



# Using simulated infectious disease outbreaks to inform site selection and sample size for individually randomized vaccine trials during an ongoing epidemic

Zachary J Madewell<sup>1</sup> , Ana Pastore Y Piontti<sup>2</sup>, Qian Zhang<sup>2</sup>, Nathan Burton<sup>3</sup>, Yang Yang<sup>1</sup>, Ira M Longini<sup>1</sup>, M Elizabeth Halloran<sup>4,5</sup>, Alessandro Vespignani<sup>2</sup> and Natalie E Dean<sup>1</sup>

## Abstract

**Background:** Novel strategies are needed to make vaccine efficacy trials more robust given uncertain epidemiology of infectious disease outbreaks, such as arboviruses like Zika. Spatially resolved mathematical and statistical models can help investigators identify sites at highest risk of future transmission and prioritize these for inclusion in trials. Models can also characterize uncertainty in whether transmission will occur at a site, and how nearby or connected sites may have correlated outcomes. A structure is needed for how trials can use models to address key design questions, including how to prioritize sites, the optimal number of sites, and how to allocate participants across sites.

**Methods:** We illustrate the added value of models using the motivating example of Zika vaccine trial planning during the 2015–2017 Zika epidemic. We used a stochastic, spatially resolved, transmission model (the Global Epidemic and Mobility model) to simulate epidemics and site-level incidence at 100 high-risk sites in the Americas. We considered several strategies for prioritizing sites (average site-level incidence of infection across epidemics, median incidence, probability of exceeding 1% incidence), selecting the number of sites, and allocating sample size across sites (equal enrollment, proportional to average incidence, proportional to rank). To evaluate each design, we stochastically simulated trials in each hypothetical epidemic by drawing observed cases from site-level incidence data.

**Results:** When constraining overall trial size, the optimal number of sites represents a balance between prioritizing highest-risk sites and having enough sites to reduce the chance of observing too few endpoints. The optimal number of sites remained roughly constant regardless of the targeted number of events, although it is necessary to increase the sample size to achieve the desired power. Though different ranking strategies returned different site orders, they performed similarly with respect to trial power. Instead of enrolling participants equally from each site, investigators can allocate participants proportional to projected incidence, though this did not provide an advantage in our example because the top sites had similar risk profiles. Sites from the same geographic region may have similar outcomes, so optimal combinations of sites may be geographically dispersed, even when these are not the highest ranked sites.

**Conclusion:** Mathematical and statistical models may assist in designing successful vaccination trials by capturing uncertainty and correlation in future transmission. Although many factors affect site selection, such as logistical feasibility, models can help investigators optimize site selection and the number and size of participating sites. Although our study

<sup>1</sup>Department of Biostatistics, University of Florida, Gainesville, FL, USA

<sup>2</sup>Laboratory for the Modeling of Biological and Socio-Technical Systems, Northeastern University, Boston, MA, USA

<sup>3</sup>Institute for Child Health Policy, University of Florida College of Medicine, Gainesville, FL, USA

<sup>4</sup>Biostatistics, Bioinformatics, and Epidemiology Program, Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

<sup>5</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA

## Corresponding author:

Zachary J Madewell, Department of Biostatistics, University of Florida, PO Box 117450, 2004 Mowry Road, Gainesville, FL 32611-7450, USA.  
Email: zmadewell@ufl.edu

focused on trial design for an emerging arbovirus, a similar approach can be made for any infectious disease with the appropriate model for the particular disease.

## Keywords

Clinical trial design, forecast model, infectious diseases, mathematical modeling, simulations, trial planning, vaccine

## Background

To observe enough events to reliably measure the efficacy of a vaccine, phase III trials often enroll thousands or tens of thousands of participants across multiple sites. For endemic diseases like rotavirus or malaria, incidence may be low but is relatively predictable. Investigators can use historical data to guide the selection of trial populations assuming that future trends will be similar. Where incidence is lower than expected during the trial, investigators can expand the sample size at existing sites or increase participant follow-up to compensate. This strategy is unlikely to work for outbreak pathogens. Historical data may be only weakly predictive of future incidence at a location. In fact, for pathogens with high attack rates, an area with a large prior outbreak may be less susceptible to a subsequent outbreak if there is a build-up of population immunity. Alternatively, that area may be more prone to another outbreak if immunity wanes or the number of susceptible individuals is replenished. The outbreaks themselves are highly unpredictable—when and where they will occur, how many will become infected, and how long they will last. The 2014–2016 West African Ebola epidemic was emblematic of this challenge, with a Phase III trial in Liberia enrolling over 8000 individuals but observing no events because the local outbreak subsided.<sup>1</sup> In this situation, expanding enrollment at existing sites or extending follow-up of participants would not be able to compensate.

