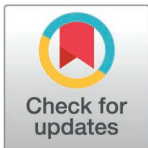


RESEARCH ARTICLE

Design of field trials for the evaluation of transmissible vaccines in animal populations

Justin K. Sheen^{1*}, Lee Kennedy-Shaffer^{2,3}, Michael Z. Levy⁴, Charlotte Jessica E. Metcalf^{1,5}

1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, United States of America, **2** Department of Mathematics and Statistics, Vassar College, Poughkeepsie, New York, United States of America, **3** Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, United States of America, **4** Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **5** School of Public and International Affairs, Princeton University, Princeton, New Jersey, United States of America

* jsheen@princeton.edu

OPEN ACCESS

Citation: Sheen JK, Kennedy-Shaffer L, Levy MZ, Metcalf CJE (2025) Design of field trials for the evaluation of transmissible vaccines in animal populations. *PLoS Comput Biol* 21(2): e1012779. <https://doi.org/10.1371/journal.pcbi.1012779>

Editor: Claudio José Struchiner, Fundação Getúlio Vargas: Fundacao Getulio Vargas, BRAZIL

Received: September 8, 2023

Accepted: January 10, 2025

Published: February 3, 2025

Copyright: © 2025 Sheen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The simulated data, code, and results that support the findings of this study are openly available in GitHub at <http://www.github.com/jsheen/netVax>.

Funding: CJEM was supported by the Princeton Catalysis Initiative. The funders had no role in study design, data collection and

Abstract

Vaccines which can transmit from vaccinated to unvaccinated animals may be especially useful for increasing immunity in hard to reach populations or in populations where achieving high coverage is logistically infeasible. However, gauging the public health utility for future use of such transmissible vaccines and assessing their risk-benefit tradeoff, given their potential for unintended evolution, hinges on accurate estimates of their indirect protective effect. Here, we establish the conditions under which a two-stage randomized field trial can characterize the protective effects of a transmissible vaccine relative to a traditional vaccine. We contrast the sample sizes required to adequately power these trials when the vaccine is weakly and strongly transmissible. We also identify how required sample sizes change based on the characteristics of host ecology such as the overdispersion of the contact structure of the population, as well as the efficacy of the vaccine and timing of vaccination. Our results indicate the range of scenarios where two-stage randomized field trial designs are feasible and appropriate to capture the protective effects of transmissible vaccines. Our estimates identify the protective benefit of using transmissible vaccines compared to traditional vaccines, and thus can be used to weigh against evolutionary risks.

Author summary

Transmissible vaccines, which are able to spread from vaccinated animals to unvaccinated animals, have the potential to vaccinate a larger proportion of animal populations than a traditional, non-transmissible vaccine. However, prior to deployment, investigators need methods that characterize these unique protective effects, such as protection via the indirect vaccination of initially unvaccinated individuals. In this work we propose a two-stage randomized trial design to evaluate transmissible vaccines and define causal estimands that capture protections conferred by a transmissible vaccine and statistical estimands that capture other features of the population impacts. Through simulation we show that these trials can identify the protective benefits of transmissible vaccines relative to traditional vaccines,

analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

and that such trials are logistically feasible for many animal populations. We also develop an analytical solution that can quickly estimate the approximate sample size required to adequately power such a trial under a variety of characteristics of the vaccine and host ecology, such as the transmissibility of the vaccine and contact structure of the host.

Introduction

Characterizing how the benefits of vaccinating one individual scales across the population is an important question in the management of pathogens that circulate in livestock and wildlife animal populations. Standard, non-transmissible vaccines directly protect vaccinated animals by reducing mortality and morbidity, and may also have secondary/indirect protective effects for the contacts of the vaccinated animal by blocking transmission chains [1-3]. Transmissible vaccines, where the vaccines may self-replicate and transmit to unvaccinated animals from a vaccinated animal within their social cluster, should provide even greater population-level protection per animal vaccinated than standard vaccines. Transmissible vaccines include live-attenuated viruses, which are genetically weakened forms of wild-type pathogens, and recombinant vector vaccines (RVV), which are harmless viruses with genetic insertions of parts of the wild-type pathogens [4]. While the potential level of protection that these vaccines might induce in a population has been the subject of much investigation [4-6], less work has been done on how to quantitatively assess these predictions in field conditions. Importantly, transmissibility of such vaccines is no guarantee of high indirect protection from disease; for example, if the population has a sparse contact structure, then the vaccine may not be able to transmit effectively in the population, although the wildtype pathogen may still circulate. Understanding whether or not transmissible vaccines deliver the benefits they promise has relevance for optimal resource use, but also in weighing evolutionary concerns. Although strongly transmissible vaccines will be more effective at reaching animals, more transmission also provides more opportunities for mutation and selection for problematic vaccine variants, e.g., pathogenic variants. Reliable estimates of the exact benefits of transmissible vaccines are required to weigh against these risks [6-10].

Over the past decade, the potential utility of these self-disseminating vaccines has led to their proposed deployment for a wide spectrum of pathogens that circulate in both livestock and wildlife animal populations [11-13]. Transmissible vaccines that target poultry viruses such as avian influenza may be especially useful given the potential for zoonotic spillover [11]. Development of transmissible vaccines against rabies has also been proposed for hard-to-vaccinate vampire bat populations [12]. The value of transmissible vaccines will be defined by the degree of indirect protection that they confer, in addition to the direct protection of vaccinated animals. Even weakly transmissible vaccines for animal populations—defined here as vaccines with a reproductive value $R_{0,v} < 1$, meaning that a vaccinated individual on average transmits the vaccine to less than one other individual—have the potential to prevent epidemics in contexts where relatively few individuals are vaccinated [5,14,15]. This characteristic makes transmissible vaccines an especially compelling option for blocking transmission and achieving herd immunity thresholds in populations likely to have low vaccination coverage; for example poultry pathogens in regions of sub-Saharan Africa with historically low vaccination coverage [16-18].

Field trials are crucial in the transmissible vaccine development pipeline, since they allow investigators to capture protections in animal populations at scale [13]. However, there is limited literature on randomized controlled trials (RCTs) [19], considered the gold standard in testing interventions [20], in evaluating transmissible vaccines. A methodology to capture the unique indirect protection of transmissible vaccines has not yet been proposed. To address this gap, we propose a two-stage randomized trial design for field trials of transmissible

vaccines in animal populations, since in addition to the overall protective effect, it allows investigators to estimate with precision the significant indirect protection of these vaccines [1-3,21-25]. The estimation of these various protections of transmissible vaccines inform the debate of their more widespread use in animal populations.

While quantifying the protections provided by transmissible vaccines is of clear policy importance, the feasibility of such designs, specifically sample sizes required to adequately power such trials, remains an open question. Our article considers two plausible trial designs: trials that compare protection conferred by transmissible vaccines relative to traditional (i.e., non-transmissible) vaccines, as well as trials that estimate the overall protection of transmissible vaccines compared to pure control groups that do not receive any vaccination. Which of these designs is used in a given setting will depend on the available comparator(s) and the current standard, as well as economic and feasibility considerations, but can be informed by the statistical properties discussed here.

We make several methodological contributions towards the successful evaluation of transmissible vaccines using field trials. First, we extend the standard two-stage randomized trial design framework to allow for causal comparisons of two vaccine types: transmissible and non-transmissible vaccines. Second, we identify and characterize two novel statistical estimands that provide epidemiological insight into reductions in risk of infection for subpopulations unique to transmissible vaccines. Third, we use both simulation and analytic approximate sample size formulae to calculate and compare the required sample sizes needed to adequately power trials depending on vaccine characteristics (e.g., strength of transmissibility), trial characteristics (e.g., vaccination timing), and host ecology factors (e.g., contact structure of the animal host population). We show that, under many conditions, reasonable sample sizes for a single post-intervention sample can adequately estimate various protective effects. We also make these methods available for investigators to determine the required sample size for their setting. Finally, from this foundation, we discuss opportunities for further investigation into effectively evaluating the protections of transmissible vaccines in animal populations.

Methods

The next three subsections describe the two-staged randomized trial design, estimands of interest, and estimators used, to evaluate the effects of a given transmissible vaccine. For sample size calculations, the following four subsections describe simulation and analytic sample size methods, grounded in an existing system. The simulation model is structured around transmission and contact structure parameters appropriate to the Newcastle disease virus transmission system in poultry, where the vaccine is a live attenuated form of the virus that is incidentally transmissible. The populations we study have no prior immunity, and the mean number of contacts for each animal is 15. The same methods for sample size estimation can be used for other pathogens or settings with appropriate changes to the parameters and model assumptions. In the Discussion, we consider how required sample sizes and trial designs may change across other study systems and pathogen and vaccine life-cycles. The trial design we propose is appropriate for field trials in the development pipeline of transmissible vaccines [13], which are proposed for geographically isolated areas such as islands or regions isolated by mountains to minimize the risk of vaccine transmission outside the study area.

Trial design

We study trials with a two-stage randomized design, where in the first stage of the trial N_T clusters are randomized to either vaccinate $\alpha\% = 5\%$ of their cluster with a transmissible vaccine or a

traditional vaccine, and in the second stage, animals are randomly assigned to be vaccinated with the vaccine assigned to the cluster in the first stage. An example schematic of the trial design is shown in Fig 1 [1]. We assume that vaccination coverage, $\alpha\%$, does not vary between treatment assignment strategies so that the estimands capture relevant quantities (see Section B.2.1 in S1 Text). The total number of recruited clusters, $N_T = N_0 + N_1$ where N_0 and N_1 are the number of clusters in each arm; we set $N_0 = N_1$. We vary the transmissibility of the vaccine to be either weak ($R_{0,v} = 0.9$) or strong ($R_{0,v} = 1.1$). In the main results, for each $R_{0,v}$, we model anticipatory trial designs, where the transmissible vaccine is allowed to circulate in the population before an outbreak of the wildtype pathogen. We also consider reactionary trial designs, where investigators react to an epidemic starting in a cluster and vaccinate susceptible animals when, on average, 0.5% of the population is actively infected (Fig 2) [5]. Additionally, we consider trials that compare protection from a transmissible vaccine to pure control groups that do not receive any vaccination.

150 days after first infection, 10% of animals in each cluster are randomly sampled and the cumulative number of infections in the sample is recorded. We choose a sampling fraction of 10% as a reasonable lower bound, since previous vaccination campaigns for poultry successfully sampled up to 50% of animals from farms at follow-up [26], and high sampling coverage for wildlife populations may be more difficult to achieve [12]. The ability to distinguish wildtype infection from vaccination is crucial, since statistical inference hinges on identifying decreases in wildtype infection burden. If infections are mild or subclinical, tools such as serological assays can be used to differentiate infected animals from vaccinated animals (DIVA) [27,28]. While the trial is not designed specifically to estimate the rate of occurrences of undesirable evolutionary outcomes, such as reversion to a pathogenic form, checks should be done during the trial itself to see whether these undesirable events have occurred. If animals that were solely seropositive for the vaccine, and not the wildtype pathogen, die at rates faster than we would expect naturally, this may be a sign that the vaccine has reverted to a pathogenic form.

