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- 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 α=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.1 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 DiseaseControl and Prevention (CDC) proposed an SWCT in view of these concerns. In theory, SWCTs...
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.

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, 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.
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. 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.

**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
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 a 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
 Nevertheless, in such spatiotemporally variable settings, a permutation test can remedy this issue. Results are less conservative than intended, although we prespecified target false-positive rate suggests that study methods are applied to SWCTs. Inflation above the expected level increases false-positive rates when standard statistical methods are applied to SWCTs. Cluster-level variation in the incidence of Ebola trends on both the validity and power of vaccine trial designs.

Leone’s Ebola virus disease epidemic has a major effect. The spatiotemporal variation of infection risk in Sierra Leone has a major effect. The spatiotemporal variation of infection risk in Sierra Leone presents not only statistical hurdles but also an ethical challenge for the vaccination rollout sequence. 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 vaccination rollout sequence. Any delays or modifications to vaccine deployment for the sake of improving statistical power would undermine its ethical foundation. 

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. 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 vaccination rollout sequence. Any delays or modifications to vaccination 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 vaccination rollout sequence. Any delays or modifications to vaccine deployment for the sake of improving statistical power would undermine its ethical foundation.

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. For example,
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.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 (ICCID) 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 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