Novel strategies are needed to make vaccine trials more robust to the uncertain epidemiology of outbreaks, such as *Aedes aegypti*-transmitted viruses.<sup>2</sup> One recommended approach is to enable the addition of new sites over time using a master or core protocol framework.<sup>3</sup> If transmission in early hotspots is brought under control before the study has reached a conclusion, the trial can continue at new hotspots. If the outbreak is declared over, the trial can be paused until a subsequent outbreak. Spatially resolved mathematical and statistical forecast models can assist investigators in selecting participating sites.<sup>4</sup> Models can incorporate site-specific features such as population size and density, socioeconomic vulnerability, sociocultural acceptance, logistic feasibility, prior immunity estimated from traditional surveillance or serosurveys, ongoing local transmission, or risk of importation. For vector-borne diseases, models can capture vector presence or abundance, sensitivity to temperature and

humidity, the spread of other diseases by the same vector, and whether other diseases interfere with the disease of interest. By integrating diverse data sources, models can help investigators identify sites at highest risk of future transmission and prioritize these for inclusion in the trial.

Another advantage of simulation models for infectious disease trials is that they enable investigators to explore a range of trial design features.<sup>5,6</sup> Projected incidence is important, but so is the uncertainty around that projection, including the probability of no or little future spread. When there is a chance that sites will have little or no transmission, it becomes more important to include multiple, geographically dispersed sites, to distribute this risk. Even where models are not able to accurately project future incidence, models can be very valuable for trial planning if they can replicate uncertainty as an epidemiological feature of the pathogen.

We illustrate the potential role of forecast modeling by using simulation data from a stochastic, highly spatially resolved, agent-based Zika virus model<sup>7</sup> that was used to inform Zika vaccine trial planning in 2016.<sup>8</sup> Although the Zika epidemic subsided so that vaccine efficacy trials were not possible,<sup>9</sup> these are the type of data investigators would have at their disposal when designing future efficacy trials for other infectious diseases. In addition to generalizable findings, we provide a plan for how future trials may analyze their modeling results to prioritize test sites, site size, and the total number of sites. We explore how disease models can be used to address key trial design questions, including how to rank sites, the optimal number of sites to include, and how to allocate participants across sites. Simulations can also be used to explore trial feasibility given financial, logistical, or time constraints. We further consider how correlation between sites due to geographic proximity or human movement impacts trial power.

## Methods

### Model

We used the Global Epidemic and Mobility model to identify the top 100 sites in the Americas with the highest projected Zika virus probability of transmission and infection rates in 2017. These projections were prepared in 2016, reflecting the type of data available to

investigators planning trials. The Global Epidemic and Mobility model, which has been described elsewhere,<sup>7,10</sup> is a discrete stochastic epidemic computational model incorporating high-resolution demographic, socioeconomic, temperature, and vector occurrence probability data. This model includes a multiscale mobility model incorporating short (e.g. daily commute) and long-range (e.g. international flights) transportation networks to reflect interactions from human travel patterns. In addition to serving as a model basis for Zika virus simulations, this model has also been used for pathogens that spread by direct contact like influenza, Ebola, and SARS-CoV-2.<sup>11–14</sup> For example, model projections were demonstrated to show strong concordance with empirical data for 2009 H1N1 influenza.<sup>13</sup> This transmission model capturing the spatiotemporal spread of Zika virus enabled estimation of correlation in outcomes between neighboring sites. Before simulating epidemics, the parameters of the Global Epidemic and Mobility model were learned from real Zika epidemics during 2015–2018. We then fed the learned parameters into the model to simulate future outbreaks in the Americas to guide site selection. The details of how uncertainty in parameters and how the model is calibrated is described in Zhang et al.<sup>7</sup> The projections were calculated using discrete time steps of one day to simulate transmission dynamics, but the results are summarized as number of infections per month. The resulting dataset included the site name, population size, and number of simulated infections (both symptomatic and asymptomatic) by month from January through December 2017 (Table S1). Population sizes for sites included all ages. Thus, we can examine both the range of projections for each site, as well as look across sites within an epidemic.

### Trial design

We describe the design of a hypothetical individually randomized Zika vaccine efficacy trial. The primary outcome is total number of confirmed symptomatic Zika virus cases. Given a set of selected sites and a fixed enrolled population for each site, for which we consider various different combinations, we simulate a trial as follows. First, we select one of the simulated epidemics, which has an associated annual infection attack rate for each site. We simulate the number of infected trial participants at each site as a binomial draw with the probability of infection set at the site-level attack rate, and then we draw the number of these with symptomatic disease assuming 20% symptomatic proportion.<sup>15</sup> This yields the total number of cases at each site, which is then added across sites. We repeat the binomial draws 50 times at each site, and then across all simulated epidemics.