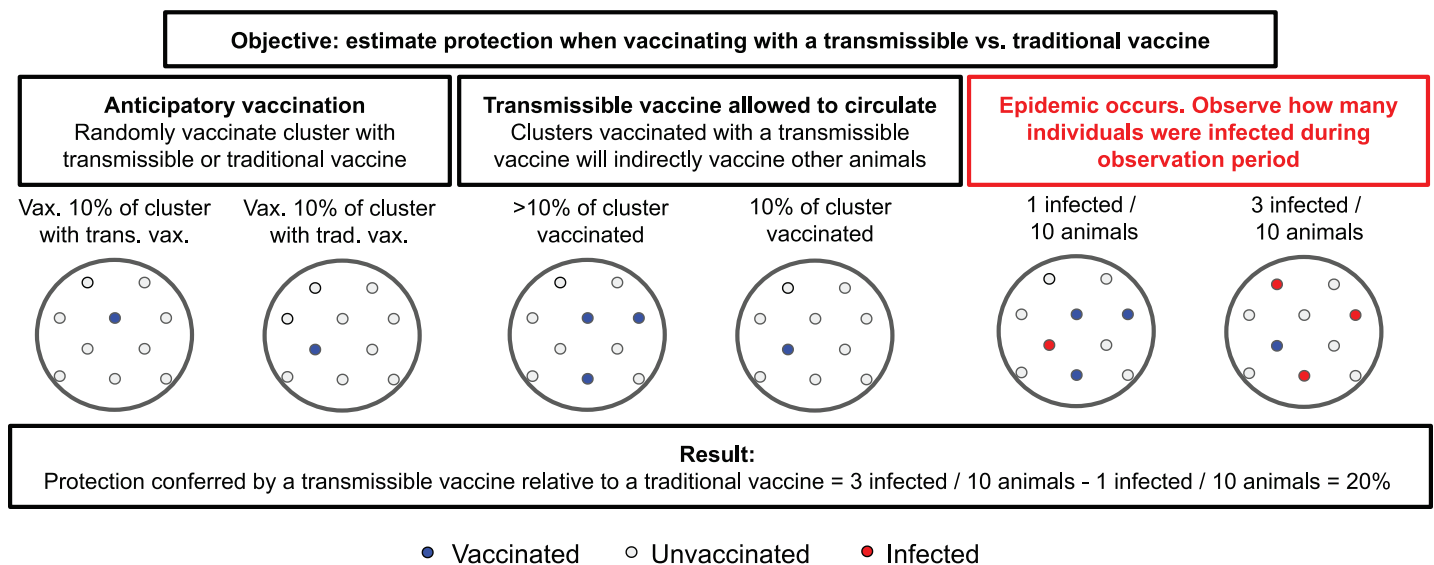


Fig 1. Simple example calculation of overall protection conferred by a transmissible vs. a traditional vaccine. First, clusters are vaccinated with either a transmissible or traditional vaccine. In the second stage, the transmissible is allowed to circulate for all animals. Finally, an epidemic occurs, and the number of infected animals of the cluster, among both initially vaccinated and initially unvaccinated animals, is counted. We assume there is no interference between clusters, i.e., The condition of vaccination or infection of an animal of one cluster affects either the condition of vaccination or probability of infection in another cluster. In this simple example 10% of each cluster is initially vaccinated for illustrative purposes, whereas in our Results 5% of each cluster is initially vaccinated.

<https://doi.org/10.1371/journal.pcbi.1012779.g001>

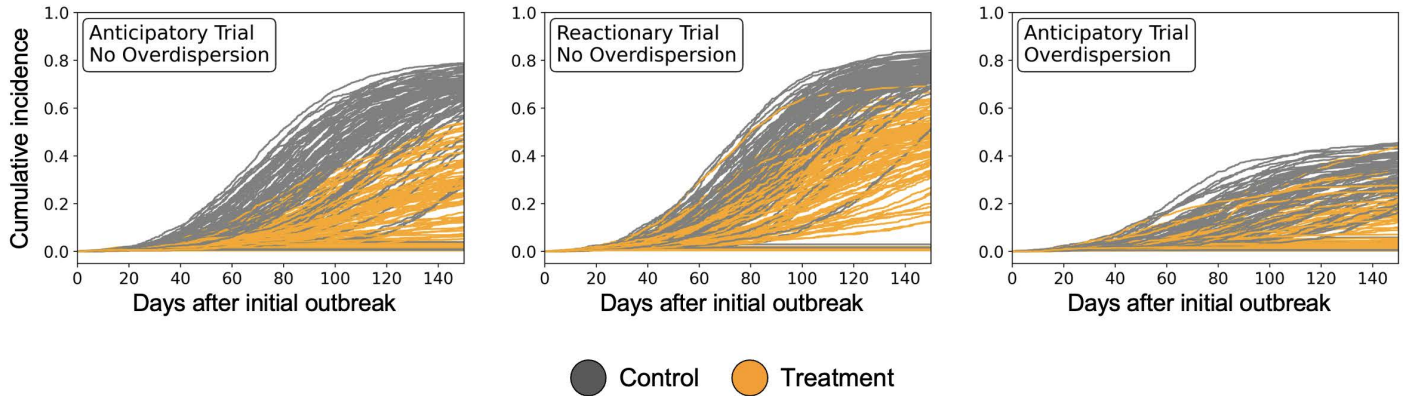


Fig 2. Differences in cumulative infection burden depend on whether the trial is anticipatory or reactionary, and if the contact structure is overdispersed. Left panel: anticipatory trial with no overdispersion in contact structure. Middle panel: reactionary trial, where vaccination starts once 0.5% of the population has been infected, with no overdispersion in contact structure. Right panel: anticipatory trial with overdispersion in contact structure ($k=1$). Observation period is 150 days. Parameters set at: $R_{0,v} = 1.1$, $R_{0,w} = 2$, proportion initially recovered = 0, vaccine efficacy = 80%, initial vaccinated proportion, $\alpha = 5\%$.

<https://doi.org/10.1371/journal.pcbi.1012779.g002>

Estimation of causal protection

In the trials we consider there are several vaccine exposures: animals can either be initially vaccinated (also referred to as directly vaccinated) with either a traditional or transmissible vaccine, or initially unvaccinated; if the cluster has been treated with a transmissible vaccine, initially unvaccinated animals are either indirectly vaccinated through transmission of the vaccine or never vaccinated. To estimate the protections of a transmissible vaccine, we first introduce the estimand, $\Delta_1(\alpha)$ which defines the overall causal effect of protection that transmissible vaccines confer when $\alpha\%$ of the population is initially vaccinated [1]. We define this estimand using the potential outcomes causal inference framework:

$$\Delta_1(\alpha) = \bar{Y}(\psi(\alpha)) - \bar{Y}(f(\alpha))$$

where the estimand $\Delta_1(\alpha)$ is an average of the difference in cluster/group potential outcomes of cluster i , $\bar{Y}_i(\psi(\alpha)) - \bar{Y}_i(f(\alpha))$ which are themselves differences in the averages of the individual potential outcomes for individual j of cluster i (further details in Section B.1.3 in S1 Text).

$\bar{Y}(\cdot)$ is the population average potential outcome of infection under a given individual treatment assignment strategy.

$\psi(\alpha)$ is a treatment assignment strategy where $\alpha\%$ of the population is initially vaccinated with a traditional vaccine.

$\Phi(\alpha)$ is a treatment assignment strategy where $\alpha\%$ of the population is initially vaccinated with a transmissible vaccine.

The estimator is:

$$\widehat{\Delta_1(\alpha)} = \frac{1}{N_0} \sum_{u=1}^{N_0} Z_{\Delta_1, \psi(\alpha), i=g(u)} / S_{\Delta_1, \psi(\alpha), i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} Z_{\Delta_1, \Phi(\alpha), i=h(u)} / S_{\Delta_1, \Phi(\alpha), i=h(u)}$$

where $S_{\Delta_1, \psi(\alpha), i=g(u)}$ and $S_{\Delta_1, \Phi(\alpha), i=h(u)}$ are the number of sampled animals from cluster i in the traditional arm ($\psi(\alpha)$) and transmissible arm ($\Phi(\alpha)$) of the trial, respectively.

$Z_{\Delta_1, \psi(\alpha), i=g(u)}$ and $Z_{\Delta_1, \Phi(\alpha), i=h(u)}$ are the number of sampled animals that were infected in the traditional arm ($\psi(\alpha)$) and transmissible arm ($\Phi(\alpha)$) of the trial, respectively.

$g(u)$ and $h(u)$ map the cluster index within each arm (from $u = 1$ to $u = N_0$ or N_1 for the traditional vaccine or transmissible vaccine arm, respectively) to the general cluster index, i .

For all causal estimands we assume that treatment assignment (i.e., initial vaccination) both on the cluster and individual levels is randomized, and that there is no interference between clusters, which is plausible for geographically isolated regions. We also assume perfect compliance, since animals would be directly vaccinated by investigators. Additionally, for all causal estimands, proper randomization of the treatment assignment, both on the cluster and individual levels, eliminates potential bias due to unmeasured confounders that have a relationship with both treatment assignment and outcome. Full details of all estimands, both causal and statistical (defined below), and estimators, including proofs of the unbiasedness, consistency, and asymptotic normality of the estimators, are provided in Appendix B in [S1 Text](#).

For all estimators, we choose to estimate cluster average effects since we are interested in protections and risk differences that will depend on the cluster-level contact structure of a given cluster [29]. If cluster size, n , varies, the interpretation of our estimands should remain cluster-level averages to maintain interpretability, consistency, and validity.

