk (figure 2B).

Simulation of trial designs

We simulated both RCT and SWCT designs within trial populations with risks of infection as described previously, with a trial start date of Feb 18, 2015, and duration of 6 months. In a sensitivity analysis, we varied the start dates from Jan 15 until April 1, 2015.

See Online for appendix

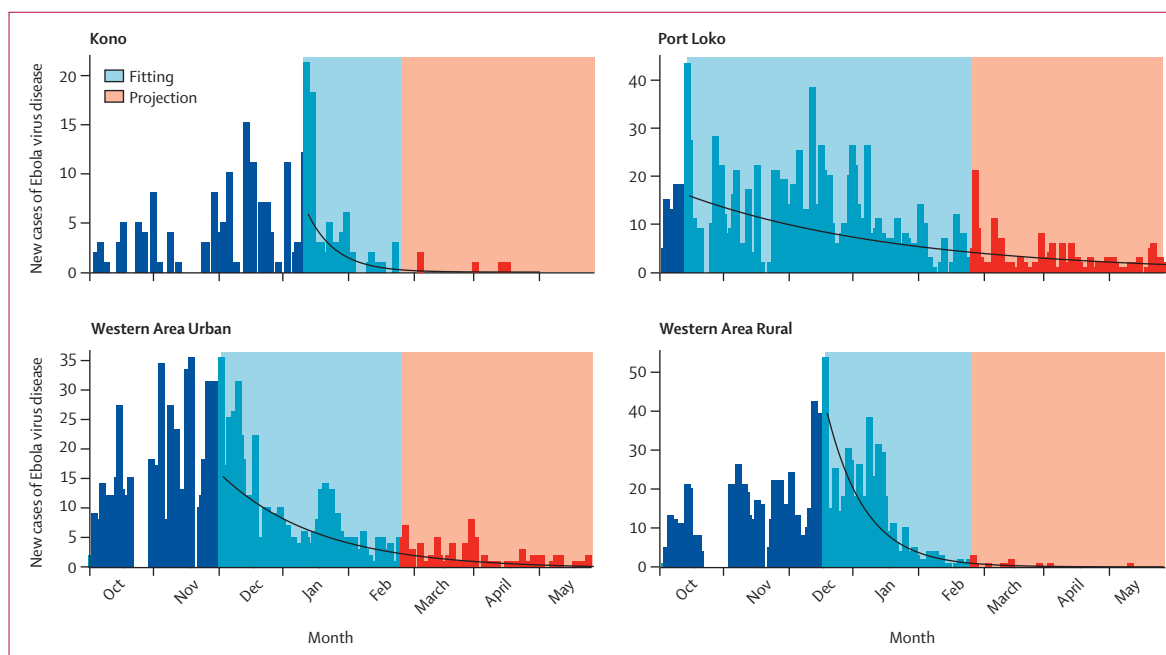


Figure 1: Fitted incidence projection models

Exponential decay functions (black line) fitted to incidence of Ebola virus disease (blue bars) within four districts of Sierra Leone, with example stochastic incidence projections (red bars). The models were fitted to district-level incidence data from the local peak until Feb 9, 2015, with negative binomial distributions. We estimated an average negative binomial overdispersion parameter of 1·2 across districts and used this estimate in projections. The timescale of the charts is October, 2014–May, 2015.

Because an SWCT is only ethically justified when logistical delays impede simultaneous vaccination of all individuals, comparisons between SWCT and RCT

designs should assume that the same logistical constraints apply to an RCT (ie, an RCT with phased vaccine rollout). We assumed that only one cluster could be vaccinated each week.¹³ However, because an RCT would vaccinate half of the individuals in each cluster, this rate of rollout would accomplish half the rate of vaccination (150 people per week) in an RCT compared with an SWCT (300 people per week). If the limiting factor is the number of individuals rather than the number of clusters vaccinated each week, then an RCT could vaccinate people at the same rate as an SWCT. To address this scenario, we also simulate a fast RCT, in which half of each of two clusters are vaccinated per week (figure 3; appendix p 10).

