HHS Public Access

Author manuscript

J R Stat Soc Ser C Appl Stat. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2022 January ; 71(1): 70–90. doi:10.1111/rssc.12522.

Using social contact data to improve the overall effect estimate of a cluster-randomized influenza vaccination program in Senegal

Gail E. Potter,
National Institute of Allergy and Infectious Diseases, National Institutes of Health, and the Emmes Company, Rockville Maryland, USA

Nicole Bohme Carnegie,
Montana State University, Bozeman Montana, USA

Jonathan D. Sugimoto,
University of Washington and Epidemiologic Research and Information Center, Veterans Affairs Puget Sound Health Care System and Fred Hutchinson Cancer Research Center, Seattle Washington, USA

Aldiouma Diallo,
Institut de Recherche pour le Développement, Niakhar Senegal

John C. Victor,
PATH, Seattle Washington, USA

Kathleen M. Neuzil,
University of Maryland School of Medicine, Baltimore Maryland, USA

M. Elizabeth Halloran
University of Washington Department of Biostatistics and Fred Hutchinson Cancer Research Center, Seattle Washington, USA

Summary.

This study estimates the overall effect of two influenza vaccination programs consecutively administered in a cluster-randomized trial in western Senegal over the course of two influenza seasons from 2009–2011. We apply cutting-edge methodology combining social contact data with infection data to reduce bias in estimation arising from contamination between clusters. Our time-varying estimates reveal a reduction in seasonal influenza from the intervention and a nonsignificant increase in H1N1 pandemic influenza. We estimate an additive change in overall cumulative incidence (which was 6.13% in the control arm) of −0.68 percentage points during Year 1 of the study (95% CI: −2.53, 1.18). When H1N1 pandemic infections were excluded from analysis, the estimated change was −1.45 percentage points and was significant (95% CI, −2.81, −0.08). Because cross-cluster contamination was low (0–3% of contacts for most villages), an estimator assuming no contamination was only slightly attenuated (−0.65 percentage points). These findings are encouraging for studies carefully designed to minimize spillover. Further work is needed to estimate contamination – and its effect on estimation – in a variety of settings.

Keywords

additive hazards; cluster-randomized; contamination; interference; overall effect; social network; spillover

1. Background

Influenza is a seasonal respiratory infection that causes a substantial global burden of morbidity and mortality, particularly among children. One meta-analysis estimated that in 2018 the global burden of influenza among children under 5 was 109.5 million influenza episodes, 870,000 hospital admissions for influenza virus-associated acute lower respiratory infection, and between 13,200 and 97,200 deaths (Wang et al., 2020). In this paper, we use novel methodology to estimate the overall effect of annual influenza vaccination of children age 6 months to 10 years —relative to polio vaccination—on the incidence of influenza in western Senegal.

The study that produced the data analyzed in this paper was a cluster-randomized trial of 20 villages in the Niakhar Demographic Surveillance System (DSS) zone. Villages were assigned to vaccination of children with either inactivated trivalent influenza vaccine or an inactivated polio vaccine as an active control. There is no national recommendation for routine influenza vaccination in Senegal, hence off-study vaccination was expected to be minimal. The trivalent influenza vaccine has been shown to be efficacious in reducing influenza infection in children in other settings (Madhi et al., 2014; Zimmerman et al., 2016); the Niakhar study was testing the effectiveness of widespread immunization of children to reduce the community burden of influenza. The primary analysis for this trial analyzed the *total effect* of the intervention (Diallo et al., 2019). The *total effect* is based on comparing outcomes of treated people in treated villages to those of untreated people in control villages and accounts for protection conferred by receipt of the vaccine as well as from reduction in exposure resulting from vaccination of others in the community. In this paper, we consider the *overall effect* of the intervention. The *overall effect* is based on comparing the average outcome in treated villages to the average outcome in control villages, so takes into account the effect of the community intervention on both treated and untreated people (Halloran et al., 1991). In the primary analysis for this trial, the "total effect" analyzed only children in the age group that was vaccinated, since the total effect combines individual-level direct and indirect effects of vaccination. In this paper, the overall effect is estimated by analyzing all participants. The overall effect quantifies the effectiveness on the entire community (including all age groups) of the TIV vaccination campaign among children.

