BIOSTAT M235 Final Project

Hiroyasu Ando[1], Bowen Zhang[1], Puyuan Liu[1], and Jiachen Ai[1]

[1]Department of Biostatistics, University of California Los Angeles, Los Angeles, CA, USA

1 Acknowledgment

JA, PL, and HA contributed to the paper's explanation and discussion sections. BZ and HA contributed to the extension section. BZ came up with the extension idea and wrote the simulation code and HA proofread it.

2 Introduction

In infectious disease clinical trials, the effectiveness of an intervention (e.g., a vaccine) can be influenced by an individual's degree of exposure to the pathogen, which often varies across populations[1]. Standard effect measures, such as the average causal effect, aggregate outcomes across heterogeneous exposure intensities and thus fail to account for variation in contact patterns, transmission dynamics, or susceptibility. As a result, they are often not transportable across settings[2]. The per-exposure effect, defined as the effect of an intervention on infection risk per exposure, addresses this limitation by providing a population-independent measure of efficacy[3,4]. Unlike conventional estimates (e.g., intention-to-treat effects), it remains interpretable across populations with differing exposure rates. This property is particularly valuable in the presence of interference, where an individual's infection risk depends on the intervention status of others[1,5]. By isolating the effect at the level of individual exposures, the per-exposure effect offers a more robust parameter for transmission modeling and public health decision-making. Despite its conceptual importance, the per-exposure effect has historically lacked a consistent definition and is often estimated without explicitly stating the assumptions necessary for its validity[6]. To address this gap, O'Hagan et al.[6] formalized the per-exposure effect

within a potential outcomes framework and demonstrated that standard estimators, such as the average hazard ratio from a Cox model, produce biased estimates under common violations[7]. These violations include time-varying effects, dependence on prior exposures, unmeasured heterogeneity in infection susceptibility, and confounding of exposure.

In the sections that follow, we review the formal mathematical definition of the average per-exposure effect, outline the conditions required for unbiased estimation, and explore the practical limitations of standard analytic approaches. We then extend this framework through simulation to investigate how overlapping social structures and sampling strategies affect vaccine efficacy estimation. By comparing standard Cox models, stratified approaches, and frailty models, we highlight key sources of bias and propose methods to improve causal inference in infectious disease trials.

3 Summary

Let E_t denote the number of infection exposures at time t (e.g., $E_t = 1$ indicates a single exposure at time t). The exposure history from baseline to time t is denoted by $\bar{E}_t = (E_0, E_1, \ldots, E_t)$. Let $Y_{t+1}^{a,\bar{e}_t}$ be the potential infection outcome at time $t+1$ had the individual received intervention value a at baseline and experienced exposure history $\bar{e}_t$ through time t. Then $\Pr\left[Y_{t+1}^{a,\bar{e}_t} = 1\right]$ denote the proportion of participants who would have become infected by $t+1$ had everyone received the intervention value a and exposure history $\bar{e}_t$ through time t. The average per-exposure effect at time t is defined as the proportionate reduction in the conditional risk of infection following a single exposure at time t, among individuals who were uninfected and unexposed prior to t, as quantified by 1 minus the hazard ratio in equation (1)[6]. This quantity represents the discrete-time hazard of infection under intervention $a = 1$ relative to $a = 0$, conditional on first exposure at time t and no prior exposure or infection. By conditioning on a specific exposure pattern, the per-exposure effect isolates the intervention's direct impact and removes dependence on population-level exposure distributions or interference[6].

$$1 - \frac{\Pr\left[Y_{t+1}^{a=1,e_t=1,\bar{e}_{t-1}=\bar{0}_{t-1}} = 1 \mid Y_t^{a=1,\bar{e}_{t-1}=\bar{0}_{t-1}} = 0\right]}{\Pr\left[Y_{t+1}^{a=0,e_t=1,\bar{e}_{t-1}=\bar{0}_{t-1}} = 1 \mid Y_t^{a=0,\bar{e}_{t-1}=\bar{0}_{t-1}} = 0\right]} \tag{1}$$