In the SWCT, by definition, clusters are vaccinated in random order to allow unbiased comparison between vaccinated and not-yet-vaccinated clusters, whereas in an RCT, we assume that clusters are either vaccinated in a random order or prioritised according to recently estimated risk. In the latter case, each week the cluster (or two clusters for a fast RCT) with the highest incidence 2 weeks previously is added to the trial. For comparison with these phased-rollout RCTs, we also considered an ideal scenario, free from logistical constraints, in which an RCT could immediately vaccinate half the trial population (simultaneous instant RCT). Notably, this situation cannot be compared fairly to an SWCT, which is predicated on the necessity of delayed rollout.

We assumed a 21-day delay between vaccination and the development of protective immunity (hereafter referred to as protective delay) but did a sensitivity analysis that considered shorter delays.^{2,16,17} We then simulated individual infections based on the hazard models described previously and individual immune status. We considered scenarios with vaccine efficacies of 0, 0.5, 0.7, or 0.9, and did not include the indirect benefits of vaccination (ie, herd immunity) within a cluster, based on the evidence that health-care workers, although at high risk of infection, rarely infect each other.¹⁸

Statistical analysis

Analyses of RCTs included only person-time within a cluster following the development of protective immunity in vaccinated individuals therein. Analyses of SWCTs included all person-time except for the protective delay, with other options explored in a sensitivity analysis (figure 3; appendix pp 11–13). We analysed simulated trial data with three types of approaches: semiparametric or parametric regression (ie, the Cox proportional hazards mixed-effect frailty model and a Poisson regression model with cluster-level effects); and two non-parametric methods based on estimates from these same regressions: bootstrap tests and permutation (randomisation) tests. For all tests, statistical significance was established through the use of a target two-tailed false-positive rate of $\alpha=0.05$, yielding a one-tailed cutoff of 0.025 for vaccine

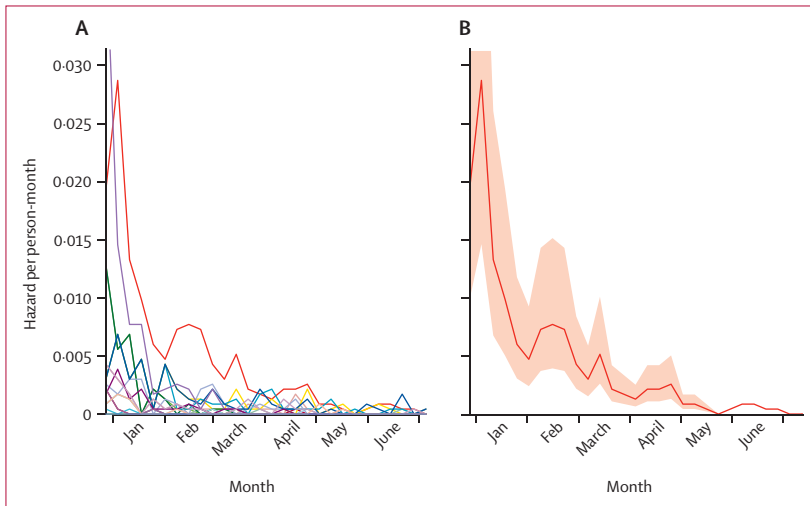


Figure 2: Ebola risk projections in simulated trial populations (A) Cluster-level projections. An example of simulated weekly Ebola virus infection hazard rate for 20 clusters based on district-level projections from Sierra Leone under the assumption that, in the absence of vaccination, 5% of district-level cases would occur in trial clusters within each district. (B) Individual-level variation. Mean (red line) and IQR (shaded area) of infection risk across individuals in one example cluster. The timescale of the graphs is January–June, 2015.