The total and overall effects are of interest scientifically because of the presence of interference in infectious disease processes. Interference —when one person's treatment can affect another's outcome—is both a boon to disease prevention and a classic inferential problem in infectious disease research. The benefit: the very nature of the process induces dependence between people's outcomes, and treating one person may prevent another's infection. The drawback: observations are no longer independent, and most mainstream

causal inference tools cannot account for the induced dependence. The main approach to dealing with interference is to use cluster-randomized trials (CRT), which allow for dependence within cluster. The assumption of no interference that would be made in a traditional individually-randomized controlled trial is thus weakened to *partial interference*—an assumption of no interference between clusters (Sobel, 2006). Violation of the partial interference condition is referred to as *contamination* (Hudgens and Halloran, 2008).

Typical methods for estimating the overall effect assume partial interference (e.g., Halloran and Struchiner (1991); Liu and Hudgens (2014)). However, for socially contagious outcomes such as infectious diseases, partial interference will not be satisfied if members of treated clusters come in contact with people from untreated clusters (and vice-versa). Recent methodological developments have explored incorporation of measured contamination data into estimation and testing methods to explicitly adjust for interference. See Halloran and Hudgens (2016) and Sävje et al. (2020) for reviews of recent efforts to develop causal inference methods that account for partial interference as well as more general forms of interference. Some of these methods incorporate detailed social network structure (Eckles et al., 2016; Toulis and Kao, 2013; Aronow et al., 2017; Ugander et al., 2013), but such detailed network data is not always available or easy to obtain. In this study, the complete social network was not observed, but information was collected on rates of contacts within and between villages. Most relevant to this data structure and to our interest in the overall effect is a method developed by Carnegie et al. (2016). It is well known that when contamination is present, the overall effect estimate is attenuated. The authors developed a method to explicitly incorporate measured contamination data into the estimation procedure and demonstrated that this adjustment removes the attenuation of the overall effect estimate. We apply this method to estimate the overall effect accounting for cross-cluster contamination and compare it to the estimate that would be obtained assuming partial interference. The analysis in this paper is post hoc and is the first to estimate the overall effect for this study and to incorporate the contact data into estimating the effectiveness of the intervention. To contextualize our findings, we also perform a simulation study demonstrating the performance of this method for a range of contamination values.

This paper continues as follows. In Section 2, we describe the data; in Section 3 we describe the causal model and data preparation. The results of causal effect estimation are given in Section 4. The simulation study parameterization and results are given in Section 4.1, and implications and limitations are discussed in Section 5.

2. Data Collection

The data were collected in a cluster-randomized clinical trial conducted in the Niakhar Demographic Surveillance System (DSS) zone from 2009–2011. Among thirty villages in the Niakhar DSS zone, twenty were selected as clusters for inclusion in the trial and randomized in a 1:1 ratio to receive a blinded vaccination campaign of either inactivated trivalent influenza vaccine (TIV) or inactivated poliovirus vaccine (IPV) as an active control. From here on, villages that received TIV will be referred to as “treated” and those that received IPV as “control”. The same villages were followed for two influenza seasons (2009–2010 and 2010–2011). Different formulations of trivalent influenza vaccine were

given during the two years; the second formulation included the H1N1 2009 “swine” pandemic strain of influenza, but the first formulation did not. A map of the twenty villages analyzed is included in Figure 1. This study, ClinicalTrials.gov NCT00893906, is closed, and the primary results for the trial have been published separately for Year 1 (Diallo et al., 2019) and Year 2 (Niang et al., 2020).

Within each treatment group the goal was to vaccinate up to 5,000 children 6 months to 10 years of age in the following approximate numbers per age-group: 1,270 children 6–35 months of age; 2,835 children 36 months to 8 years of age; and 895 children 9–10 years of age. Vaccinees received age-specific doses. In villages assigned to receive influenza vaccine, 3,906 (78.1% of target number for vaccination) were vaccinated with Dose 1, while 3,843 (76.9% of the target) of those in control villages were vaccinated with IPV. These numbers comprised 66.6% and 66.2% of age-eligible children, respectively.