In practice, however, the exact timing and occurrence of exposures are often unknown due to restricted feasibility and ethical concerns in conducting human challenge studies[6,8–10]. Consequently, the average per-exposure effect is often approximated by one minus the intention-to-treat average hazard ratio[11], estimated via a Cox proportional hazards model as equation (2). The hazard ratio ($\frac{\Pr[Y_{t+1}=1|Y_t=0,A=1]}{\Pr[Y_{t+1}=1|Y_t=0,A=0]}$) represents a weighted average of time-specific hazard ratios across all t.

$$1 - \frac{\Pr\left[Y_{t+1} = 1 \mid Y_t = 0, A = 1\right]}{\Pr\left[Y_{t+1} = 1 \mid Y_t = 0, A = 0\right]} \tag{2}$$

For the intention-to-treat hazard ratio to provide an unbiased estimate of the per-exposure effect, several conditions must be satisfied: (1) The average per-exposure effect must be constant over time; (2) The effect must not depend on the number of previous exposures to the pathogen during the follow-up; (3) There must be no unmeasured heterogeneity in infection susceptibility across individuals. All individual-level factors that influence susceptibility to infection (e.g., behavioral or biological risk factors) are either measured or evenly distributed across treatment groups; and (4) There must be no unmeasured confounding of the exposure–outcome relationship, i.e., no unobserved variable should influence both exposure to infection and the likelihood of infection itself[6].

Violations of these assumptions can introduce bias into hazard ratio estimates[6]. For instance, conditioning on remaining uninfected up to time t may induce selection bias, particularly in the presence of unmeasured susceptibility ($\bar{E}_{t-1}$ is unobserved). Likewise, if exposure is confounded by unmeasured variables, or if prior exposures influence both subsequent exposure and outcome risk, then the resulting estimates will fail to represent the true per-exposure effect. Therefore, while hazard-based estimators remain widely used due to feasibility, they often fail to recover the causal per-exposure effect without

further design or analytic adjustments[7,12,13].

4 Discussion

The standard Cox proportional hazards model estimates the hazard of infection at time t as $\lambda(t \mid A) = \lambda_0(t)\exp(\beta A)$, where A denotes intervention status and $\exp(\beta)$ is the hazard ratio[14]. In infectious disease studies, the vaccine efficacy (VE) is often estimated as $1 - \exp(\hat{\beta})$.

However, this estimate relies on strong assumptions that are frequently violated in practice: (1) *Proportional hazards.* If the true effect of the intervention varies over time, as commonly observed with waning vaccine efficacy, the Cox model yields a weighted average of time-specific hazard ratios, disproportionately influenced by periods with more events rather than those of greater causal relevance[7]. (2) *No selection bias.* Conditioning on being uninfected at time t (i.e., $Y_t = 0$) may induce collider bias, as Y_t is influenced by both treatment and unmeasured susceptibility U, thereby opening a non-causal path from A to Y_{t+1} (i.e., $A \rightarrow Y_t \leftarrow U \rightarrow Y_{t+1}$)[13]. (3) *No unmeasured confounding.* If unobserved variables affect both exposure risk and infection, the estimated hazard ratio conflates causal and non-causal effects, leading to biased estimates[12,15]. Figure 1 illustrates an example of unmeasured confounding affecting both exposure and infection outcome.

In particular, points (2) and (3) deserve special mention. In many cases, randomized controlled trials (RCTs) are considered the gold standard for eliminating bias from unmeasured susceptibility or confounding factors. However, it is important to recognize that in the context of infectious diseases, clinical trials often face additional challenges: it is necessary to measure all these susceptibility and confounding factors to obtain unbiased estimates. These challenges arise partly because the exposure status of trial participants is typically unobservable. If the exposure status were available, vaccine effectiveness (VE) could be estimated directly by simply regressing infection outcomes on exposure status.