Approximately 60 symptomatic infection events are needed to have 90% power to reject the null hypothesis

that the vaccine efficacy  $\leq 30\%$  when it is actually 70% using a 1:1 allocation to vaccine or placebo. We therefore defined a successful (i.e. adequately powered) trial as finding  $\geq 60$  cases across all sites in 1 year; we also explored trial designs targeting 50 to 150 events.

In a sensitivity analysis, we consider the feasibility of trials when attack rates are uniformly lower than projected by the model. To explore this scenario, we restrict analyses to the 25% of simulated epidemics with lowest overall infection attack rates across all sites.

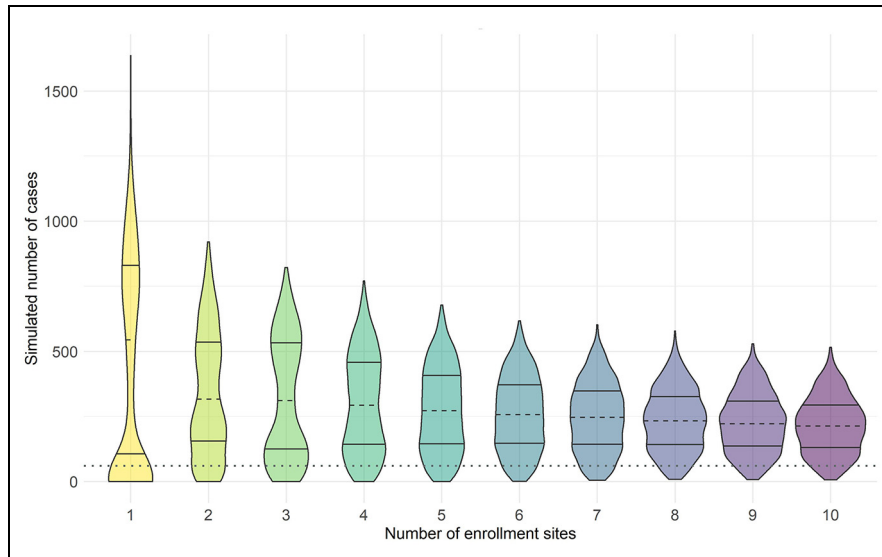
## Results

### Number of sites

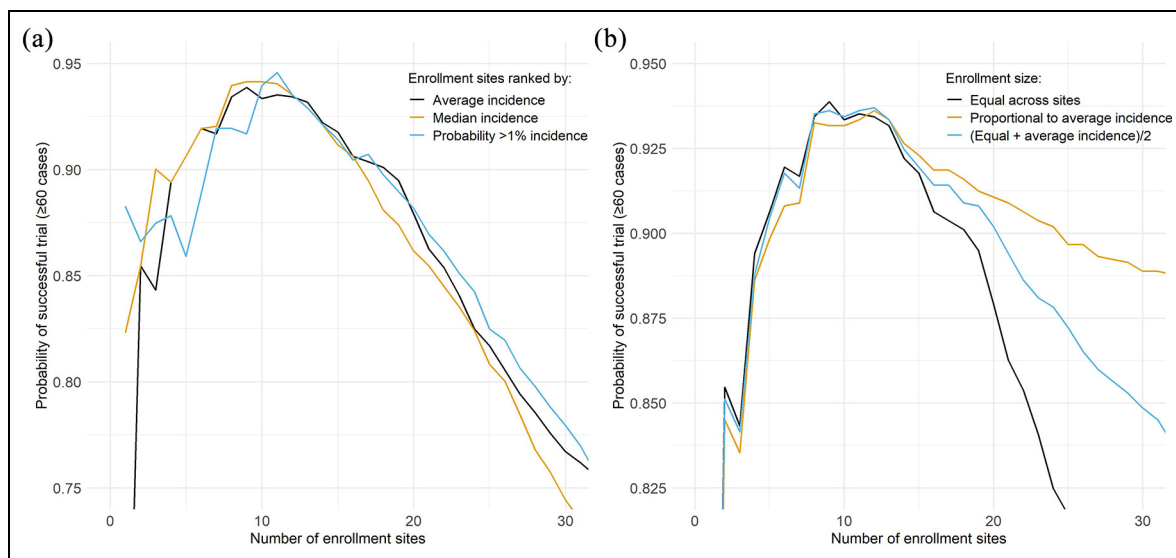
The first key design choice is the number of participating sites. Sites are ranked by mean incidence of infection across all simulated epidemics (Figure S1), and we consider designs including the top site, the top two sites, and so on. For this example, we constrain the overall sample size at 15,000 participants and allocate these participants equally across selected sites. We plot the distribution of the simulated number of cases for each design in Figure 1. Starting on the left side of the figure, the bimodal nature of outbreaks is apparent when five or fewer sites are included. While the median number of cases of the one-site design is highest relative to other designs, with a high upper tail observed for large outbreaks, there is notable mass near zero cases, when little transmission occurs at the site. As the number of sites increases, this bimodal phenomenon disappears; the probability of having zero cases decreases, but the median expected number of cases also decreases because lower incidence sites are included.