We introduce two additional causal estimands: $\Delta_2(\alpha)$, which defines the causal protection that transmissible vaccines confer to initially vaccinated individuals relative to initially vaccinated individuals using a traditional vaccine; and $\Delta_3(\alpha)$, which defines the causal protection the transmissible vaccines confer to initially unvaccinated individuals relative to initially unvaccinated individuals using a traditional vaccine. Specifically, the estimands are:

$$\Delta_2(\alpha) = \bar{Y}(z = 1; \psi(\alpha)) - \bar{Y}(z = 2; \Phi(\alpha))$$

$$\Delta_3(\alpha) = \bar{Y}(z = 0; \psi(\alpha)) - \bar{Y}(z = 0; \Phi(\alpha))$$

where the estimand $\Delta_2(\alpha)$ is an average of the difference in cluster/group potential outcomes of cluster i , $\bar{Y}_i(z = 1; \psi(\alpha)) - \bar{Y}_i(z = 2; \Phi(\alpha))$ which are themselves differences in the averages of the individual potential outcomes for individual j of cluster i , $\bar{Y}_{ij}(z = 1; \psi(\alpha)) - \bar{Y}_{ij}(z = 2; \Phi(\alpha))$

The estimand $\Delta_3(\alpha)$ is an average of the difference in cluster/group potential outcomes of cluster i , $\bar{Y}_i(z = 0; \psi(\alpha)) - \bar{Y}_i(z = 0; \Phi(\alpha))$ which are themselves differences in the averages of the individual potential outcomes for individual j of cluster i (further details in Section B.1.3 in [S1 Text](#)).

$\bar{Y}(\cdot)$ is the population average potential outcome of infection under a given treatment assignment strategy for an individual of treatment assignment z .

$z = 1$ indicates an individual initially vaccinated with a traditional vaccine.

$z = 2$ indicates an individual initially vaccinated with a transmissible vaccine.

$z = 0$ indicates an individual initially unvaccinated.

$\psi(\alpha)$ and $\Phi(\alpha)$ are the same as for $\Delta_1(\alpha)$

Similar to the above overall causal effect, the population average potential outcome is an average of the group potential outcome which is itself an average of the individual potential outcome, defined as the infection outcome of individual j of initial vaccine exposure z of group i under the given treatment strategy. Note that because we do not model decreased infectiousness of vaccinated individuals, where vaccinated individuals infected with the wildtype pathogen have decreased subsequent transmission of the wildtype pathogen, the estimated indirect spillover effect $\Delta_3(\alpha)$ is due to the protective effect of contagion, where the

prevention of a vaccinated individual from being infected with the wildtype pathogen prevents subsequent infection to their neighbors [30]. The estimators are:

$$\widehat{\Delta_2(\alpha)} = \frac{1}{N_0} \sum_{u=1}^{N_0} Z_{\Delta_2, \psi(\alpha), i=g(u)} / S_{\Delta_2, \psi(\alpha), i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} Z_{\Delta_2, \Phi(\alpha), i=h(u)} / S_{\Delta_2, \Phi(\alpha), i=h(u)}$$

$$\widehat{\Delta_3(\alpha)} = \frac{1}{N_0} \sum_{u=1}^{N_0} Z_{\Delta_3, \psi(\alpha), i=g(u)} / S_{\Delta_3, \psi(\alpha), i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} Z_{\Delta_3, \Phi(\alpha), i=h(u)} / S_{\Delta_3, \Phi(\alpha), i=h(u)}$$

where, for $\widehat{\Delta_2(\alpha)}$ $S_{\Delta_2, \psi(\alpha), i=g(u)}$ and $S_{\Delta_2, \Phi(\alpha), i=h(u)}$ are the number of sampled initially vaccinated animals in cluster i that are in the traditional arm ($\psi(\alpha)$) and transmissible arm ($\Phi(\alpha)$) of the trial, respectively.

$Z_{\Delta_2, \psi(\alpha), i=g(u)}$ and $Z_{\Delta_2, \Phi(\alpha), i=h(u)}$ are the number of sampled initially vaccinated animals that were infected in cluster i in the traditional arm ($\psi(\alpha)$) and transmissible arm ($\Phi(\alpha)$) of the trial, respectively.

Z and S expressions for $\widehat{\Delta_3(\alpha)}$ are the same as for $\widehat{\Delta_2(\alpha)}$ except that they measure the proportion infected among initially unvaccinated animals (as specified by their subscript, Δ_3).

Estimation of differences in subpopulation risks of infection

Besides causal estimands, we introduce two statistical estimands to provide differences in the risk of infection for indirectly vaccinated animals—i.e., those animals who were not initially vaccinated and were subsequently vaccinated by the transmissible vaccine—and never vaccinated animals—i.e., those animals who were not initially vaccinated and did not get vaccinated by the transmissible vaccine—in clusters where the transmissible vaccine was implemented. We compare the risk of infections among these populations to the initially unvaccinated population among clusters where the transmissible vaccine was not implemented. We use θ to denote statistical estimands to differentiate them from Δ , which refers solely to causal estimands defined using potential outcomes (see Section B.3 in [S1 Text](#)). Similarly, we do not use the overline notation when defining these estimands since they are not potential outcomes.

Trials should aim to identify the causal effects of transmissible vaccines defined in “Estimation of causal protection.” However, if there are sufficient resources, these estimands provide further epidemiological information of the differences in risk of infection for subpopulations unique to transmissible vaccines. We caution that these estimands may not be directly interpretable as protection solely due to the individual-level status of indirect vaccination or never vaccination since there may be population-specific characteristics that can bias this interpretation. Since they use non-randomized groups in the transmissible vaccine arm, these estimates may be biased from the causal protective effect due to confounding of the transmission of vaccine and transmission of the pathogen. Rather than causal vaccine effects, these estimands represent true differences in the cluster-average risk of infection for subpopulations unique to the utilization of transmissible vaccines.

$\theta_1(\alpha)$ defines the difference in risk of infection between indirectly vaccinated animals in clusters that receive the transmissible vaccine, relative to unvaccinated animals in clusters that receive the traditional vaccine. We define the estimand as:

$$\theta_1(\alpha) = \frac{1}{N_0} \sum_{u=1}^{N_0} T_{\theta_1, \psi(\alpha), i=g(u)} / K_{\theta_1, \psi(\alpha), i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} T_{\theta_1, \Phi(\alpha), i=h(u)} / K_{\theta_1, \Phi(\alpha), i=h(u)}$$

where $K_{\psi(\alpha),i=g(u)}$ is the total number of initially unvaccinated animals in cluster i that are in the traditional vaccine arm ($\psi(\alpha)$) where $\alpha\%$ of the cluster is initially vaccinated with a traditional vaccine and $T_{\psi(\alpha),i=g(u)}$ is the number of those animals that were infected.

$K_{\Phi(\alpha),i=h(u)}$ is the total number of indirectly vaccinated animals in the transmissible vaccine arm ($\Phi(\alpha)$) where $\alpha\%$ of the cluster is initially vaccinated with a transmissible vaccine and $T_{\Phi(\alpha),i=h(u)}$ is the number of those animals that were infected.

The functions $g(u)$ and $h(u)$ map the cluster index within each arm (from $u = 1$ to $u = N_0$ or N_1 for the traditional vaccine or transmissible vaccine arm, respectively) to the general cluster index, i . In other words, we take the difference in average cluster-level risks of infections among the two arms of the trial. The estimator is:

$$\widehat{\theta}_1(\alpha) = \frac{1}{N_0} \sum_{u=1}^{N_0} Z_{\psi(\alpha),i=g(u)} / S_{\psi(\alpha),i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} Z_{\Phi(\alpha),i=h(u)} / S_{\Phi(\alpha),i=h(u)}$$

where $S_{\psi(\alpha),i=g(u)}$ is the number of sampled initially unvaccinated animals in cluster i that are in the traditional arm ($\psi(\alpha)$) and $Z_{\psi(\alpha),i=g(u)}$ are the number of those sampled animals that were infected.

$S_{\Phi(\alpha),i=h(u)}$ is the number of sampled indirectly vaccinated animals in the transmissible arm ($\Phi(\alpha)$) and $Z_{\Phi(\alpha),i=h(u)}$ are the number of those sampled animals that were infected.

$\theta_2(\alpha)$ defines the difference in risk of infection between never vaccinated animals of clusters treated with the transmissible vaccine vs. initially unvaccinated animals of clusters treated with the traditional vaccine. We define the estimand as:

$$\theta_2(\alpha) = \frac{1}{N_0} \sum_{u=1}^{N_0} T_{\psi(\alpha),i=g(u)} / K_{\psi(\alpha),i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} T_{\Phi(\alpha),i=h(u)} / K_{\Phi(\alpha),i=h(u)}$$

where $K_{\psi(\alpha),i=g(u)}$ is the total number of initially unvaccinated animals in cluster i of the traditional arm ($\psi(\alpha)$) where $\alpha\%$ of the cluster is initially vaccinated with a traditional vaccine and $T_{\psi(\alpha),i=g(u)}$ is the number of those animals that were infected.

$K_{\Phi(\alpha),i=h(u)}$ is the total number of never vaccinated animals (neither directly nor indirectly vaccinated) in cluster i of the transmissible arm ($\Phi(\alpha)$) where $\alpha\%$ of the cluster is initially vaccinated with a transmissible vaccine and $T_{\Phi(\alpha),i=h(u)}$ is the number of those animals that were infected.

The functions $g(u)$ and $h(u)$ map the cluster index within each arm (from $u = 1$ to $u = N_0$ or N_1 for the traditional vaccine or transmissible vaccine arm, respectively) to the general cluster index i . The estimator is:

$$\widehat{\theta}_2(\alpha) = \frac{1}{N_0} \sum_{u=1}^{N_0} Z_{\psi(\alpha),i=g(u)} / S_{\psi(\alpha),i=g(u)} - \frac{1}{N_1} \sum_{u=1}^{N_1} Z_{\Phi(\alpha),i=h(u)} / S_{\Phi(\alpha),i=h(u)}$$

where $S_{\psi(\alpha),i=g(u)}$ and are the number of sampled initially unvaccinated animals in cluster i that are in the traditional arm ($\psi(\alpha)$) and $Z_{\psi(\alpha),i=g(u)}$ are the number of those sampled animals that were infected.

$S_{\Phi(\alpha),i=h(u)}$ is the number of sampled never vaccinated animals in the transmissible arm ($\Phi(\alpha)$) and $Z_{\Phi(\alpha),i=h(u)}$ are the number of those sampled animals that were infected.

Because of the complexity of variance estimation for clusters with different contact structures, all hypothesis testing is conducted using randomization-based inference; variance estimation and confidence intervals are calculated using resampling-based inference.