Stratified Cox models specify the hazard as $\lambda(t \mid A, S = s) = \lambda_0^{(s)}(t) \exp(\beta A)$, allowing the baseline hazard $\lambda_0^{(s)}(t)$ to vary across strata defined by measured confounders or baseline risk factors[16]. This approach relaxes the proportional hazards assumption by accounting for heterogeneity in baseline risk, thereby improving robustness to violations of constant hazard ratios. While stratification helps reduce bias due to observed confounders and risk heterogeneity, it does not address bias from unmeasured confounding or unknown exposure histories[17].

5 Extension

5.1 Background and Motivation

Even when individuals are randomized within a vaccine trial and are exchangeable at baseline, heterogeneity in infection hazard can emerge stochastically due to differences in community-level epidemic dynamics. In a simulation study by Kahn et al., individuals are grouped into communities with different inter-connectedness. Stochastic transmission within communities can lead to disproportionate outbreaks in some clusters, thus inducing time-varying susceptibility heterogeneity across groups[18]. They employed a stratified Cox proportional hazards model, allowing the baseline hazard function to vary across communities.

While the stratified Cox model is useful in mitigating bias arising from differences in exposure rates, between communities[18], it is not a perfect methodology. It can be argued that within each community, there are further subdivisions—subcommunities—each of which may have a distinct hazard. In reality, individuals belong to multiple overlapping social groups, such as family, work, and friends, which can influence their exposure differently.

The study by Kahn et al. does not account for these subcommunity effects, and their

impact on vaccine effectiveness (VE) estimates remains unknown. To address this gap, in the following sections, we will develop a simulation framework that incorporates sub-communities and examine their influence on VE estimates. We will also explore methods to mitigate the biases that may arise due to this additional layer of social structure.

5.2 Methodology : Frailty Models for Susceptibility Heterogeneity from Overlapping Cluster

We are considering another setting in which each individual simultaneously belongs to two clusters. Specifically, we assume that every individual belongs to both a household group and a social group, such as a workplace, school, or other community setting. Household clusters tend to be small in size but exhibit higher per-contact transmissibility due to close and prolonged contact, whereas social groups are typically larger but involve lower transmission risk. Unlike the disjoint community structure in Kahn et al., where each individual is nested within a single group, our setting reflects a cross-classified cluster structure such that more closely mirrors real-world social organization, where individuals interact within both household and broader community contexts. Under this structure, nearly all individuals are connected either directly or indirectly through overlapping clusters. Therefore, our setting can be viewed as a substructure embedded within a larger community network, consistent with the regimes in Kahn et al. This interpretation is illustrated in Figure 2, where the left panel depicts disjoint community clusters and the right panel illustrates our overlapping household–social group structure.

While stratified Cox models provide a useful way to account for heterogeneity across known groups, they become impractical in settings involving overlapping clusters. To account for both household and social group effects, one would need to define strata based on the intersection of household and social group memberships. With sparse data in most strata, the model cannot reliably estimate separate baseline hazards, leading to unstable or undefined estimates for the vaccine efficacy.

In contrast, frailty models offer a more flexible alternative by introducing random effects for each cluster type. For individual i, the hazard at time t could be written as:

$$\lambda_i(t) = \lambda_0(t) \cdot \exp\left(\beta \cdot \mathbf{1}\left\{\text{i-th subject was vaccinated}\right\} + b_{\text{HH}_i} + b_{\text{SG}_i}\right) \tag{3}$$

where $b_{\text{HH}_i} \sim N(0, \sigma^2_{\text{HH}})$, $b_{\text{SG}_i} \sim N(0, \sigma^2_{\text{SG}})$. Frailty models allow each cluster to contribute independently to an individual's hazard via multiplicative random effects, and it is more suitable to capture susceptibility heterogeneity in complex, overlapping structures.

5.3 Simulation

In order to evaluate the performance of frailty models under overlapping cluster structures, we conducted a simulation study that mirrors the dual cluster setting described above. Suppose there is a community of 2000 people, each person belongs to a household of size between 2 and 6, and a social group of size between 5 and 30.