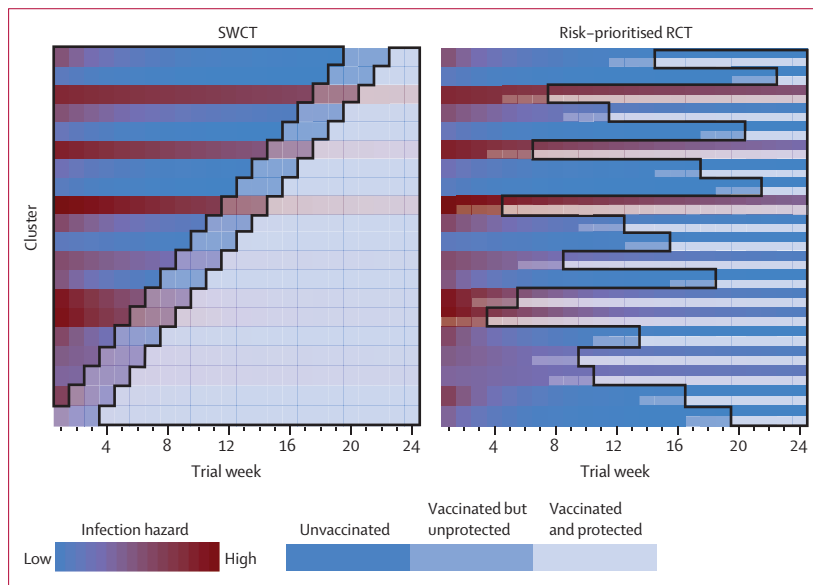


Figure 3: Trial designs Diagrammatic representation of the SWCT and the risk-prioritised RCT on the same trial population, with the assumption that only one cluster can receive vaccination per week. Each row corresponds to one cluster over the 24 weeks of the trial, with colour representing the level of infection risk within the cluster (ranging from red [high risk] to blue [low risk]) and transparency indicating vaccination status (opaque: unvaccinated; intermediate: vaccinated but not yet protected; transparent: vaccinated and protected). Areas outlined in black indicate the person-time at risk analysed in either design, with both analyses excluding the 21-day delay between vaccination and protection. In the SWCT, entire clusters are vaccinated in random order to allow unbiased comparison between vaccinated and not-yet-vaccinated clusters. In the risk-prioritised RCT, half of the individuals in each cluster are chosen at random to be vaccinated and we assume that clusters are prioritised in order of recently estimated risk (the reddest cluster 2 weeks preceding each round of vaccination). For diagrams of the random-ordered and simultaneous instant RCT and the fast RCT in the same trial population and alternative analyses of the SWCT, see appendix pp 10–11. SWCT=stepped wedge cluster trial. RCT=randomised controlled trial.

efficacy. For each of 300 combinations of assumed vaccine efficacy, protective delay, trial design, start date, and hazard levels, we simulated 2000 trials (totalling 600 000 trials), with each trial based on a unique set of stochastically generated district-level Ebola virus disease projections for Sierra Leone. All codes necessary to replicate these analyses and a more detailed discussion of the methods are provided in the appendix.

Role of the funding source

The funding sources had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The authors were not paid by any pharmaceutical company or other agency to write this report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

First, we consider trials of a vaccine with no effect on infection risk and assess the false-positive rate—ie, the frequency at which trials erroneously conclude that the vaccine affects risk of infection. All RCT scenarios yielded false-positive rates lower than the target $\alpha=0.05$. By contrast, SWCTs showed increased false-positive rates as high as 0.09 when analysed with the Cox proportional hazards mixed-effect frailty model, 0.11 with Poisson regression, and 0.15 when analysed with bootstrap methods (figure 4). These high rates were caused by type I errors in both directions (ie, erroneous conclusions that the vaccine either decreases or increases risk when it actually does not affect it). The false-positive rate of the Cox proportional hazards mixed-effect frailty model increased with Ebola incidence in the trial population (figure 4A) and arose from cluster-level variation in hazard trends (appendix pp 15–16). By contrast, the false-positive rate of the bootstrap approach decreased with increasing incidence of Ebola (figure 4B) and was unrelated to hazard trends or variation therein (appendix p 16). Permutation test analyses of SWCTs maintained the prespecified target false-positive rate (figure 4C).