Simulation model: cluster assumptions

We study trials that are conducted for clusters of moderately large size; the number of animals in each simulated cluster is size $n = 1000$. We define a cluster as a contiguous region where animals are in possible contact with one another, such as farms with free-roaming poultry. We assume that animals within each cluster can potentially interact through a network contact structure, but not between clusters [31,32]; thus, there is interference within clusters but not between clusters. Although all animals are expected to have a mean number of contacts, we assume the exact number is drawn from a distribution, and thus there is heterogeneity in the number of contacts per animal. Animals that were once infected or vaccinated cannot become infected or vaccinated again. Picking cluster characteristics that will reveal the protection and risk differences of transmissible vaccines is important; for example, in our case, picking larger $n = 1000$ animals ensures a higher probability of imported infection via exposure to external sources of infections [33,34]. We assume no dependency between cluster contact structure and importation rate, but note that investigators should be aware that this may potentially affect the importation rate of infections: if animals have a high number of contacts within clusters these animals may also have higher contact outside of the cluster, increasing infection importation. In practice, if n varies across the trial there may be informative cluster sizes where treatment effects differ across clusters [29]; our estimators would average over these treatment effects across clusters since they estimate cluster-average effects; interpretation of the causal effects would need to change accordingly. Within each cluster we assume that there is no prior immunity to wildtype infection and vaccination, as has been assumed in previous transmissible vaccine models [5,12].

We model the contact structure of the clusters using a network, where animals are nodes of the network, and edges represent contact between animals. The degree of each animal is either Poisson distributed (mean degree = 15) or moderately overdispersed (mean degree = 15, overdispersion parameter $k = 1$) with no nodes of degree 0 (Fig 2) [35,36]. We choose a mean degree of 15 to approximately match the average number of poultry per farm which experience outbreaks of poultry pathogens such as Newcastle disease virus and avian influenza [27,37]. While a Poisson-distributed degree distribution may be appropriate to model certain ecological contexts, we additionally model a moderately overdispersed degree distribution (contact structure) due to the overdispersion that has been observed in the connectivity of hosts on a regional scale, e.g., within a trading network where Newcastle disease virus circulates [38]. We assume that both the wildtype pathogen and vaccine transmit on this same contact structure of the population.

For overdispersed contact structures, in order to create each cluster according to the desired degree distribution, we use a configuration model (CM) algorithm that does not allow singletons in the cluster, nor self-loops and multi-edges between nodes [39]. Because of our modifications, the effective degree distribution of the cluster nearly, but does not exactly, match the desired degree distribution. The number of connected components in each cluster is low; we have observed at most three connected components in a single cluster. Thus, we account for the possibility of disjoint groupings of animals within a single cluster, where there is no interference, which may occur in practice.

Simulation model: wildtype and vaccine transmission assumptions

We model transmission through an animal population with a continuous-time, stochastic SEIR model (Fig A in S1 Text) [40]. Although our aim is to provide results that are

applicable across a broad spectrum of livestock and wildlife pathogens, we initially model an acute infectious pathogen similar to poultry pathogens that infect chickens, and discuss generalization of these results below. In the model, an animal's mean incubation period and mean infectious period are 5 days and 10 days respectively, both for the wildtype pathogen and vaccine, mirroring the average incubation and infectious period of Newcastle disease virus [41,42]. The wildtype infection and vaccine infection differ solely in their rate of transmission. The incubation period, infectious period, and time between onward transmission events are exponentially distributed [43]. The memoryless property of the exponential distribution allows us to interrupt the epidemic simulation to implement vaccination in reactionary trials without interfering with the epidemic progression. We assume $R_{0,w}$, or the R_0 of the wildtype strain, is 2 in order to model a highly transmissible wildtype strain, such as Newcastle disease virus [38,44]. Parameter values and references are detailed in Table A in [S1 Text](#).

We assume that the vaccine confers either perfect immunity or “leaky” immunity, where the vaccine efficacy for vaccinated animals against wildtype infection is 80% (i.e., their risk of infection on each contact is reduced by 80%). We assume vaccinated animals cannot subsequently become reinfected by the vaccine. Additionally, animals infected with the wildtype cannot become reinfected by either the wildtype or vaccine. We do not model vaccine failure, where some proportion of vaccinated animals do not have any protection against infection (“all-or-nothing” vaccine effects). Similarly, we do not model vaccine transmission failure, where the vaccine systematically fails to transmit for some vaccinated animals, nor a decreased infectiousness of vaccinated individuals [30]. We assume both immunity from wildtype and vaccine infection last for the entire trial duration. In our simulations, there is also no superinfection, meaning that no animal can simultaneously be infected by both the vaccine and wildtype strain.

We seed the epidemic of each cluster with four initial infections. Both when vaccinating 5% of susceptible animals at intervention and when seeding the epidemics with wildtype infection, we select animals randomly to vaccinate and infect, respectively. We use the *EoN* python package to simulate the epidemic [45]. Code used for simulations is provided at <http://www.github.com/jsheen/netVax>.

Required sample sizes, N_T^* through simulation

For each estimand in a given scenario we report the required sample size, N_T^* , which we define as the minimum number of total clusters across both arms of the trial needed to achieve at least 80% power to detect a statistically significantly lower infection burden (p -value ≤ 0.05) in the transmissible vaccine treatment arm compared to the second arm of the trial, which is either a traditional vaccine treatment or pure control with no vaccination (one-sided test). The estimands across scenarios are for the overall causal effect $\Delta_1(\alpha)$ and spillover causal effect $\Delta_3(\alpha)$, as well as the differences in risk for indirectly vaccinated animals and never vaccinated animals, which we refer to as the statistical estimands (see “Estimation of causal protection” and “Estimation of differences in subpopulation risks of infection”). The total vaccine causal effect on the vaccinated $\Delta_2(\alpha)$ could be measured here, but is more relevant when vaccine efficacy varies between arms of the trial; we leave that exploration to future work. For a given number of clusters, the power of the trial is the proportion of 1000 simulated trials with p -value ≤ 0.05 under the alternative hypothesis of a true effect. To obtain the p -value of each simulated trial for hypothesis testing, we conduct a randomization (permutation) test. We first create a null distribution by performing 1000 random permutations of the original labels of clusters to each of the two arms of the trials, in each permutation recalculating the targeted estimate of protection or difference in risk of

infection. When there are 12 or fewer clusters, all possible permutations are used instead of randomly choosing 1000. The p-value is:

$$\text{p-value} = 1 - p_n$$

where p_n is the proportion of the estimates in this null distribution that are less than or equal to the original estimate.

To identify N_T^* , we start with eight clusters as the lower bound on the number of clusters across both arms of the trial and increase by two clusters, one for each arm of the trial, until 80% power is achieved. A minimum of eight clusters is used since this is the minimum number of clusters needed for the permutation test to be able to yield a p-value under 0.05. Note that because of the coarse scale of the evaluation of the design (e.g., adding relatively large clusters), the power obtained can be noticeably above 80%. To ensure statistical validity, we also provide the Type I error rate of N_T^* , i.e., the false positive rate, by calculating the power of the trial if the vaccine has no effect across 3000 trial simulations.

For each N_T^* for each estimand of a given scenario, we also provide the mean estimate of the protective effect, or risk difference, and mean width of the bootstrapped 95% confidence interval for each estimate. We choose a bootstrapped confidence interval to estimate variance since existing analytical variance estimators may not be appropriate in our study setting, for example due to the assumption of stratified interference [1]. To obtain the 95% confidence intervals of each estimate, we create 1000 bootstrap samples and find the 2.5 and 97.5 percentiles of the samples. To create each bootstrap sample, we first draw with replacement the clusters of each arm of the trial in the original sample [46,47]. Next, we draw with replacement the same number of sampled animals, s , from each resampled cluster, which may be a conservative estimate of the variance [48-49]. Alternatively, investigators may also include all originally sampled animals from each resampled cluster, among other resampling approaches [46].

Clusters where less than three sampled animals (i.e., less than 3% of the sample) were observed to have an infection are excluded in our analysis. For cluster randomized trials and infectious disease outbreak studies more generally, bias or reduced efficiency can arise from including communities with few or no events [50,51]. Observing at least three infected animals in our sample indicates transmission in the population, ensuring exposure to the wild-type pathogen, and improving the efficiency of estimation and power of the hypothesis test, especially for overdispersed populations. If the vaccine has no effect, the exclusion should not introduce any bias. When the vaccine is effective enough to cause less than 3% of the population to be infected on average, the exclusion may create a downward bias by excluding clusters where the vaccination successfully prevented almost all transmission. Thus, the exclusion can potentially raise N_T^* in those settings, since treated clusters with little to no infection will not be used for analysis. However, because infection rates for highly transmissible pathogens such as Newcastle disease virus are high [17], and because we model settings where a moderately small portion of the population is vaccinated (5%), this scenario is unlikely, and we specifically study these scenarios to avoid bias. With a larger initially vaccinated population or even more transmissible vaccine, where the vaccinated proportion may reach larger proportions of the population, the bias and efficiency loss may be more pronounced since there may be more clusters with little to no infection (noting that this will yield a conservative estimate of N_T^*). Our reported sample sizes account for the potential exclusion: for a given N_T^* across all trial simulations at least 80% are statistically significant, even if for any one trial simulation clusters may be excluded before analysis. When calculating the risk difference for indirectly vaccinated animals ($\theta_1(\alpha)$), we do not use this constraint, since the effect necessarily relies on vaccinated animals who have little to no risk of infection.

Approximate formula for the required sample size, N_T^*

In addition to the simulation study results, which are based on the power of a permutation test to detect a significant risk difference of infection between arms of the trial, we also provide an analytical solution for the approximate required sample size, N_T^* , to adequately power trials that use a Welch's t-test on the risk differences of infection between a transmissible vaccine and traditional vaccine. When the traditional vaccine confers no protection, this comparison is equivalent to comparing the protection or infection risk of transmissible vaccines to control groups that do not receive any vaccine. This approximate sample size formula demonstrates how N_T^* depends on both ecological and study design factors such as the R_0 of the vaccine, R_0 of the wildtype pathogen, the cluster size, n , number of animals sampled from each cluster, s , the initial recovered population, initial vaccinated population, vaccine efficacy, and the between-cluster variance in final size proportions. Although the approximate formula is useful for understanding sample size relationships, in practice, because the variances of the estimators we calculate for the approximate formula are lower bounds based on distributional and epidemic assumptions (see Appendix C in [S1 Text](#)) [52], the bootstrapped confidence intervals defined in the simulation study methods above are a more appropriate estimate of variance. All methods to derive the formula are provided in [S2 Appendix C in S1 Text](#). An R Shiny app that shows the changes in N_T^* according to this formula is available at <https://j-k-s.shinyapps.io/shinyTrans/>.

Results

The required sample sizes across both arms of the trial, N_T^* , to achieve 80% power for anticipatory trials in non-overdispersed settings when comparing the overall and spillover causal effects of a transmissible vaccine with $R_{0,v} = 0.9$ or 1.1 vs. a non-transmissible traditional vaccine when the vaccine efficacy is either 80% or 100% are broadly logistically feasible ([Table 1](#)) [26,53,54]. If there are sufficient resources, the risk differences for subpopulations unique to transmissible vaccines are also logistically feasible to test ([Table 1](#)). The minimum number of total animals that need to be sampled is 800 animals from 8 clusters, and the maximum number of total animals that would need to be sampled is 1800 animals from 18 clusters when the vaccine has 80% efficacy and is weakly transmissible ([Table 1](#)). The Type I error across all scenarios is $\leq 5\%$.