While it is theoretically possible to enroll from only a single site, this presents an unacceptable risk of failing to accrue the needed endpoints. It may also not be practically feasible if the site has a small population. Furthermore, while a very high attack rate in a trial could shorten the trial duration or increase study precision, our primary goal is to meet our target number of events, not dramatically exceed it. Thus, rather than median expected number of cases, it is preferable to examine the probability that the design is adequately powered. The curves in Figure 2(a) plot the probability of success (here defined as exceeding the target of 60 cases) as a function of the number of sites. We observe a local maximum around 8–11 sites, such that too few or too many sites are suboptimal with respect to trial success. In practice, the exact location of this maximum will depend on the specific epidemiological setting.

The location of this local maximum for number of sites is reasonably stable to the ranking criterion, even if we change the target number of events or total sample size. Figure 3 shows the minimum trial size required to achieve at least 90% probability of success for different target numbers of events and sites. As the target



**Figure 1.** Violin plot of the simulated number of Zika virus cases for the top 1–10 sites with the highest average site-level incidence of infection across all simulated outbreaks in 1 year (2017). We assume an enrolled population of 15,000 across all enrollment sites with enrollment size spread evenly across all sites. Median number of cases (dashed line), 25th and 75th percentiles (solid lines) are shown. The threshold for a successful trial, defined as  $\geq 60$  cases across all sites in 1 year, is indicated by the dotted line.

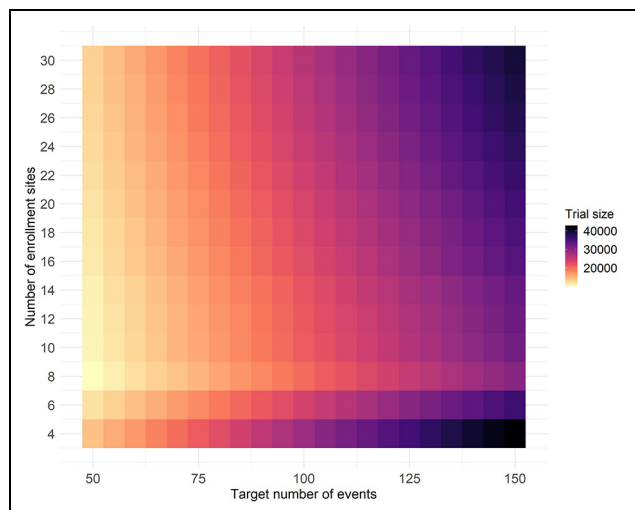


**Figure 2.** Probability of a successful trial (defined as  $\geq 60$  cases for an enrolled population of 15,000 across all enrollment sites in 1 year (2017) as function of the cumulative number of enrollment sites. In Panel (a), enrollment size was spread evenly across all sites and sites were added sequentially based on their ranking by the (1) average site-level incidence, (2) median incidence, and (3) probability of exceeding 1% site-level incidence of infection across all simulated outbreaks. In Panel (b), sites were added sequentially based on their ranking of average site-level incidence of infection across all simulated outbreaks and enrollment size was (1) spread evenly across all sites, (2) proportional to the average site-level incidence of each site, and (3) average of equal enrollment and proportional to mean incidence.

number of events increases, there is an expected increase in the total number of participants needed. Yet for any specified target number of events, the desired 90% probability of success is achieved with the smallest overall sample size when around eight sites are included. Too many sites include sites with lower risk

profiles, whereas too few sites risk having little transmission at those sites.

The optimal design depends on the underlying simulation data through the site-level attack rates. We consider a sensitivity analysis where the overall epidemic is smaller than projected by restricting to the 25% of



**Figure 3.** The minimum enrollment size required for an individually randomized vaccine efficacy trial to achieve at least 90% probability of success for different target numbers of events and numbers of sites. Sites were ranked by average site-level incidence of infection across all simulated outbreaks in 1 year (2017) with enrollment proportional to average site-level incidence of infection.

simulations with lowest overall infection attack rates across all sites. Figure 4 compares the probability of success of different designs for all simulations versus the low incidence subset. While many of the same relationships persists, the probability of success drops dramatically. Even increasing the total number of sites and overall sample size, the probability of success does not