The primary outcome of the study was laboratory-confirmed symptomatic influenza infection. A combination of active and passive surveillance was used for the primary outcome in the Niakhar DSS zone. In this geographic area, residences are organized in compounds, clusters of dwellings typically housing an extended family. For the twenty villages randomized in the study, field workers visited compounds on a weekly basis to inquire about the occurrence of influenza symptoms. If the person had experienced influenza-like illness in the past 7 days, then the field worker consented them into the surveillance study and documented symptoms and epidemiologic data. Influenza-like illness was defined as follows: (1) among children under 2 years of age, the sudden onset of fever ($>37.5^{\circ}$ C axillary) or subjective (parent-reported) feverishness, plus at least 1 other symptom (cough, sore throat, nasal congestion, rhinorrhea, or difficulty breathing), and (2) among individuals 2 years and older, the sudden onset of fever ($>37.5^{\circ}$C axillary) or subjective (parent- or participant-reported) feverishness, plus either a cough or sore throat. Cases of influenza-like illness were reported to the study center, and nasal and throat swab specimens were collected. In addition, individuals seeking medical care at any of the three Niakhar DSS health posts at any time throughout the year were assessed by health post medical staff or a study physician to determine if the person had influenza-like illness. These individuals were consented into the surveillance study, their symptoms were documented, and nasal and throat swab specimens were obtained for influenza testing.

When individuals with influenza-like illness enrolled into the surveillance study, they also responded to a survey about their travel and social contact patterns during the prior three days. The contact survey defined a “contact” as a conversation occurring between two people in the same location. The contact survey collected numbers of contacts in various locations at two time points (AM and PM) for three consecutive days: the survey day and the two prior days. Numbers of contacts recorded on the survey day are subject to truncation bias because most surveys were administered in the morning and exclude contacts occurring after the time of the survey. Contact patterns for asymptomatic participants are included in the data since some participant’s symptoms began on the day of or the day before the survey. For each day, the respondent provided the number of people she contacted in her own compound in the morning and the afternoon/evening. In addition, she indicated yes or no to whether she had visited a list of locations: another compound (up to five could be

identified in the survey), a market, mosque or church, field, school, sports field or public place, outside the study zone, or another location. For each location visited, the village identification code (and compound identification number, where applicable), the time of day visited (AM, PM, or both), and the number of persons the respondent spoke with during the visit were recorded. For additional details, refer to the example survey form in the Supplementary Material.

Village of residence was recorded during quarterly censuses conducted by the Niakhar DSS (Delaunay et al., 2002, 2013). If participants moved during the trial, their departure date, arrival date, and village of their new residence were recorded. Those who moved a second time had their departure date (but not residence after second move) recorded as well. The cleaning that was performed after receiving the residence data from the DSS is described in the Supplementary Material.

3. Analytic Methods

3.1. Causal effect estimation

In this paper, we consider two estimators for the overall effect of influenza vaccination relative to polio vaccination. The first estimator assumes partial interference (i.e., no contamination), and we refer to it as the *no-contamination estimator*. The second explicitly accounts for interference generated by contacts to villages of the opposite treatment assignment; we refer to this as the *contamination-adjusted estimator*.

To account for contamination, we use the method developed in Carnegie et al. (2016). This approach uses an additive hazards model (Aalen, 1989) for the time to first event but includes a modified treatment variable to account for contacts occurring between clusters in a cluster-randomized trial. Typically, the treatment variable Z is a binary indicator such that $Z = 1$ for participants from treated villages and $Z = 0$ for those from control villages. This is the treatment variable used to calculate the no-contamination estimator. To account for interference between clusters, we use an alternate treatment variable M, which is the proportion of contacts of residents of the participant's village that are with treated villages. It can be thought of as a village-level intensity of exposure to the treatment conditions, and will range from 0 (if all contacts reported in a village are with control villages) to 1 (if all of the contacts reported in a village are with treated villages). Note that if no contamination is present, then this modified treatment variable reduces to the binary treatment variable used to calculate the no-contamination estimator. The additive hazards model used to obtain the no-contamination estimator for an individual in cluster j is

$$\lambda_j(t \mid Z) = \beta_0(t) + \beta_Z(t) z_j,$$

where z_j is a binary treatment indicator for cluster j. The contamination-adjusted estimator is obtained from the following model for individual in cluster j:

$$\lambda_j(t \mid M) = \beta_0(t) + \beta_M(t) m_j,$$

where m_j is the total percentage of contacts of susceptibles in cluster j that are with treated clusters. Note that m_j is a cluster-level variable, but the model is an individual-level model, with individuals in the same cluster taking the same value for m_j.