We focused our subsequent analysis of statistical power on two methods that retained the target false-positive rate—specifically, the Cox proportional hazards mixed-effect frailty model for RCTs and permutation tests for SWCTs (figure 5). For each design, power is largely determined by the number of cases recorded in the trial (appendix pp 17–18).

However, power differs greatly between trial designs. An RCT that rolls out vaccination to clusters in a random order provides modest gains in statistical power over an SWCT, despite accumulating vaccinated person-time at half the rate of an SWCT, in which the full cluster is vaccinated. When risk varies both in time and across clusters, the RCT allows more direct and sustained comparison of vaccinated and unvaccinated individuals in similar high-risk settings. The RCT also maintains a

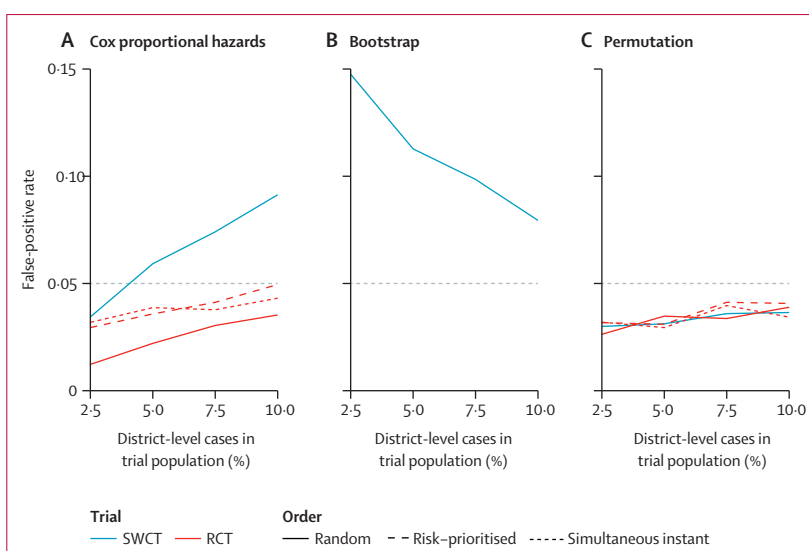


Figure 4: False-positive rates by trial design and analysis

False-positive rates are shown by trial design, analytical method, proportion of district-level cases occurring in trial participants, and order of cluster vaccination (for the RCT). Increases above the target false-positive rate of 0.05 (horizontal dashed line) indicates an invalid study for SWCT when analysed with a Cox proportional hazards frailty model (A) or with cluster-level bootstrap CIs (B), whereas all analyses of RCTs or of an SWCT with a permutation test (C) maintain prespecified false-positive rates. Line types indicate whether RCT clusters are vaccinated in random order (random; solid line), in order of highest risk (risk-prioritised; dotted line), or all simultaneously at the start of the trial (simultaneous instant; dashed line). Underlying hazards are based on district-level incidence projections (figure 2). We did not apply cluster-level bootstrapping to the RCT designs because they rely on individual-level randomisation. The fast RCT results are not shown here but were qualitatively similar to those of the RCT. The simultaneous instant design is a hypothetical idealised design and is presented here for comparison, but is not regarded as a feasible trial design in this context. SWCT=stepped-wedge cluster trial. RCT=randomised controlled trial.