Infection dynamics was simulated using an SEIR framework (Susceptible Exposed Infectious Recovered). We assumed a latent period of 3 days and an infectious period of 5 days. Each day, individuals interact with all members of their own household and social group. If a contact occurs with an infectious individual, per-contact transmissibility is 0.3 within households and 0.1 within social groups. At the start of the simulation, 10 individuals are randomly seeded as infected. Additionally, to reflect external risk, all uninfected individuals have a 0.1% chance of becoming infected each day due to unknown sources outside the defined clusters.

The population was followed over a 60-day period. From the entire community, a subset of individuals was randomly sampled to represent participants in a simulated clinical trial. These participants were then randomly assigned in a 1:1 ratio to either a vaccine group or a control group. We assumed a leaky vaccine with a true efficacy of 60%, meaning that vaccination reduced the per-contact transmissibility of infection by 60% among vaccinated

individuals. Consequently, the per-contact transmissibility within households decreased from 0.30 to 0.12, and that within social groups decreased from 0.10 to 0.04 for vaccinated individuals.

Due to time constraints, we conducted 100 simulation runs. Vaccine efficacy (VE) was estimated as $1-\beta$, where β denotes the coefficient associated with the vaccine group indicator in the fitted model. Statistical power was calculated as the proportion of simulation replicates in which the p-value was less than 0.05 for $H_0 : \beta = 0$.

Figure 3 displays the estimated vaccine efficacy (VE) across different analysis methods and varying trial sample sizes. When the sample size is small, stratified Cox models perform poorly due to the limited number of participants within each cluster, leading to unstable estimates. As the sample size increases, both the regular Cox model and stratified Cox models exhibit downward bias in VE estimation, whereas the frailty model demonstrates less bias. This observation is consistent with the underlying data-generating mechanism, where stratification by either household or workspace accounts for only partial heterogeneity in susceptibility. Figure 4 shows the power under different analysis methods and trial sample sizes. Regular Cox model and frailty model have higher power compared with stratified Cox models.

5.4 Sensitivity Analysis

We then examined the impact of different sampling strategies on vaccine efficacy estimation and statistical power. Specifically, we compared individual-level random sampling with cluster-based sampling approaches, including household-level and workplace-level sampling. These strategies were evaluated to assess how the underlying sampling design influences the performance of frailty models.

Figure 5 presents the estimated VE under different sampling strategies and sample sizes. As the sample size increases, the variance of VE estimates decreases across all strategies. However, household-based sampling tends to exhibit an upward bias relative to the other

two strategies. Corresponding power estimates are shown in Figure 6. Cluster-based sampling yields higher power at smaller sample sizes. In particular, workplace-based sampling—which involves larger cluster sizes—achieves higher power than household-based sampling. However, when the sample size is limited, workplace-based sampling may introduce a slight downward bias in VE estimates.

5.5 Discussion and Future Work

In settings characterized by overlapping cluster and relatively small cluster sizes, the frailty model demonstrated superior performance compared to both regular and stratified Cox models. Specifically, our simulations showed that the frailty model consistently yielded lower bias in VE estimates and achieved higher statistical power across a range of sample sizes.

In the second simulation, we found that household-based sampling tends to exhibit an upward bias. One plausible explanation lies in the time-dependent dynamics of infection within small clusters. In another study discussing frailty model, the authors modeled the hazard rate in the unvaccinated people as

$$\lambda_0(t) = c\pi p(t) \tag{4}$$

where c denotes the number of contacts per time, π is the transmission probability to an unvaccinated susceptible host during contact with an infectious person, and $p(t)$ represents the probability that a contact is infectious[19]. The hazard rate in the vaccinated people is $\lambda_1(t) = \theta c\pi p(t) = \theta\lambda_0(t)$, where $1 - \theta$ is the vaccine efficacy.