N_T^* depends on the estimand of interest. N_T^* for the overall and spillover causal effects were found to be the same for anticipatory trials in non-overdispersed settings ([Table 1](#)). Additional estimation of the difference in risk of infection for indirectly vaccinated animals ($\theta_1(5)$) requires the smallest sample size (approximately 800 animals from 8 clusters), while estimating the difference in risk of infection for never vaccinated animals ($\theta_2(5)$) requires the largest (approximately 10–18 clusters sampling 1000–1800 animals total).

Mean effect estimates increase as the vaccine transmissibility and efficacy increase. For example, the overall causal effect increases from 25% to 41% with increased transmissibility from 0.9 to 1.1 when the vaccine efficacy is leaky and increases from 32% to 49% when the vaccine is perfectly effective instead of 80% effective. These increases in vaccine transmissibility and vaccine efficacy can reduce N_T^* , which may be crucial in resource-limited areas. For most scenarios the mean 95% bootstrapped confidence interval width was less than or equal to the mean effect estimate.

When the contact structure is overdispersed ([Table 2](#)), or when trial designs are reactionary ([Table 3](#)), N_T^* for the causal effects, as well as estimates of risk differences for subpopulations unique to transmissible vaccines, can increase due to more clusters with little to no infections in the former, and increased competition of the vaccine with the wildtype pathogen for susceptible animals (and, thus, less opportunity for indirect protection) in the latter. For overdispersed populations, even for strongly transmissible vaccines, N_T^* can increase from 10 clusters

Table 1. Estimated protection or risk difference from and required sample sizes, N_T^* , for trials comparing a transmissible vaccine to a traditional vaccine. Prior to the outbreak, $\alpha=5\%$ of each cluster is vaccinated with either a transmissible or traditional, non-transmissible vaccine. The contact structure is Poisson distributed and trials are anticipatory. The estimand of interest is noted in the first column: $\Delta_1(5)$ refers to the overall protection, $\Delta_3(5)$ refers to the spillover protection, $\theta_1(5)$ refers to the difference in risk of infection for indirectly vaccinated animals in the transmissible case compared to those unvaccinated in the traditional case, and $\theta_2(5)$ refers to the difference in risk of infection for never vaccinated animals in the transmissible case. Vax. Eff. is the true direct efficacy of the vaccine to decrease susceptibility of vaccinated animals. $R_{0,v}$ refers to the R_0 of the vaccine. Estimate protection or risk difference is the mean estimate of the noted estimand comparing transmissible vaccines to traditional vaccines. $\langle . \rangle$ indicates the mean 95% confidence interval width. Req. sample size from simulation, N_T^* , is the number of clusters and number of animals required to estimate the effect in order to achieve at least 80% power. Power is the percentage of simulations with a p-value $\leq 5\%$ across 1000 trial simulations. Type I error is the percentage of simulations with a p-value $\leq 5\%$ across 3000 trial simulations for this sample size with a traditional vaccine.

	Vax. Eff.	$R_{0,v}$	Estimate protection or risk difference: mean $\langle CI \rangle$	Req. sample size from simulation, N_T^*	Power	Type I Error
$\Delta_1(5)$	0.8	0.9	25% $\langle 25\% \rangle$	10 clus., 1000 indiv.	81%	4%
		1.1	41% $\langle 26\% \rangle$	10 clus., 1000 indiv.	95%	4%
	1	0.9	32% $\langle 26\% \rangle$	10 clus., 1000 indiv.	91%	4%
		1.1	49% $\langle 24\% \rangle$	10 clus., 1000 indiv.	92%	4%
$\Delta_3(5)$	0.8	0.9	26% $\langle 26\% \rangle$	10 clus., 1000 indiv.	83%	4%
		1.1	42% $\langle 27\% \rangle$	10 clus., 1000 indiv.	95%	4%
	1	0.9	33% $\langle 27\% \rangle$	10 clus., 1000 indiv.	91%	4%
		1.1	52% $\langle 25\% \rangle$	10 clus., 1000 indiv.	94%	5%
$\theta_1(5)$	0.8	0.9	58% $\langle 30\% \rangle$	8 clus., 800 indiv.	80%	4%
		1.1	64% $\langle 23\% \rangle$	8 clus., 800 indiv.	80%	4%
	1	0.9	69% $\langle 19\% \rangle$	8 clus., 800 indiv.	91%	4%
		1.1	70% $\langle 18\% \rangle$	8 clus., 800 indiv.	90%	4%
$\theta_2(5)$	0.8	0.9	19% $\langle 21\% \rangle$	18 clus., 1800 indiv.	81%	5%
		1.1	35% $\langle 30\% \rangle$	10 clus., 1000 indiv.	88%	4%
	1	0.9	26% $\langle 27\% \rangle$	12 clus., 1200 indiv.	86%	4%
		1.1	45% $\langle 29\% \rangle$	10 clus., 1000 indiv.	91%	4%

<https://doi.org/10.1371/journal.pcbi.1012779.t001>

Table 2. Estimated overall protection or risk difference from and required sample sizes, N_T^* , for trials comparing a transmissible vaccine to a traditional vaccine when clusters have overdispersed contact structures. Prior to the outbreak, $\alpha=5\%$ of each cluster is vaccinated with either a transmissible or traditional, non-transmissible vaccine. The contact structure is moderately overdispersed ($k=1$) and trials are anticipatory. The estimand of interest is noted in the first column: $\Delta_1(5)$ refers to the overall protection, $\Delta_3(5)$ refers to the spillover protection, $\theta_1(5)$ refers to the difference in risk of infection for indirectly vaccinated animals in the transmissible case compared to those unvaccinated in the traditional case, and $\theta_2(5)$ refers to the difference in risk of infection for never vaccinated animals in the transmissible case. Vax. Eff. is the true direct efficacy of the vaccine to decrease susceptibility of vaccinated animals. $R_{0,v}$ refers to the R_0 of the vaccine. Estimate Protection or risk difference is the mean estimate of the noted estimand comparing transmissible vaccines to traditional vaccines. $\langle . \rangle$ indicates the mean 95% confidence interval width. Req. sample size from simulation, N_T^* is the number of clusters and number of animals required to estimate the effect in order to achieve at least 80% power. Power is the percentage of simulations with a p-value $\leq 5\%$ across 1000 trial simulations. Type I error is the percentage of simulations with a p-value $\leq 5\%$ across 3000 trial simulations for this sample size with a traditional vaccine.

	Vax. Eff.	$R_{0,v}$	Estimate protection or risk difference: mean $\langle CI \rangle$	Req. sample size from simulation, N_T^*	Power	Type I Error
$\Delta_1(5)$	0.8	1.1	19% $\langle 18\% \rangle$	22 clus., 2200 indiv.	80%	5%
$\Delta_3(5)$	0.8	1.1	20% $\langle 19\% \rangle$	22 clus., 2200 indiv.	84%	5%
$\theta_1(5)$	0.8	1.1	24% $\langle 24\% \rangle$	14 clus., 1400 indiv.	84%	5%
$\theta_2(5)$	0.8	1.1	19% $\langle 19\% \rangle$	24 clus., 2400 indiv.	83%	5%

<https://doi.org/10.1371/journal.pcbi.1012779.t002>

sampling 1000 animals total to 22 clusters sampling 2200 animals total. Similarly, when the vaccine is weakly transmissible and vaccine efficacy is imperfect, if the trial design is reactionary rather than anticipatory, N_T^* for the overall causal effect increases from 10 clusters sampling 1000 animals to 18 clusters sampling 1800 animals total. But although sample sizes increase in these situations, they remain feasible as they are no greater than twice the maximum number we would expect to sample across all three estimands in our best case scenario of anticipatory trials in environments with non-overdispersed contacts (Table 1).

Table 3. Estimated overall protection from and required sample sizes, N_T^* for reactionary trials comparing a transmissible vaccine to a traditional vaccine. When 0.5% of the cluster is infected, $\alpha=5\%$ of each cluster is vaccinated with either a transmissible or traditional, non-transmissible vaccine. The contact structure is Poisson distributed. $\Delta_1(5)$ is the overall protection. Vax. Eff. is the true direct efficacy of the vaccine to decrease susceptibility of vaccinated animals. $R_{0,v}$ refers to the R_0 of the vaccine. Estimate protection is the mean estimate of the overall protection conferred by transmissible vaccines compared to traditional vaccines. $\langle \cdot \rangle$ indicates the mean 95% confidence interval width. Req. sample size from simulation, N_T^* , is the number of clusters and number of animals required to estimate the effect in order to achieve at least 80% power. Power is the percentage of simulations with a p-value $\leq 5\%$ across 1000 trial simulations. Type I error is the percentage of simulations with a p-value $\leq 5\%$ across 3000 trial simulations for this sample size with a traditional vaccine.

	Vax. Eff.	$R_{0,v}$	Estimate protection: mean $\langle CI \rangle$	Req. sample size from simulation, N_T^*	Power	Type I Error
$\Delta_1(5)$	0.8	0.9	14% $\langle 16\% \rangle$	18 clus., 1800 indiv.	82%	5%
		1.1	23% $\langle 23\% \rangle$	10 clus., 1000 indiv.	84%	4%

<https://doi.org/10.1371/journal.pcbi.1012779.t003>

N_T^* of approximately 10 clusters sampling 1000 animals total for the overall causal effect do not change for trials that compare the transmissible vaccine to pure control groups that do not receive any vaccine (Table 4). Both the power and the estimate of protection increase in this scenario, and the greater power in these cases may indicate that for other parameter settings N_T^* would be lower.

Approximate sample size formula results

Fig 3 shows the relationships between N_T^* estimated by an approximate analytic formula for trials of the overall causal effect of transmissible vaccines ($\Delta_1(5)$), with factors unique to transmissible vaccines, as well as with the trial design. As the $R_{0,v}$ of the vaccine increases or the between-cluster variance decreases, N_T^* decreases. Similarly, we find that as the R_0 of the wildtype pathogen increases, N_T^* first decreases, since larger exposure of the pathogen will highlight differences in protection conferred by the two vaccine types, then eventually increases as differences in the final proportion of infected individuals between the two vaccine types decrease (Fig C in S1 Text). Thus, although exposure to the pathogen is necessary to detect a protective effect, as the R_0 of the wildtype pathogen increases, it is harder to observe differences in the proportions of infected individuals. Decreasing the vaccine efficacy will also increase sample sizes (Fig C in S1 Text). Sampling more animals from clusters can increase the power to detect the protection of these transmissible vaccines, and sampling less from more clusters may have higher power than sampling more from fewer clusters (middle panel Fig 3).