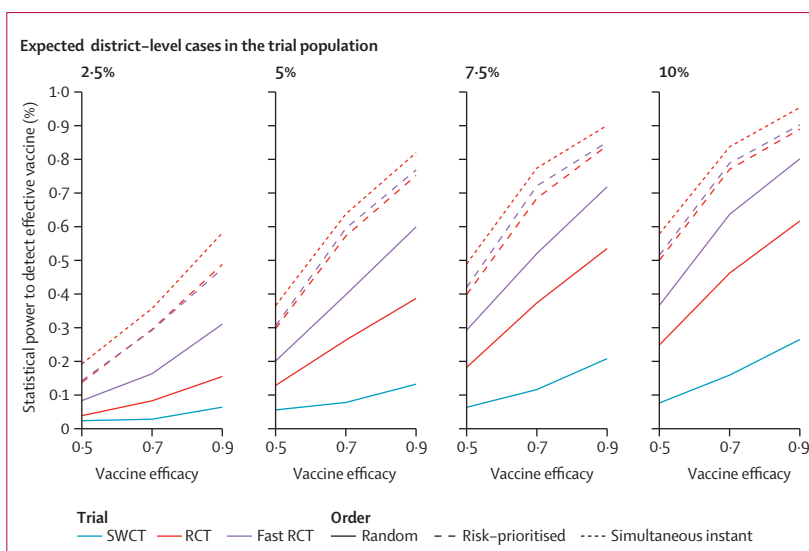


Figure 5: Statistical power by trial design and expected infection risk to trial participants

Panels differ by expected proportion of district-level cases occurring in trial participants in the absence of vaccination. Line types indicate whether RCT clusters are vaccinated in random order (random; solid line), in order of highest risk (risk-prioritised; dotted line), or all simultaneously at the start of the trial (simultaneous instant; dashed line). RCT or fast RCT data are analysed with a Cox proportional hazards frailty model; SWCT data are analysed with a permutation test over estimates from the Cox proportional hazards frailty model. The simultaneous instant design is a hypothetical idealised design and is presented here for comparison, but is not regarded as a feasible trial design in this context. SWCT=stepped-wedge cluster trial. RCT=randomised controlled trial.

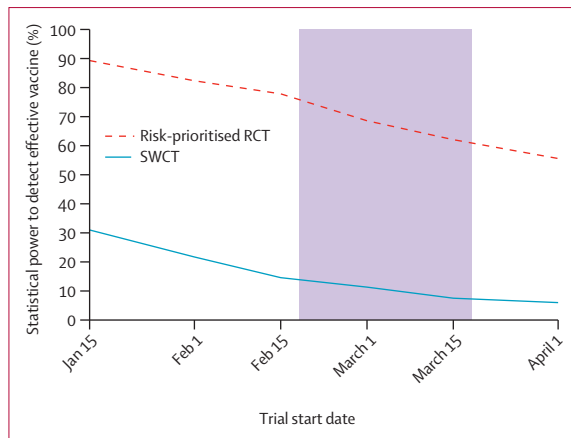


Figure 6: Statistical power by trial design and start date

Estimated power by trial start date, under the assumption that, in the absence of vaccination, 5% of district-level cases occur in the trial, the candidate vaccine is 90% effective, and the vaccine protective delay is 21 days. The shaded area highlights the effect of a 1-month delay in the trial start date. SWCT=stepped-wedge cluster trial. RCT=randomised controlled trial.

balanced ratio of person-time in vaccinated and control groups, by contrast with the SWCT, in which this ratio changes throughout the trial. A fast RCT, in which vaccinated person-time accumulates at the same rate as the SWCT, has even greater power than a normal RCT. For RCT designs, prioritisation of high-risk clusters for vaccination increases power; risk-prioritised RCT and fast RCT designs achieve similarly high power, which is substantially higher than that of either random-order design. With the assumption of a trial start date of Feb 18, 2015, and duration of 6 months, we estimate the power to detect a 90% effective vaccine to be between 49% and 89% for a risk-prioritised RCT, and between 6% and 26% for a SWCT, depending on the proportion of incidence that occurs within the trial population. Under the assumption that 5% of district-level cases occur in the trial population, we estimate that a 1-month delay in trial start date will reduce the power of RCT designs from 75% to 62% and of SWCT designs from 13% to 8% (figure 6). Under this same assumption, assessment of a vaccine with a 5-day versus 21-day protective delay would increase the power of RCT and SWCT designs from 75% to 81% and from 13% to 22%, respectively (appendix p 19).

Discussion

The spatiotemporal variation of infection risk in Sierra Leone's Ebola virus disease epidemic has a major effect on both the validity and power of vaccine trial designs. Cluster-level variation in the incidence of Ebola trends increases false-positive rates when standard statistical methods are applied to SWCTs. Inflation above the prespecified target false-positive rate suggests that study results are less conservative than intended, although we show that a permutation test can remedy this issue. Nevertheless, in such spatiotemporally variable settings,

the power of an SWCT to detect an effective vaccine is three-to-ten-times lower than that of a risk-prioritised RCT in the same trial population, given identical logistical constraints.