Importantly, $p(t)$ is time-varying, and are heterogeneous across clusters. At the beginning of the outbreak, $p(t)$ is low due to the few infectious individuals. It then increases as transmission occurs within clusters, and eventually declines as individuals recover. In a tightly connected household, once several members are infected early in the follow-up

period, the remaining members, may be disproportionately composed of individuals in the vaccine group. This depletion of susceptibly is particularly problematic in small clusters, where $p(t)$ will rapidly drop to 0 in this household. Consequently, vaccinated people are likely to be the remaining people, and may exhibit a lower cumulative hazard of infection than $\theta c\pi \int p(t)\ dt$. This phenomenon violates the overlap assumption required for causal inference, leading to an overestimation of vaccine efficacy. This bias is more pronounced in household-based sampling due to the smaller cluster size and higher proportion of vaccinated individuals within the sampled households.

Future work may proceed along several directions. First, a more comprehensive literature review could help justify the use of overlapping cluster structures in the simulation design. Second, our simulations focused only on leaky vaccines; future studies should also include all-or-nothing effects and mixtures of leaky and all-or-nothing mechanisms. Third, the simulation hyperparameters in this study were limited. Future work should examine a broader set of parameter combinations. Fourth, our conclusions were primarily based on empirical simulation results, and further mathematical work is needed to establish properties of the estimators, such as consistency.

Data Availability

The simulation code can be found in https://github.com/BowenZhang2001/Dual-Cluster.

References

[1] M.Elizabeth Halloran, Michael Haber, Jr Longini, Ira M., and Claudio J. Struchiner. Direct and indirect effects in vaccine efficacy and effectiveness. *American Journal of Epidemiology*, 133(4):323–331, February 1991.

[2] Geoffrey P Garnett, Simon Cousens, Timothy B Hallett, Richard Steketee, and Neff Walker. Mathematical models in the evaluation of health programmes. *The Lancet*, 378(9790):515–525, 2011.

[3] M. Elizabeth Halloran and Claudio J. Struchiner. Causal inference in infectious diseases. *Epidemiology*, 6:142–151, 1995.

[4] Claudio J. Struchiner and M. Elizabeth Halloran. Randomization and baseline transmission in vaccine field trials. *Epidemiology and Infection*, 135:181–194, 2007.

[5] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012.

[6] Justin J. O'Hagan, Marc Lipsitch, and Miguel A. Hernán. Estimating the per-exposure effect of infectious disease interventions. *Epidemiology (Cambridge, Mass.)*, 25(1):134–138, January 2014.

[7] Miguel A. Hernán. The hazards of hazard ratios. *Epidemiology*, 21(1):13–15, 2010.

[8] M. Roestenberg, M. McCall, J. Hopman, et al. Protection against a malaria challenge by sporozoite inoculation. *N Engl J Med*, 361:468–477, 2009.

[9] B. Killingley, J. Enstone, R. Booy, et al. Potential role of human challenge studies for investigation of influenza transmission. *Lancet Infect Dis*, 11:879–886, 2011.

[10] C. Harro, S. Chakraborty, A. Feller, et al. Refinement of a human challenge model for evaluation of enterotoxigenic escherichia coli vaccines. *Clin Vaccine Immunol*, 18:1719–1727, 2011.

[11] M. Elizabeth Halloran, Ira M. Longini, and Claudio J. Struchiner. *Design and Analysis of Vaccine Studies*. Springer, New York, 2009.

[12] S. Greenland. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*, 7:498–501, 1996.

[13] M. A. Hernán, S. Hernández-Díaz, and J. M. Robins. A structural approach to selection bias. *Epidemiology*, 15:615–625, 2004.

[14] N. E. Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review / Revue Internationale de Statistique*, 43(1):45–57, 1975.

[15] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

[16] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[17] James M. Robins, Miguel A. Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[18] Rebecca Kahn, Matt Hitchings, Steven Bellan, and Marc Lipsitch. Impact of stochastically generated heterogeneity in hazard rates on individually randomized vaccine efficacy trials. *Clinical trials (London, England)*, 15(2):207–211, April 2018.