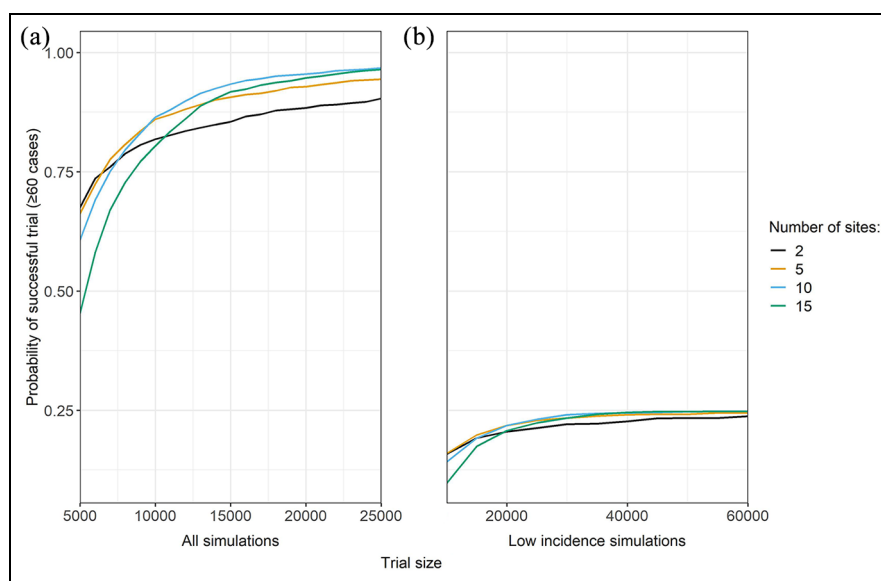
exceed 25% in the designs explored. Thus, this approach is also useful for exploring the feasibility of trials.

### Site prioritization

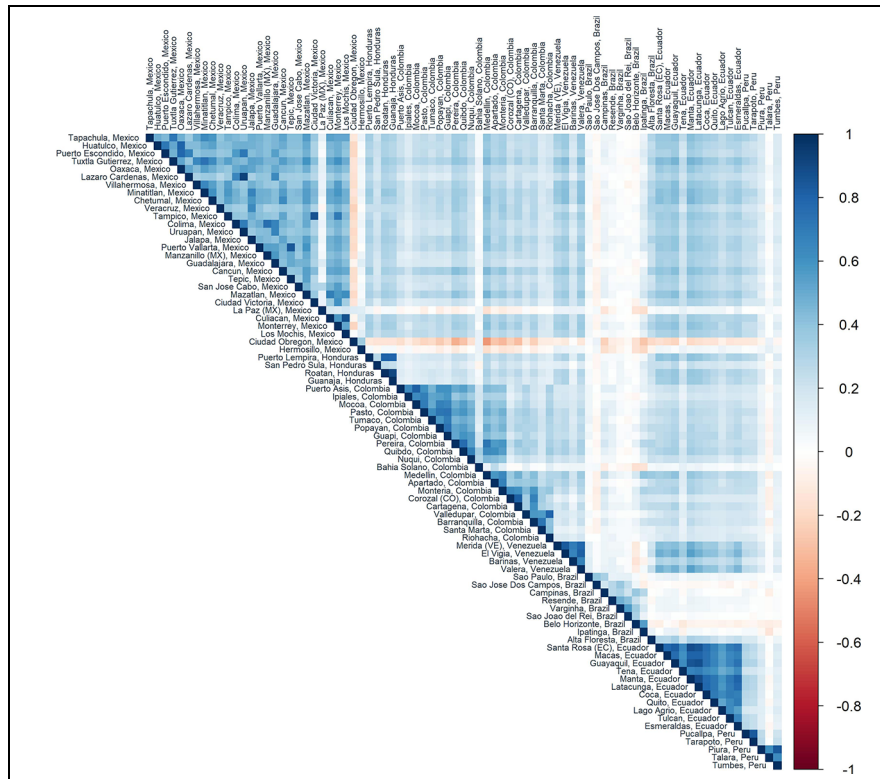
In the previous section, sites are ranked by average model-projected site-level incidence of infection. We examined other ranking strategies, including median model projected site-level incidence, and the proportion of simulated outbreaks where site-level incidence exceeds a threshold, such as 1%. The latter strategy is intended to capture the bimodal nature of outbreaks, and that a few very large outbreaks could drive a high average incidence. In general, these measures are well-correlated, but they can yield different rankings (Figure S2, Table S2). Small sites may have higher attack rates, but may also have a higher probability of having zero infections across all participants. Nonetheless, we found similar performance across the different ranking strategies (Figure 2).