Although an SWCT, by design, must vaccinate clusters in random order, an RCT can vaccinate clusters in order of the highest to lowest risk, thereby providing the most information about vaccine efficacy. In fact, prioritisation of high-risk clusters increased power far more than did rolling out vaccines at double the speed in a fast RCT. Thus, a risk-prioritised RCT might still have sufficient power to definitively identify an effective vaccine, although power will continue to decrease rapidly if incidence declines continue at present rates. The imminent CDC-led phased-rollout design is very similar to the risk-prioritised RCTs simulated here and will include several geographically distinct clusters of front-line caregivers in regions of ongoing Ebola virus transmission.² Individuals within clusters will be randomly assigned to receive immediate vaccination or delayed vaccination after 6 months (without placebo), with vaccine rolled out to clusters sequentially at the fastest logistically feasible speed. The design allows flexible addition of clusters to improve power and vaccine distribution as transmission patterns shift—an approach that would not be possible in an SWCT, which needs random a-priori specification of the vaccination rollout sequence.

Despite the statistical advantages of an RCT, the SWCT might remain preferable in some scenarios. In weighing up practical and ethical considerations at the height of the Ebola epidemic, policy makers proposed an SWCT for Sierra Leone in response to the concern that an RCT would withhold potentially lifesaving interventions from the control group (and would therefore lack equipoise). The ethical linchpin of the SWCT is that it delivers vaccines as quickly as possible to maximise public health benefits, while monitoring vaccine efficacy as a non-competing secondary objective. Any delays or modifications to vaccine deployment for the sake of improving statistical power would undermine its ethical foundation. We argue that the ongoing decline and spatial variation in risk in Sierra Leone presents not only statistical hurdles but also an ethical challenge for the SWCT design. If the goal is to maximise the public health effect, then clusters or individuals should be prioritised according to their risk of infection, which is not allowed in an SWCT. This approach would not, however, pose a problem in scenarios where infection risk is homogeneous or unpredictable across clusters. We conclude that the RCT is the more promising design for Sierra Leone, in view of its greater statistical power than the SWCT and the absence of ethical advantage for the SWCT.

However, permissibility of an RCT relies on its own equipoise considerations. We note that RCT equipoise is a function of anticipated protective or adverse effects of vaccination, combined with infection risk, which modulates their relative importance.¹⁹ For example,

equipoise is more achievable in low-risk settings because the potential risks and benefits of vaccination are more balanced than in high-risk settings. Thus, in view of the substantial declines in Ebola virus disease incidence in Sierra Leone, the imminent phased-rollout RCT might now be ethically viable. Finally, we reiterate our claim¹⁹ that, for fatal diseases like Ebola virus disease, equipoise is easier to achieve when other life-saving resources are dedicated to caring for trial participants who become infected.

Our power estimates are based on epidemic projections from a simple model that assumes that the incidence of Ebola virus disease is falling exponentially in all regions, albeit at different rates. These projections should not be taken out of context, since they merely extrapolate recent trends, which might change. We did not aim to forecast incidence accurately; rather, we aimed to assess vaccine trial power and validity in the context of realistic spatiotemporal variation. An increase in Ebola virus disease incidence would increase the power of any trial, whereas a sudden end to the epidemic would decrease the power. Our straightforward model captures realistic variation in infection risk at both the individual and cluster level, but does not consider underlying stochastic dynamics in the transmission process, movement of Ebola virus disease cases between districts, or the indirect benefits of vaccination within a cluster.

The most important determinant of power is the expected number of cases that would occur in the trial population (in the absence of vaccination), which is established both by the trial population size and by the infection risk experienced by individuals in the trial. We have considered a single trial population size, and we have done a sensitivity analysis in which the hazard of infection varies. In our analysis, the risk of infection in trial participants was established by both district-level incidence projections and the proportion of cases assumed to occur within the trial population. To account for variation in the former, we implemented stochastic projections and assumed random distribution of trial clusters across districts. For variation in the proportion of cases assumed to occur in the trial population, we considered a range of values, based on the proportion of total cases occurring within health-care worker populations so far.