[19] M. Elizabeth Halloran, Ira M. Longini Jr, and Claudio J. Struchiner. Estimability and interpretation of vaccine efficacy using frailty mixing models. *American Journal of Epidemiology*, 144(1):83–97, 1996.

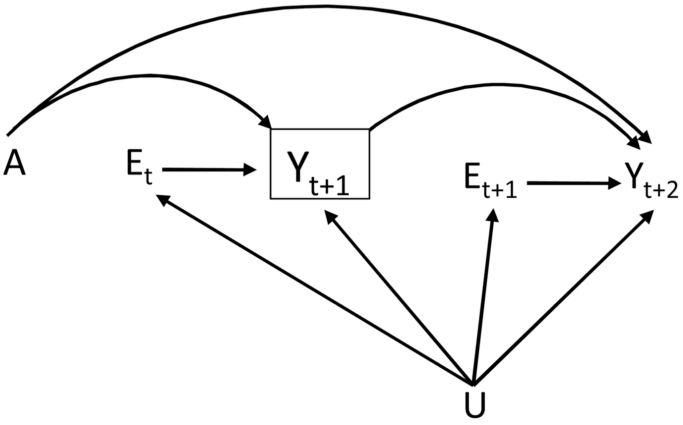

Figure 1: Causal diagram illustrating a double-blind randomized trial of vaccine A and infection outcome Y. E denotes exposure to infection and U represents unmeasured risk factors. Subscripts indicate time periods; only two time points are shown for simplicity.

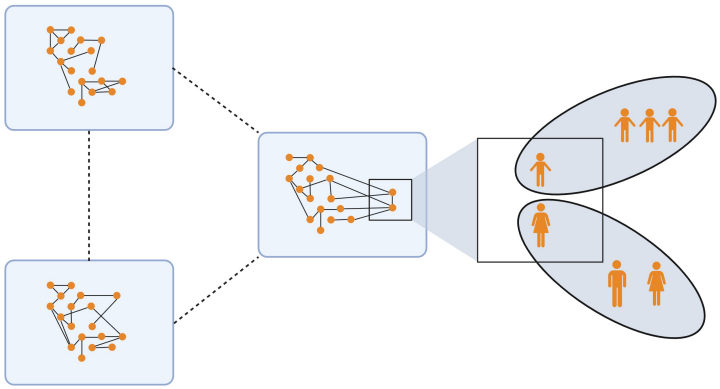

Figure 2: Overlapping cluster structure with household and social group membership.

The left side of the figure illustrates a single-group setting, where each individual belongs to only one community. Within-group connections are dense, while between-group connections are sparse, representing traditional disjoint community structures. The right side depicts our setting, in which each individual is a member of both a household group (represented by the square) and a social group (represented by ellipses). This cross-classified structure can be viewed as a substructure nested within a broader community.