### Allocation strategies

Next, we considered different strategies for allocating the total sample size across multiple sites, again considering designs including the top site with highest average model-projected site-level incidence, the top two sites, etc. The strategies are to distribute enrollment size (1) evenly across all sites, (2) proportional to mean incidence of infection, and (3) a middle-ground strategy using the average of sample sizes obtained from the



**Figure 4.** Probability of a successful trial (defined as  $\geq 60$  cases) as function of trial size, with enrollment size spread evenly across all sites. Sites were added sequentially based on their ranking by average site-level incidence of infection across all simulated outbreaks in 1 year (2017). Panel A includes all simulations, whereas Panel B is restricted to the 25% of simulated epidemics with lowest overall incidence across all sites.



**Figure 5.** Spearman's correlation of average site-level incidence of infection across all simulated outbreaks in 1 year (2017) between sites identified by the Global Epidemic and Mobility model (countries with at least four sites included). Sites are sorted by country and then by latitude.

previous two strategies (see File S1 for equations). An example with five sites is shown in Table S3. The difference between the strategies will depend upon how similar projected incidence is across the top-ranked sites.

In this example, enrolling participants proportional to average incidence did not outperform the other strategies when fewer than 15 sites were included (Figure 2(b)). This covers the range where the probability of success maximizes, around 8–11 sites. With larger numbers of enrollment sites, enrolling participants proportional to average incidence outperformed the other strategies as fewer individuals are enrolled from sites that are expected to have lower attack rates. However, designs with more sites are sub-optimal based on their reduced probability of success. As expected, the middle-ground strategy performs in between the others, but it may be desirable for logistical reasons by balancing enrollment across sites. The similar performance across the strategies for fewer than 15 sites may reflect that sites had similar enough risk and large enough uncertainty that proportional allocation provided no worthwhile advantage.

### Correlation between sites

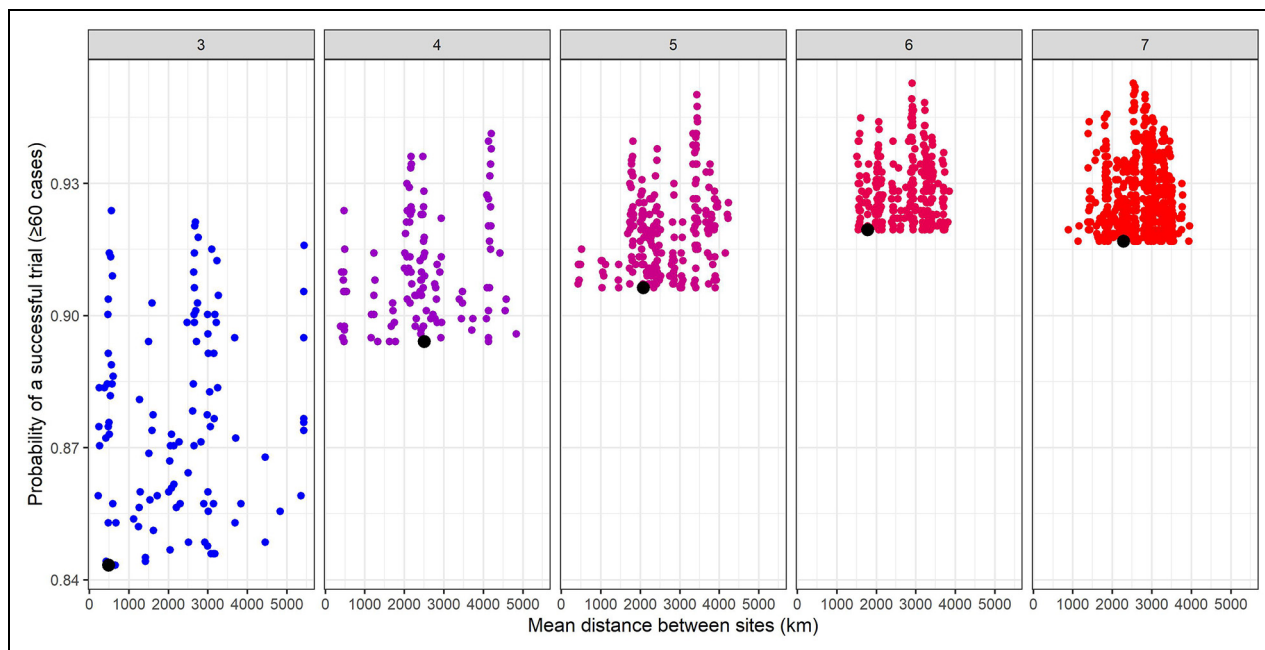
It is conceivable that incidence rates are similar among sites in the same geographic region. Aside from Brazil, Figure 5 shows that correlation in incidence is highest

among sites from the same country. It may therefore be that the sites with the highest simulated incidence are all from the same geographic region. To reduce the chance of enrolling sites from only one geographic area that may, by chance, have a smaller than expected outbreak, it may be prudent to simultaneously enroll participants from other geographically dispersed sites.