Several other important design considerations exist that are not included in our models. For example, RCT—but not SWCT—designs can use placebo or comparator vaccines to avoid bias induced by behavioural changes in participants when they become aware of their vaccination status. However, an SWCT could be more acceptable for high-risk communities, in which vaccination of only half of trial participants might be unacceptable. The greater power of an RCT over an SWCT could also increase the speed with which a vaccine is definitively identified as effective and subsequently rolled out to the wider population.

Crossover cluster-randomised designs (ie, those in which each cluster is observed both pre-intervention and post-intervention), such as the SWCT, are based on the ability to make comparisons between clusters that control for underlying time trends. We show that, when trends differ between clusters, the statistical validity of this approach is at least partly compromised because of confounding between the timing of intervention and time-dependent changes in risk to clusters. In particular, classical regression approaches, including Cox proportional hazards or generalised linear mixed models, that ignore this time dependency within clusters have inflated false-positive rates. This problem might be more common than has previously been recognised, especially during acute infectious disease epidemics or outbreaks, in which risk is highly variable in space and time.^{20,21} For example, this issue could be a concern for the ring vaccination trial planned in Guinea, which is another trial with a crossover cluster-randomised design. In addition to permutation tests, other non-parametric approaches have also been suggested to handle complex, and potentially unknown, spatiotemporal dependencies in risk.²²

Our findings show the usefulness of simulation approaches to assess trial designs and analyses under realistic and variable scenarios. Classical power analyses rely on analytical calculations that make strong assumptions, such as the absence of spatiotemporal dependency in infection risk. Analysis of estimates from simulated data, in which the underlying true parameters are known, provides a powerful device with which to discover biases and assess the robustness of estimators when not all assumptions are met. Here, we have leveraged modern computational approaches to identify problems with conventional methods and new analytical approaches that can be used to resolve them.

In conclusion, because of the observed temporal and geographical variation in infection risk, an SWCT design to assess candidate Ebola vaccines in Sierra Leone would have little power to detect efficacy and might not retain anticipated ethical advantages over other trial designs. By contrast, an RCT might have sufficient power to detect efficacy but needs to be started soon to avoid substantial reductions in power as the epidemic declines. Adaption of the basic RCT design to prioritise high-risk clusters substantially increases statistical power and ensures rapid distribution of a potentially effective vaccine to the groups that would benefit most.

Contributors

SEB, JRCP, CABP, MG, BAL, TCP, LAM, and JD developed the conceptual modelling framework. SEB and CABP did the analyses and created the figures. SEB, SJF, and LS reviewed the published literature. SEB wrote the first draft. All authors contributed to the interpretation and presentation of results, and the writing and approval of the final report.

Declaration of interests

We declare no competing interests.

Acknowledgments

SEB, SJF, LS, APG, and LAM were supported by a National Institute of General Medical Sciences MIDAS grant to LAM and APG (U01GM087719). JRCP is supported by a National Science Foundation Rapid Response Research Program grant and the Research and Policy on Infectious Disease Dynamics Program of the Fogarty International Center, National Institutes of Health and Science and Technology Directorate, Department of Homeland Security. SEB and JRCP were additionally supported by the International Clinics on Infectious Disease Dynamics and Data (ICI3D) programme, which is funded by a National Institute of General Medical Sciences of the National Institutes of Health grant (R25GM102149) to JRCP and Alex Welte. DC and JD are supported by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada. LS and APG were also supported by a National Science Foundation Rapid Response Research Program grant (1514673). TCP was funded by National Institute of General Medical Sciences MIDAS grant U01GM087728. We thank Jason Asher and Molly Davies for useful discussions. This project arose as a consequence of discussions at the conference on Modeling the Spread and Control of Ebola in West Africa, which took place at the Georgia Institute of Technology in Atlanta (GA, USA) on Jan 22–23, 2015. We also acknowledge the Texas Advanced Computing Center at The University of Texas at Austin (TX, USA) for providing high-performance computational resources that have contributed to the research results reported within this Article. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References