Discussion

The development and utilization of transmissible vaccines for protection of both wildlife and domestic animal populations against pathogens remains a contentious issue [9]. Identifying the protective effects conferred by transmissible vaccines relative to traditional vaccines in real-world epidemiological conditions is necessary for the successful development of transmissible vaccines [13,15,20]. Towards this aim, we propose a field trial design that utilizes causal inference methods to capture the overall, spillover, and vaccine protective effects of transmissible vaccines. We also provide estimates of differences in the risks of infection for subpopulations unique to transmissible vaccines: animals indirectly vaccinated with the transmissible vaccine, and never vaccinated animals. Trials should be powered to identify causal effects, but with sufficient resources, differences in risk of infection for subpopulations unique to transmissible vaccines can also be estimated. We prove the unbiasedness, consistency, and asymptotic normality for estimators of both causal and statistical effects and perform a simulation study of several of the estimators to test the feasibility of this trial design in realistic epidemic scenarios [1]. In our simulations we vary the timing of intervention, transmissibility of the vaccine, efficacy of the vaccine, contact structure of the population.

Table 4. Estimated overall protection from and required sample sizes, N_T^* , for anticipatory trials comparing a transmissible vaccine to a pure control group with no vaccination. Clusters are either vaccinated at a rate of $\alpha=5\%$ prior to the outbreak with a transmissible vaccine or do not receive any vaccination. The contact structure is Poisson distributed and trials are anticipatory. $\Delta_1(5)$ is the overall protection. Vax. Eff. is the true direct efficacy of the vaccine to decrease susceptibility of vaccinated animals. $R_{0,v}$ refers to the R_0 of the vaccine. Estimate protection is the mean estimate of the overall protection conferred by transmissible vaccines compared to traditional vaccines. $\langle \cdot \rangle$ indicates the mean 95% confidence interval width. Req. sample size from simulation, N_T^* , is the number of clusters and number of animals required to estimate the effect in order to achieve at least 80% power. Power is the percentage of simulations with a p-value $\leq 5\%$ across 1000 trial simulations. Type I error is the percentage of simulations with a p-value $\leq 5\%$ across 3000 trial simulations for this sample size with a traditional vaccine.

	Vax. Eff.	$R_{0,v}$	Estimate protection: mean $\langle CI \rangle$	Req. sample size from simulation, N_T^*	Power	Type I Error
$\Delta_1(5)$	0.8	0.9	30% $\langle 25\% \rangle$	10 clus., 1000 indiv.	95%	4%
		1.1	47% $\langle 26\% \rangle$	10 clus., 1000 indiv.	97%	4%

<https://doi.org/10.1371/journal.pcbi.1012779.t004>

We find that the sample sizes needed to adequately power our proposed anticipatory trial design to capture overall and spillover effects are feasible in groups of animals with non-overdispersed contact structures, since approximately 10 clusters sampling 1000 animals total are needed to adequately power the trial. Although trials should primarily aim to estimate the overall and spillover causal effects of the transmissible vaccine, such trials would be adequately powered to estimate the difference in risk of infection for indirectly vaccinated animals (required sample size, $N_T^* = 8$ clusters sampling 800 animals total) and if the sample size increases to 10–18 clusters sampling 1000–1800 animals total, depending on the vaccine efficacy and R_0 of the transmissible vaccine, the trial would capture the difference in risk of infection for never vaccinated animals. When the trial compares transmissible vaccines to pure control groups that receive no vaccination, N_T^* for the overall causal effect remains at 10 clusters sampling 1000 animals total but with higher than the nominal power. If investigators use the maximum N_T^* across all four estimands (overall causal effect, spillover causal effect, difference in risk for indirectly vaccinated animals, and difference in risk for never vaccinated animals) for a given scenario, all four estimands can be estimated, although this ignores multiple testing effects. Through an approximate sample size formula we derive, we also find that N_T^* needed to power these trials will decrease with increases in the strength of the transmissible vaccine, $R_{0,v}$, number sampled from each cluster, s , and vaccine efficacy. These sample sizes will also decrease with decreases in the between cluster variance of the proportion infected and R_0 of the wildtype pathogen. These results can be used to quickly estimate N_T^* for other settings, before confirmation using a full simulation model.

Estimates of these effects may be especially sensitive to cluster-level factors such as heterogeneity in contact structure. Despite this sensitivity, we find that generally, N_T^* needed to power trials in order to capture the protective effects are achievable, even for weakly transmissible vaccines [14]. Many real-world populations for which transmissible vaccines have been proposed are expected to contain heterogeneities in contact rates, yet measurement of protection is still feasible, at most 10 cluster sampling 1000 animals total for the overall causal effect [55]. If vaccines compete with the wildtype pathogen for susceptible animals, as they do in reactionary trials, protection could decrease, increasing N_T^* to 18 clusters sampling 1800 animals total. Similarly, in environments with overdispersed contact structures, N_T^* may increase to sampling 22 clusters sampling 2200 animals total. The trial design remains feasible even for these more challenging settings, as previous campaigns and field trials for poultry vaccines have been able to vaccinate hundreds, thousands, and in one case over hundreds of thousands of poultry at the village level within a year [26,53,54]. Given this historical precedent of large scale campaigns of poultry vaccination, the sample sizes we propose, from the hundreds to thousands, are comparable and thus logistically feasible.

Our results show that anticipatory trials may be a useful tool to decrease N_T^* to power the trial, but that sample size increases may be required to buffer against unknown parameters,

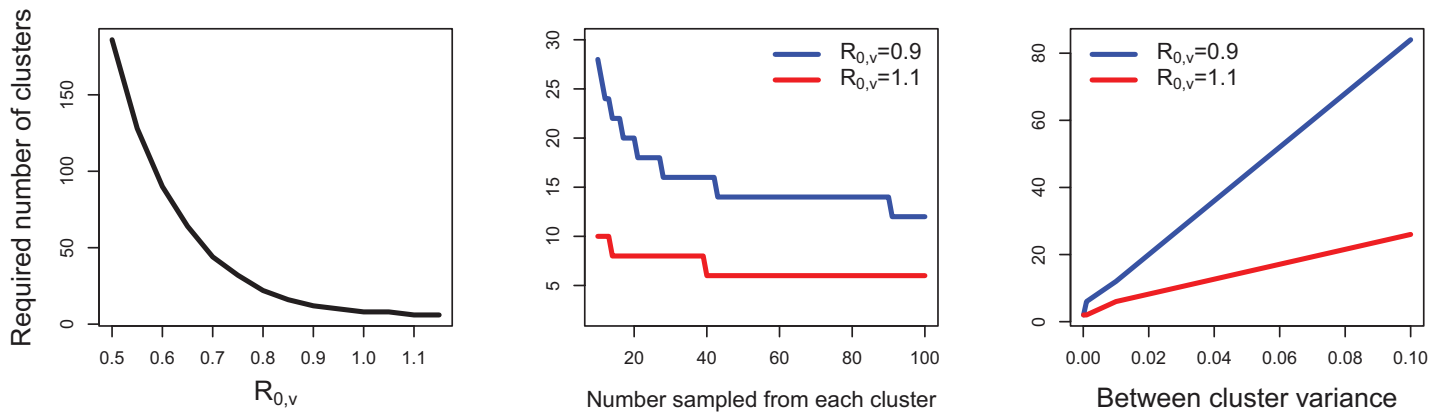


Fig 3. Required sample sizes, N_T^* , for overall protection estimand ($\Delta_i(5)$) will depend on the transmissible vaccine's protective strength, as well as on the trial design. Our approximate formula reveals that as the R_0 of the vaccine increases (left panel) or number sampled in each cluster, s , increases (middle panel), N_T^* for trials that measure the overall protection that transmissible vaccines confer relative to traditional vaccines will decrease. N_T^* will increase as the between-cluster variance in the final size proportion increases. All other parameters are set at $R_{0,v} = 1.1$, number sampled from each cluster, $s = 100$, between cluster variance = 0.01, $R_{0,w} = 2$, proportion initially recovered = 0, vaccine efficacy = 80%, initial vaccinated proportion, $\alpha = 5\%$.

<https://doi.org/10.1371/journal.pcbi.1012779.g003>

such as the exact contact structure of the cluster and transmissibility of the vaccine, when conducting the trial. In such cases, investigators could choose the maximum sample size across all possible scenarios to minimize the risk of conducting an underpowered trial. To estimate the overall protective effect, this amounts to 22 clusters sampling 2200 animals total, which more than doubles the number of animals from the lower bound of 10 clusters sampling 1000 animals total. The most practical trial design will depend on these three ecological factors. For example, in cases where the R_0 of the wildtype is high, such as for poultry pathogens (thus raising N_T^* , as shown in the analytical sample size results), investigators may prefer to estimate the effects of strongly transmissible vaccines, rather than weakly transmissible vaccines (all else being equal), in order to decrease N_T^* .

Although evolution within the vaccine strain might first be identified in laboratory studies, with monitoring of health outcomes across individuals the field trial designs we derive allow us to observe if unintended evolution occurs at scale [13]. Further, the protective efficacy estimates from these trials can be used to quantify the protective benefits of these vaccines in order to weigh them against potential unintended evolution. For example, with the trial design and setting we propose, strongly transmissible vaccines with 80% efficacy are expected to reduce the infection burden of the cluster by 41% on average compared to traditional vaccines. The 41% overall protection beyond that of the traditional vaccine can then be weighed against the potential for unintended evolution. If a second trial can be done comparing a weakly transmissible vaccine to a traditional vaccine, these results can also be used to weigh against the evolutionary risk of using a vaccine with higher transmissibility. For example, weakly transmissible vaccines are expected to reduce infection burden by 25% compared to a traditional vaccine, and this reduction can be compared to the 41% reduction of strongly transmissible vaccines. Future work that makes further assumptions of rates of reversion to wildtype could address which of the proposed estimands is most relevant for assessing unintended evolution and what sample size is required to estimate it. These characterizations would lend further support (if there was no observed undesirable evolution) or strong discouragement for a proposed transmissible vaccine (if there was observed undesirable evolution).