We explored whether alternative combinations of sites (that may be more geographically dispersed) could have a higher probability of success than those based only on rankings. Figure 6 displays combinations of three, four, five, six, and seven site trials that achieved higher probability of success than a design that selects sites based solely on average site-level incidence. For example, the top three sites based on incidence are relatively close together, as seen by the low mean pairwise distance between sites plotted on the X-axis. Many other combinations of three sites return higher probability of success, and these tend to be more geographically dispersed (higher mean pairwise distance). A similar pattern is observed for higher numbers of sites, although the gains become more modest.

### Number of countries

While it may be prudent to recruit sites from at least a few different countries (under the intuition that sites within a country are correlated), an important



**Figure 6.** Probability of a successful trial (defined as  $\geq 60$  cases for an enrolled population of 15,000 across all enrollment sites in 1 year) by the mean distance between sites (in kilometers). The panels represent numbers of sites and points represent combinations of sites from the top 15 sites with the highest average site-level incidence of infection across all simulated outbreaks in 1 year (2017) that had a higher probability of success than the combination of sites with the highest projected incidence (represented by the black dot). Enrollment is assumed to be spread evenly across sites. We have included combinations of 3–7 sites of the top 15 sites for illustration, but this process could be extended to any number of sites.

operational consideration is the number of countries enrolled. For each country included, the logistical burden for the trial increases substantially because it involves engaging with multiple ministries of health and the protocol needs to be approved by country-level institutional review boards. Thus, investigators may prefer to pursue trials in countries that have many high-risk sites. Figure S3 visualizes the number and incidence of sites by country. For example, in this scenario, it may be practical to select several sites from Peru, Mexico, Ecuador, and Colombia since many candidate sites are high risk.

## Conclusion

We describe the use of infectious disease modeling data to inform site selection and sample size planning for individually randomized vaccine trials during an ongoing epidemic. Mathematical models allow us to capture a range of possible outcomes, from small to large outbreaks, and incorporate correlation between sites connected by human movement. Models generate a stochastic distribution for future incidence beyond the mean estimate typically used in trial planning. This allows investigators to explore trial robustness to the considerable risk of no or very low future transmission at sites.

We find that the optimal number of sites to enroll represents a balance between mitigating the risk of a smaller than expected outbreak at any one site, versus enrolling participants from lower risk sites.<sup>7</sup> This optimal number of sites stayed relatively constant even when increasing the targeted number of events. In our example, different methods of prioritizing sites (by average incidence, median incidence, and probability of exceeding a threshold) returned different rankings but overall performed similarly as they generally returned the same set of prioritized sites. Investigators can prioritize participant enrollment from high-risk sites, though this may only be worthwhile when there are large differences in risk across sites.

We used a single model to identify the top 100 sites in the Americas with the highest projected Zika virus transmission probability and infection rates in 2017 and leveraged those data to analyze vaccine efficacy trial design strategies. However, ensemble forecast modeling (combining projections from independent modeling groups) has been shown to be overall more robust than relying on a single model.<sup>16</sup> Another advantage of ensemble modeling is that it enables projections to be compared across models.<sup>4,8</sup> We may feel more confident about prioritizing sites for inclusion or larger enrollment that are consistently ranked highest. Where there is considerable disagreement across models, or

where real-time validation indicates that models are not performing well enough, we may opt to rely less upon models for planning. In many real-life settings, model uncertainty may be too high to provide stable and reliable rank-orderings. The main value of models then is to characterize uncertainty, in order to assess whether the trial has enough sites to feasibly accrue the targeted number of events.

Of course projected incidence is not the only consideration when selecting sites. Vaccine efficacy trials are logistically complex, and a site with existing clinical trial infrastructure could be far preferable even if it is lower risk. Practical considerations include the availability of technical support, records maintenance, cold chain storage, and commitment from regulatory authorities to a speedy and high quality review.<sup>17</sup> Model projected incidence is thus viewable as one factor among a larger set that must be weighed together. One strategy is to use models as a first step to screen sites that is then followed by site visits to explore logistical capacity, collect additional surveillance data, or conduct a baseline serosurvey to better understand risk.