- Mohammadi D. Ebola vaccine trials back on track. *Lancet* 2015; **385**: 214–15.
- Centers for Disease Control and Prevention. STRIVE (Sierra Leone Trial to Introduce a Vaccine Against Ebola). <https://clinicaltrials.gov/ct2/show/NCT02378753> (accessed March 20, 2015).
- Enserink M. High hopes for Guinean vaccine trial. *Science* 2015; **347**: 219–20.
- Van der Graaf R, van Delden JJM. Equipoise should be amended, not abandoned. *Clin Trials* 2011; **8**: 408–16.
- Adebamowo C, Bah-Sow O, Binka F, et al. Randomised controlled trials for Ebola: practical and ethical issues. *Lancet* 2014; **384**: 1423–24.
- Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006; **6**: 54.
- Mdege ND, Man MS, Taylor C, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol* 2011; **64**: 936–48.
- Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978; **108**: 100–02.
- Kerry SM, Bland JM. Sample size in cluster randomisation. *BMJ* 1998; **316**: 549.
- Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol* 1999; **28**: 319–26.
- Humanitarian Data Exchange. Sub-national time series data on Ebola cases and deaths in Guinea, Liberia, Sierra Leone, Nigeria, Senegal and Mali since March 2014. 2015. <https://data.hdx.rwllabs.org/dataset/rowca-ebola-cases> (accessed Feb 20, 2015).
- Pearson C, Porco TC. Web-based Ebola forecasting tool. 2014. http://ebola.ici3d.org/ebola_forecast (accessed Feb 10, 2015).
- Lopman B. Can effectiveness be accurately estimated using a phased introduction of an Ebola vaccine? In: Modeling the Spread and Control of Ebola in West Africa. Atlanta, 2015. <http://bioinformatics.gatech.edu/ebola-modeling-workshop-program> (accessed Feb 10, 2015).
- Cohen J, Kupferschmidt K. Ebola vaccine trials raise ethical issues. *Science* 2014; **346**: 289–90.
- Kilmarx PH, Clarke KR, Dietz PM, Hamel MJ, Husain F, McFadden JD. Ebola virus disease in health care workers—Sierra Leone, 2014. *Morb Mortal Wkly Rep* 2014; **63**: 1168–71.
- Geisbert TW, Feldmann H. Recombinant vesicular stomatitis virus-based vaccines against Ebola and marburg virus infections. *J Infect Dis* 2011; **204**: S1075–81.
- Rampling T, Ewer K, Bowyer G, et al. A monovalent chimpanzee adenovirus ebola vaccine—preliminary report. *N Engl J Med* 2015; published online Jan 28. DOI:10.1056/NEJMoa1411627.
- Faye O, Boëlle P, Heleze E, et al. Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect Dis* 2015; **15**: 320–26.
- Bellan SE, Pulliam JRC, Dushoff J, Meyers LA. Ebola virus vaccine trials: the ethical mandate for a therapeutic safety net. *BMJ* 2014; **349**: g7518.
- Wolbers M, Kleinschmidt I, Simmons CP, Donnelly CA. Considerations in the design of clinical trials to test novel entomological approaches to dengue control. *PLoS Negl Trop Dis* 2012; **6**: e1937.
- Van der Tweel I, van der Graaf R. Issues in the use of stepped wedge cluster and alternative designs in the case of pandemics. *Am J Bioeth* 2013; **13**: 23–24.
- Davies MM, Laan MJ Van Der. A new approach to variance estimation for time-ordered dependent data. *U.C. Berkeley Division of Biostatistics Working Paper Series*. Working Paper 322. <http://biostats.bepress.com/ucbbiostat/paper322> (accessed Feb 15, 2015).

For the Texas Advanced
Computing Center see
<http://www.tacc.utexas.edu>