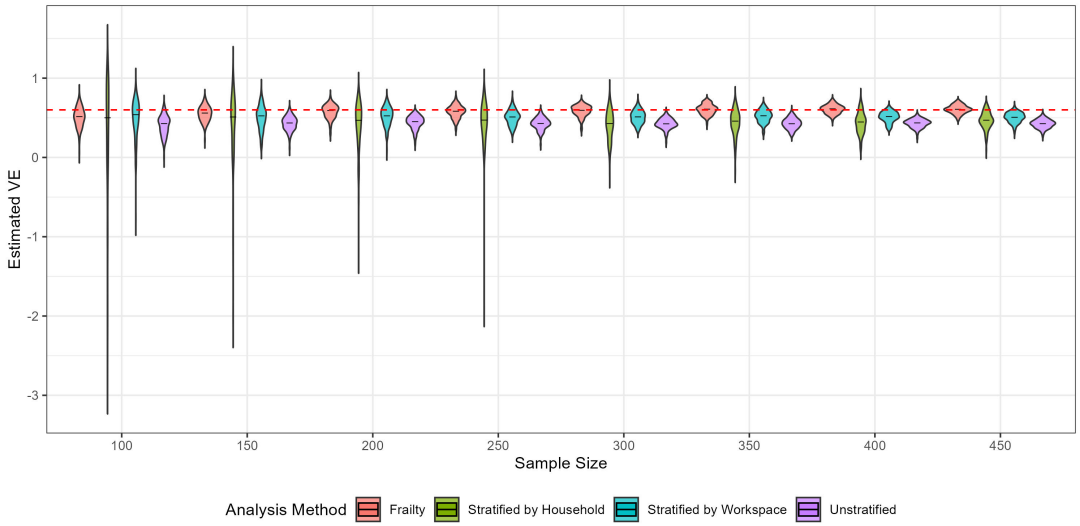

Figure 3: Distribution of estimated vaccine efficacy (VE) across 100 simulation runs under different analysis methods and sample sizes.

Each violin plot represents the distribution of VE estimates for a given sample size and analysis method: frailty model (red), stratified Cox models by household (green) or by workspace (blue), and unstratified Cox model (purple). The red dashed line indicates the true vaccine efficacy of 60%. The black segments represent the median of estimated VE.

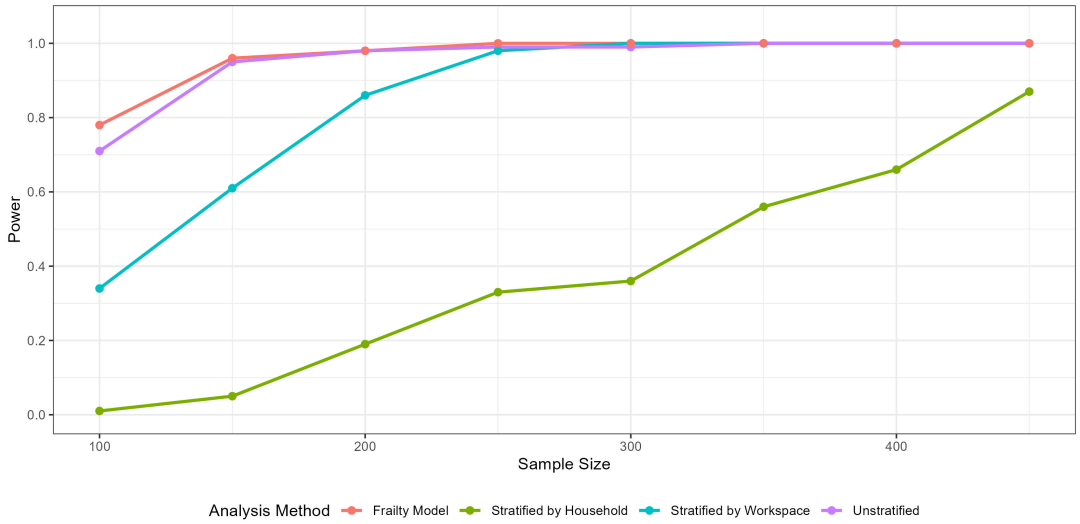

Figure 4: Statistical power across varying sample sizes for different analysis methods based on 100 simulation runs.

Power was defined as the proportion of simulation replicates in which the p-value was less than 0.05 for $H_0 : \beta = 0$. The frailty model (red), stratified Cox models by household (green) and by workspace (blue), and the unstratified Cox model (purple) are compared.