The trial design we propose is most suitable for populations where the wildtype pathogen is newly introduced in the population, sparking an epidemic, rather than circulating

endemically. For example, since the wildtype rabies virus circulates endemically in these populations, standard epidemiological theory would predict that the vaccine strain will not circulate greatly in the population if it has a lower R_0 than the wildtype, which is usually the case [56–58]. When pathogens are more episodic, there are periods without circulating pathogen, and thus the vaccine has time to spread further in the population without competing with the wildtype pathogen for susceptible animals. Thus, transmissible vaccines against poultry pathogens which create seasonal outbreaks may be an appropriate candidate for such trials. Future work may study how characteristics of endemicity, such as prior immunity, can affect the sample sizes we provide.

The models used here can also be extended to account for setting-specific effects, such as the distribution of immunity within explicit subpopulations and less than perfect protectiveness of prior immunity. Modeling explicit sub-clusters, where there is greater within-cluster connectivity and less between-cluster connectivity may also change our results, since transmissible vaccines may be more easily transmitted within sub-clusters, and less easily transmitted between clusters. Since contact structures for poultry networks as well as bat colonies have been constructed [39,59] these more specific dependencies can potentially be incorporated in future work. Informative cluster size may also increase N_T^* if the rate of imported infections decreases, and thus avoiding recruitment of smaller clusters may be needed to mitigate this risk. Characteristics of leaky vaccines may affect N_T^* : while vaccinated animals may still get infected if vaccine efficacy is imperfect, the vaccine may still reduce infection severity, decreasing the time of the infectiousness period and probability of forward transmission. This may further reduce infection burden and N_T^* if transmissible vaccines spread far in a population. Finally, the approximate sample size formula assumes normality of the effect and null distributions, and deterministic solutions for the final size of outbreaks, in contrast to the simulation study results. Thus it may be difficult to directly compare these results to N_T^* from the simulation study. Accounting for these differences may allow more direct comparison between the analytical solution and simulation study results. Both the simulation results as well as analytical results we provide can easily be extended to test other epidemiological settings and design features; the code we use and instructions for how to implement it is freely available in the Github repository we provide. For example, in this study we obtain the required sample size to adequately power the trial, but the code we provide can also be used to identify the required sample size to reduce the confidence interval uncertainty to a specific width.

The statistical estimands provide epidemiological information for indirectly vaccinated animals, $\theta_1(5)$, and never vaccinated animals, $\theta_2(5)$, in clusters where the transmissible vaccine is implemented and are potentially useful for modeling and surveillance. We leave to future work detailed characterization of the interpretation of these estimands and their relationship with causal protective effects for these groups. As these estimands do not lend themselves to the exchangeability assumption on the individual level since indirect vaccination is not randomly applied, and thus there is potential confounding, observational study methods may be useful to estimate causal effects. For example, if the degree of animals is known, then a matching could be done on degree to account for the potential confounder of higher degree animals being more likely to be indirectly vaccinated as well as infected with the wildtype pathogen. Identification of these causal estimands within an exposure-mediation interaction framework, where the mediator is whether or not an initially unvaccinated animal has been indirectly vaccinated, may also be successful [60]. Accounting for explicit network structures into these estimands is also warranted, given that they will certainly depend on the underlying network contact structure.

In this paper, we have laid the groundwork for the design of field trials to evaluate transmissible vaccines. Future studies that aim to identify ways to relax assumptions of this study

could help further frame and specify trial designs needed for the potential introduction of this promising tool to lower the infection burden for many wildlife and livestock animal populations.

Supplementary information

S1 Text. Appendix A: Additional Figures and Tables. Appendix B: Estimands and Estimators Appendix C: Approximate sample size formula Fig A: SEIR model with vaccination. Fig B: Direct and indirect protection of a transmissible vaccine broken down by individual. Fig C: Required sample sizes, N_T^* , depend on R_0 of the wildtype pathogen as well as the vaccine efficacy. Table A: Epidemiological parameter values. Table B: Estimated overall protection from and required sample sizes, N_T^* , for trials comparing a transmissible vaccine to a traditional vaccine when cluster size $n = 500$.

(PDF)

Author contributions

Conceptualization: Justin K. Sheen, Michael Z. Levy, Charlotte Jessica E. Metcalf.

Data curation: Justin K. Sheen.

Formal analysis: Justin K. Sheen, Lee Kennedy-Shaffer.

Investigation: Justin K. Sheen, Lee Kennedy-Shaffer, Charlotte Jessica E. Metcalf.

Methodology: Justin K. Sheen, Lee Kennedy-Shaffer.

Software: Justin K. Sheen, Lee Kennedy-Shaffer.

Visualization: Justin K. Sheen.

Writing – original draft: Justin K. Sheen, Charlotte Jessica E. Metcalf.

Writing – review & editing: Justin K. Sheen, Lee Kennedy-Shaffer, Michael Z. Levy, Charlotte Jessica E. Metcalf.

References

1. Hudgens MG, Halloran ME. Toward causal inference with interference. *J Am Stat Assoc.* 2008 Jun 1;103(482):832–42. <https://doi.org/10.1198/016214508000000292> PMID: 19081744
2. Barkley BG, Hudgens MG, Clemens JD, Ali M, Emch ME. Causal inference from observational studies with clustered interference, with application to a cholera vaccine study. *Ann Appl Stat.* 2020;14(3). <https://doi.org/10.1214/19-aos1314>
3. Basse G, Feller A. Analyzing two-stage experiments in the presence of interference. *J Am Stat Assoc.* 2018 Jan 2;113(521):41–55. <https://doi.org/10.1080/01621459.2017.1323641>
4. Bull J, Smithson M, Nuismer S. Transmissible viral vaccines. *Trends Microbiol.* 2018 Jan 1;26(1):6–15.
5. Nuismer SL, May R, Basinski A, Remien CH. Controlling epidemics with transmissible vaccines. *PLoS One.* 2018;13(5):e0196978. <https://doi.org/10.1371/journal.pone.0196978> PMID: 29746504
6. Nuismer SL, Bull JJ. Self-disseminating vaccines to suppress zoonoses. *Nat Ecol Evol.* 2020 Sep;4(9):1168–73. <https://doi.org/10.1038/s41559-020-1254-y> PMID: 32719452
7. Burns CC, Diop OM, Sutter RW, Kew OM. Vaccine-derived polioviruses. *J Infect Dis.* 2014 Nov 1;210 Suppl 1:S283–93. <https://doi.org/10.1093/infdis/jiu295> PMID: 25316847
8. Kew OM, Sutter RW, de Gourville EM, Dowdle WR, Pallansch MA. Vaccine-derived polioviruses and the endgame strategy for global polio eradication. *Annu Rev Microbiol.* 2005 Oct 13;59(1):587–635.
9. Sandbrink JB, Watson MC, Hebbeler AM, Esvelt KM. Safety and security concerns regarding transmissible vaccines. *Nat Ecol Evol.* 2021 Apr;5(4):405–6. <https://doi.org/10.1038/s41559-021-01394-3> PMID: 33542476
10. Metzger VT, Lloyd-Smith JO, Weinberger LS. Autonomous targeting of infectious superspreaders using engineered transmissible therapies. *PLoS Comput Biol.* 2011 Mar 17;7(3):e1002015. <https://doi.org/10.1371/journal.pcbi.1002015> PMID: 21483468

11. Murphy A, Redwood A, Jarvis M. Self-disseminating vaccines for emerging infectious diseases. *Expert Rev Vaccines*. 2016 Jan 2;15(1):31–9.
12. Griffiths ME, Meza DK, Haydon DT, Streicker DG. Inferring the disruption of rabies circulation in vampire bat populations using a betaherpesvirus-vectored transmissible vaccine. *Proc Natl Acad Sci U S A*. 2023 Mar 14;120(11):e2216667120. <https://doi.org/10.1073/pnas.2216667120> PMID: [36877838](https://pubmed.ncbi.nlm.nih.gov/36877838/)
13. Streicker D, Griffiths M, Antia R, Bergner L, Bowman P, de Moraes M, et al. Developing transmissible vaccines for animal infections. *Science*. 2024 Apr 10;384(6693):275–7.
14. Nuismer SL, Althouse BM, May R, Bull JJ, Stromberg SP, Antia R. Eradicating infectious disease using weakly transmissible vaccines. *Proc R Soc B: Biol Sci*. 2016 Oct 26;283(1841):20161903.
15. Varrelman TJ, Remien CH, Basinski AJ, Gorman S, Redwood A, Nuismer SL. Quantifying the effectiveness of betaherpesvirus-vectored transmissible vaccines. *Proc Natl Acad Sci U S A*. 2022 Jan 25;119(4):e2108610119. <https://doi.org/10.1073/pnas.2108610119> PMID: [35046024](https://pubmed.ncbi.nlm.nih.gov/35046024/)
16. Griffiths ME, Broos A, Bergner LM, Meza DK, Suarez NM, da Silva Filipe A, et al. Longitudinal deep sequencing informs vector selection and future deployment strategies for transmissible vaccines. *PLoS Biol*. 2022 Apr 10;20(4):e3001580. <https://doi.org/10.1371/journal.pbio.3001580> PMID: [35439242](https://pubmed.ncbi.nlm.nih.gov/35439242/)
17. Annapragada A, Borgerson C, Iams S, Ravelomanantsoa MA, Crawford GC, Helin M, Anjaranirina EJ, Randriamady HJ, Golden CD. Modeling the impact of newcastle disease virus vaccinations on chicken production systems in northeastern madagascar. *Front Vet Sci*. 2019 Sep 26;6:305.
18. Kariithi HM, Ferreira HL, Welch CN, Ateya LO, Apopo AA, Zoller R, et al. Surveillance and genetic characterization of virulent newcastle disease virus subgenotype V.3 in indigenous chickens from backyard poultry farms and live bird markets in Kenya. *Viruses*. 2021 Jan 13;13(1):103. <https://doi.org/10.3390/v13010103> PMID: [33451125](https://pubmed.ncbi.nlm.nih.gov/33451125/)
19. Torres JM, Sánchez C, Ramírez MA, Morales M, Bárcena J, Ferrer J, et al. First field trial of a transmissible recombinant vaccine against myxomatosis and rabbit hemorrhagic disease. *Vaccine*. 2001 Aug 14;19(31):4536–43. [https://doi.org/10.1016/s0264-410x\(01\)00184-0](https://doi.org/10.1016/s0264-410x(01)00184-0) PMID: [11483281](https://pubmed.ncbi.nlm.nih.gov/11483281/)
20. Akobeng AK. Understanding randomised controlled trials. *Arch Dis Child*. 2005 Aug 1;90(8):840–4. <https://doi.org/10.1136/adc.2004.058222> PMID: [16040885](https://pubmed.ncbi.nlm.nih.gov/16040885/)
21. Hitchings MDT, Lipsitch M, Wang R, Bellan SE. Competing effects of indirect protection and clustering on the power of cluster-randomized controlled vaccine trials. *Am J Epidemiol*. 2018 Aug 1;187(8):1763–71. <https://doi.org/10.1093/aje/kwy047> PMID: [29522080](https://pubmed.ncbi.nlm.nih.gov/29522080/)
22. Baird S, Bohren J, McIntosh C, Özler B. Optimal design of experiments in the presence of interference. *Rev Econ Stat*. 2018 Dec 1;100(5):844–60.
23. Jiang Z, Imai K, Malani A. Statistical inference and power analysis for direct and spillover effects in two-stage randomized experiments. *Biometrics*. 2023 Sep;79(3):2370–81. <https://doi.org/10.1111/biom.13782> PMID: [36285364](https://pubmed.ncbi.nlm.nih.gov/36285364/)
24. Halloran ME, Hudgens MG. Dependent happenings: a recent methodological review. *Curr Epidemiol Rep*. 2016 Dec 3;3(4):297–305. <https://doi.org/10.1007/s40471-016-0086-4> PMID: [28133589](https://pubmed.ncbi.nlm.nih.gov/28133589/)
25. Kahan BC, Li F, Copas AJ, Harhay MO. Estimands in cluster-randomized trials: choosing analyses that answer the right question. *Int J Epidemiol*. 2023 Feb 1;52(1):107–18. <https://doi.org/10.1093/ije/dyab131> PMID: [35834775](https://pubmed.ncbi.nlm.nih.gov/35834775/)
26. Nwanta JA, Umoh JU, Abdu PA, Ajogi I, Egege SC, Adeiza AA. Field trial of Malaysian thermostable Newcastle disease vaccine in village chickens in Kaduna State, Nigeria. *Sokoto J Vet Sci*. 2008;7(2).
27. Bello MB, Yusoff K, Ideris A, Hair-Bejo M, Peeters BPH, Omar AR. Diagnostic and vaccination approaches for newcastle disease virus in poultry: the current and emerging perspectives. *Biomed Res Int*. 2018;2018:7278459. <https://doi.org/10.1155/2018/7278459> PMID: [30175140](https://pubmed.ncbi.nlm.nih.gov/30175140/)
28. Suarez D. DIVA vaccination strategies for avian influenza virus. *Avian Dis*. 2012 Dec 1;56(4s1):836–44.
29. Kahan BC, Li F, Blette B, Jairath V, Copas A, Harhay M. Informative cluster size in cluster-randomised trials: a case study from the TRIGGER trial. *Clin Trials*. 2023 Dec;20(6):661–9. <https://doi.org/10.1177/17407745231186094> PMID: [37439089](https://pubmed.ncbi.nlm.nih.gov/37439089/)
30. VanderWeele TJ, Tchetgen EJ, Halloran ME. Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. *Epidemiology*. 2012 Sep 1;23(5):751–61.
31. Olwande PO, Ogara WO, Okothe SO, Muchemi G, Okoth E, Odindo MO, et al. Assessing the productivity of indigenous chickens in an extensive management system in southern Nyanza, Kenya. *Trop Anim Health Prod*. 2010 Feb;42(2):283–8. <https://doi.org/10.1007/s11250-009-9418-4> PMID: [19680773](https://pubmed.ncbi.nlm.nih.gov/19680773/)