We use a model that captures correlation between sites due to shared characteristics and connectivity by human movement. We demonstrated that optimal combinations of sites may be those that are geographically dispersed to reduce this dependency. However, including multiple countries increases trial costs and logistical complexities. The problem has parallels with the design of cluster randomized trials, where there is a trade-off between the cost of adding new clusters and statistical efficiency. Like cluster randomized trials, these decisions can be explored in a cost-effectiveness analysis where the costs of engaging new countries or regions are quantified and compared.<sup>18</sup>

Our simulations had several limitations. Our model did not explicitly account for vaccination at the sites, which could impact projected incidence if population coverage is sufficiently high.<sup>19</sup> Though the models should integrate all available data that is predictive of risk, there may be other interventions such as vector control or education at specific sites or countries that affect their risk profiles that are not captured. In practice, sites that are projected to have high incidence may also institute other preventive measures (e.g. vector control) that reduce their risk profile independent of the model projections. It is important to conduct sensitivity analyses where incidence is lower than expected, as such a scenario presents the greatest barrier to establishing vaccine efficacy.

This study focused on trial design for an emerging arbovirus, but these methods can apply to vaccine trials for other communicable diseases, such as COVID-19. In using this approach for other diseases, a key factor is the predictability of future transmission and our ability to resolve differences in risk across sites. Yet, even where unpredictability is very high, models offer a

valuable structure for exploring trial design decisions while integrating infectious disease epidemiological insights.<sup>5</sup> In particular, models can help investigators optimize site selection and the number and size of participating sites.


### Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: A.P.Y.P., Q.Z., and A.V. report grants from Metabiota Inc, outside the submitted work.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institutes of Health R01-AI139761.

### ORCID iD

Zachary J Madewell  <https://orcid.org/0000-0002-5667-9660>

### Supplemental material

Supplemental material for this article is available online.

### References

1. Widdowson M-A, Schrag SJ, Carter RJ, et al. Implementing an Ebola vaccine study—Sierra Leone. *MMWR Suppl* 2016; 65: 98–106.
2. Dean NE, Gsell P-S, Brookmeyer R, et al. Design of vaccine efficacy trials during public health emergencies. *Sci Transl Med* 2019; 11: eaat0360.
3. Dean NE, Gsell P-S, Brookmeyer R, et al. Creating a framework for conducting randomized clinical trials during disease outbreaks. *Mass Medical Soc* 2020; 382: 1366–1369.
4. Dean NE, Piontti APY, Madewell ZJ, et al. Ensemble forecast modeling for the design of COVID-19 vaccine efficacy trials. *Vaccine* 2020; 38: 7213–7216.
5. Halloran ME, Auranen K, Baird S, et al. Simulations for designing and interpreting intervention trials in infectious diseases. *BMC Med* 2017; 15: 223.
6. Hitchings MDT, Grais RF and Lipsitch M. Using simulation to aid trial design: ring-vaccination trials. *Plos Negl Trop Dis* 2017; 11(3): e0005470.
7. Zhang Q, Sun K, Chinazzi M, et al. Spread of Zika virus in the Americas. *Proc Natl Acad Sci USA* 2017; 114: E4334–E4343.
8. Asher J, Barker C, Chen G, et al. Preliminary results of models to predict areas in the Americas with increased likelihood of Zika virus transmission in 2017. *bioRxiv*, 2017, <https://www.biorxiv.org/content/10.1101/187591v1>
9. Vannice KS, Cassetti MC, Eisinger RW, et al. Demonstrating vaccine effectiveness during a waning epidemic: a WHO/NIH meeting report on approaches to development and licensure of Zika vaccine candidates. *Vaccine* 2019; 37: 863–868.



10. Balcan D, Gonçalves B, Hu H, et al. Modeling the spatial spread of infectious diseases: the GLObal Epidemic and Mobility computational model. *J Comput Sci* 2010; 1: 132–145.
11. Poletto C, Gomes M, Piontti APY, et al. Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro Surveill* 2014; 19: 20936.
12. Gomes MFC, Piontti APY, Rossi L, et al. Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *Plos Curr* 2014; 6: 2360.
13. Tizzoni M, Bajardi P, Poletto C, et al. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med* 2012; 10: 165.
14. Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 2020; 368: 395–400.
15. Duffy MR, Chen TH, Hancock WT, et al. Zika virus outbreak on Yap Island, federated states of Micronesia. *N Engl J Med* 2009; 360: 2536–2543.
16. Reich NG, McGowan CJ, Yamana TK, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the US. *PLoS Comput Biol* 2019; 15: e1007486.
17. Kabineh AK, Carr W, Motevalli M, et al. Operationalizing international regulatory standards in a limited-resource setting during an epidemic: the Sierra Leone trial to introduce a vaccine against Ebola (STRIVE) experience. *J Infect Dis* 2018; 217: S56–S59.
18. Hayes RJ and Moulton LH. *Cluster randomised trials*. 2nd ed. Boca Raton, FL: CRC Press, 2017.
19. Hitchings MDT, Lipsitch M, Wang R, et al. Competing effects of indirect protection and clustering on the power of cluster-randomized controlled vaccine trials. *Am J Epidemiol* 2018; 187: 1763–1771.