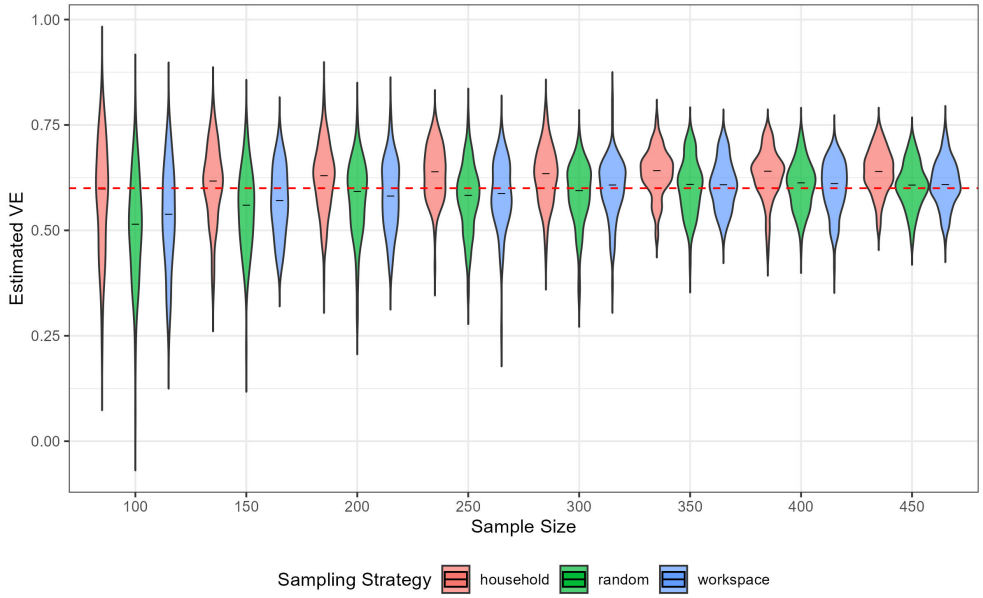

Figure 5: Distribution of estimated vaccine efficacy (VE) across different sample sizes and sampling strategies based on 100 simulation replicates per condition.

Three sampling strategies were compared: random individual sampling (green), household-level cluster sampling (red), and workplace-level cluster sampling (blue). The red dashed line indicates the true vaccine efficacy of 60%. All estimates were obtained using the frailty Cox model.

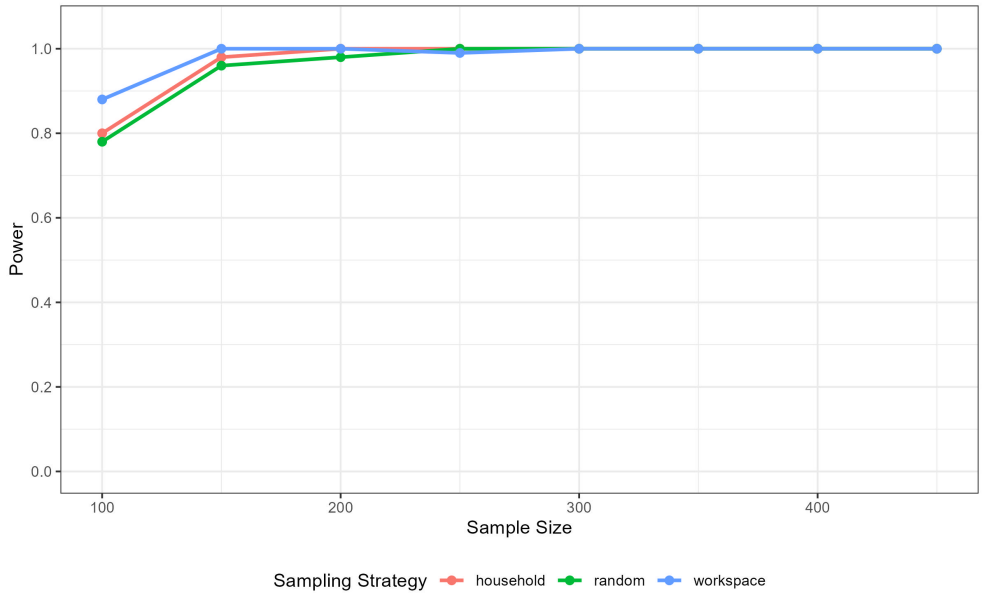

Figure 6: Power estimates under three sampling strategies across varying sample sizes.

Strategies include household-based (red), random individual (green), and workplace-based (blue) sampling. Power was defined as the proportion of simulations where, the p-value was less than 0.05 for $H_0 : \beta = 0$, based on 100 replicates per condition. All estimates were obtained using the frailty Cox model.