32. Campbell ZA, Marsh TL, Mpolya EA, Thumbi SM, Palmer GH. Newcastle disease vaccine adoption by smallholder households in Tanzania: identifying determinants and barriers. *PLoS One*. 2018;13(10):e0206058. <https://doi.org/10.1371/journal.pone.0206058> PMID: 30356260
33. Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford JM Jr, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. *Int J Epidemiol*. 2018;47(1):332–47. <https://doi.org/10.1093/ije/dyx201> PMID: 29106568
34. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton university press; 2008.
35. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005 Nov 17;438(7066):355–9. <https://doi.org/10.1038/nature04153> PMID: 16292310
36. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Superspreading events in the transmission dynamics of SARS-CoV-2: opportunities for interventions and control. *PLoS Biol*. 2020 Nov;18(11):e3000897. <https://doi.org/10.1371/journal.pbio.3000897> PMID: 33180773
37. Okeno TO, Kahi AK, Peters KJ. Characterization of indigenous chicken production systems in Kenya. *Trop Anim Health Prod*. 2012 Mar;44(3):601–8. <https://doi.org/10.1007/s11250-011-9942-x> PMID: 21805308
38. Rasamoelina-Andriamanivo H, Duboz R, Lancelot R, Maminiaina OF, Jourdan M, Rakotondramaro TM, et al. Description and analysis of the poultry trading network in the Lake Alaotra region, Madagascar: implications for the surveillance and control of Newcastle disease. *Acta Tropica*. 2014 Jul 1;135:10–8.
39. Newman MEJ. The structure and function of complex networks. *SIAM Rev*. 2003;45(2):167–256. <https://doi.org/10.1137/s003614450342480>
40. Kennedy DA, Dunn PA, Read AF. Modeling Marek's disease virus transmission: a framework for evaluating the impact of farming practices and evolution. *Epidemics*. 2018 Jun 1;23:85–95.
41. Chukwudi OE, Chukwuemeka ED, Mary U. Newcastle disease virus shedding among healthy commercial chickens and its epidemiological importance. *Pak Vet J*. 2012 Jul 1;32(3):354–6.
42. Brown VR, Bevins SN. A review of virulent Newcastle disease viruses in the United States and the role of wild birds in viral persistence and spread. *Vet Res*. 2017 Dec;48(1):68. <https://doi.org/10.1186/s13567-017-0475-9> PMID: 29073919
43. Chuma F, Mwanga GG, Masanja VG. Application of optimal control theory to newcastle disease dynamics in village chicken by considering wild birds as reservoir of disease virus. *J Appl Math*. 2019;2019:1–14. <https://doi.org/10.1155/2019/3024965>
44. Ferreira HL, Taylor TL, Dimitrov KM, Sabra M, Afonso CL, Suarez DL. Virulent newcastle disease viruses from chicken origin are more pathogenic and transmissible to chickens than viruses normally maintained in wild birds. *Vet Microbiol*. 2019 Aug 1;235:25–34.
45. Miller J, Ting T. EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *JOSS*. 2019;4(44):1731. <https://doi.org/10.21105/joss.01731>
46. Huang FL. Using cluster bootstrapping to analyze nested data with a few clusters. *Educ Psychol Meas*. 2018 Apr;78(2):297–318. <https://doi.org/10.1177/0013164416678980> PMID: 29795957
47. Field CA, Welsh AH. Bootstrapping clustered data. *J R Stat Soc Series B: Stat Methodol*. 2007 Jun;69(3):369–90. <https://doi.org/10.1111/j.1467-9868.2007.00593.x>
48. Bachmann MO, Fairall L, Clark A, Mugford M. Methods for analyzing cost effectiveness data from cluster randomized trials. *Cost Eff Resour Alloc*. 2007 Dec;5:12. <https://doi.org/10.1186/1478-7547-5-12> PMID: 17822546
49. Ng ES-W, Grieve R, Carpenter JR. Two-stage nonparametric bootstrap sampling with shrinkage correction for clustered data. *The Stata J: Prom Commun Stat Stata*. 2013 Mar;13(1):141–64. <https://doi.org/10.1177/1536867x1301300111>
50. Thompson JA, Leyrat C, Fielding KL, Hayes RJ. Cluster randomised trials with a binary outcome and a small number of clusters: comparison of individual and cluster level analysis method. *BMC Med Res Methodol*. 2022 Aug 12;22(1):222. <https://doi.org/10.1186/s12874-022-01699-2> PMID: 35962318
51. Kahn R, Kennedy-Shaffer L, Grad YH, Robins JM, Lipsitch M. Potential biases arising from epidemic dynamics in observational seroprotection studies. *Am J Epidemiol*. 2021;190(2):328–35. <https://doi.org/10.1093/aje/kwaa188> PMID: 32870977
52. Arino J, Brauer F, van den Driessche P, Watmough J, Wu J. A final size relation for epidemic models. *Math Biosci Eng*. 2007 Jan 31;4(2):159–75. <https://doi.org/10.3934/mbe.2007.4.159> PMID: 17658921

53. Henning J, Morton J, Pym R, Hla T, Meers J. Evaluation of strategies to improve village chicken production: controlled field trials to assess effects of Newcastle disease vaccination and altered chick rearing in Myanmar. *Prevent Vet Med.* 2009 Jul 1;90(1–2):17–30.
54. Msoffe PLM, Bunn D, Muhairwa AP, Mtambo MMA, Mwamhehe H, Msago A, et al. Implementing poultry vaccination and biosecurity at the village level in Tanzania: a social strategy to promote health in free-range poultry populations. *Trop Anim Health Prod.* 2010 Feb;42(2):253–63. <https://doi.org/10.1007/s11250-009-9414-8> PMID: [19688307](https://pubmed.ncbi.nlm.nih.gov/19688307/)
55. Varrelman TJ, Basinski AJ, Remien CH, Nuismer SL. Transmissible vaccines in heterogeneous populations: implications for vaccine design. *One Health.* 2019 Jun 1;7:100084. <https://doi.org/10.1016/j.onehlt.2019.100084> PMID: [30859117](https://pubmed.ncbi.nlm.nih.gov/30859117/)
56. van Baalen M, Sabelis MW. The dynamics of multiple infection and the evolution of virulence. *The Am Nat.* 1995;146(6):881–910. <https://doi.org/10.1086/285830>
57. Anderson RM, May RM. Coevolution of hosts and parasites. *Parasitology.* 1982;85(Pt 2):411–26. <https://doi.org/10.1017/s0031182000055360> PMID: [6755367](https://pubmed.ncbi.nlm.nih.gov/6755367/)
58. Blackwood JC, Streicker DG, Altizer S, Rohani P. Resolving the roles of immunity, pathogenesis, and immigration for rabies persistence in vampire bats. *Proc Natl Acad Sci U S A.* 2013 Dec 17;110(51):20837–42. <https://doi.org/10.1073/pnas.1308817110> PMID: [24297874](https://pubmed.ncbi.nlm.nih.gov/24297874/)
59. Rocha F, Dias RA. The common vampire bat *Desmodus rotundus* (Chiroptera: Phyllostomidae) and the transmission of the rabies virus to livestock: a contact network approach and recommendations for surveillance and control. *Prev Vet Med.* 2020 Jan 1;174:104809. <https://doi.org/10.1016/j.prevetmed.2019.104809> PMID: [31756671](https://pubmed.ncbi.nlm.nih.gov/31756671/)
60. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol.* 2013 Oct 1;42(5):1511–9. <https://doi.org/10.1093/ije/dyt127> PMID: [24019424](https://pubmed.ncbi.nlm.nih.gov/24019424/)