The coefficient of interest in the additive hazards model—corresponding to the treatment variable—is potentially time-varying. For this reason, we report both that coefficient (visually) and the difference in cumulative hazard of influenza due to the treatment. Because the cumulative hazard is low, this is approximately equal to the difference in cumulative incidence due to treatment. The time-varying coefficients are visualized by displaying the value of their integrals, $\int_0^T \beta_Z(t)dt$ and $\int_0^T \beta_M(t)dt$, as a function of time. These integrals represent the cumulative hazard difference over the time interval [0,t] and are estimated using the nonparametric approach proposed by Aalen (1989). Since the nonparametric estimation approach (based on step functions) produces curves that are not always differentiable, the additive treatment effect is not explicitly estimated, but it is visualized as the slope of the curve (Aalen, 1989). Estimation is implemented with the aalen function in the R package timereg to fit the additive hazards models (Scheike and Zhang, 2011; R Core Team, 2017), and the R code used is provided in the Supplementary Material. The effects are displayed together with confidence intervals based on robust (sandwich) standard errors which take into account the statistical dependency arising from the clustering; these are also provided by the aalen function.

The estimand of interest, which we will denote $\beta(t)$, is the population-averaged difference in hazard of laboratory-confirmed symptomatic influenza infection associated with a change from 0% to 100% exposure to treatment. While $\hat{\beta}_Z(t)$ is a consistent estimator for $\beta(t)$ in the absence of contamination, Carnegie et al. (2016) proved that $\hat{\beta}_M(t)$ is a consistent estimator for $\beta(t)$ in the presence of measured contamination. While the actual hazards experienced by different individuals may differ as an epidemic progresses (due to its stochasticity in time and space), the estimand of interest is averaged over all individuals and clusters. As such, this method does not account for interference localized in time and space (e.g., the increased risk a household member has when another member is infected). Rather, it accounts for interference arising from the contact between people in clusters of the opposite treatment assignments. The common assumption of “partial interference” allows interference within but not between clusters. This method applies to situations where that assumption is violated by adjusting for the between-cluster interference; it does not adjust for within-cluster interference.

This additive hazards model for interference has a natural correspondence to a compartmental epidemic model such as an SIR model (Susceptible-Infectious-Recovered; see, e.g., Keeling and Rohani (2008)). This relationship results from the assumption of the compartmental model that the transmission rate is a product of the contact rate and the per-contact transmission probability. We provide further details on this relationship in the Supplementary Material. This correspondence supports application of our method to influenza, which is frequently modelled with an SIR or SEIR (Susceptible-Exposed-Infectious-Recovered) model (Coburn et al., 2009). Since the length of the exposure state is

irrelevant to modelling disease-free survival, this method gives identical results under SIR and SEIR assumptions (Carnegie et al., 2016).

Although the hazards are permitted to be time-varying, the model assumes identical hazards for different individuals (with the same attributes) at a given time point. As such, the model does not take into account the differences in individual hazards due to different numbers of infections among neighbors at that time point. Survival analysis models applied in influenza vaccine trials typically make this assumption (Ainslie et al., 2019). We perform a simulation study, described in Section 4.1, showing that this estimation approach performs well in estimating the overall effectiveness even when the true process is stochastic.

While Cox regression is frequently used for survival analysis, the Cox proportional hazards model does not share this natural correspondence to epidemic compartmental models. Another advantage that the additive hazards model has over the proportional hazards model is collapsibility, which implies that the treatment effect is the causal effect of interest whether or not covariates are included in the model. A drawback of the additive instead of proportional hazards model is that the estimated hazard, or the lower limit of its confidence interval, is not mathematically restricted to be nonnegative. However, we did not observe negative hazard estimates or negative lower bounds for the confidence interval in the models that we fit.

Analyses were performed separately for Year 1 and Year 2 of the study. Inputs to the additive hazards model are the time to event (or censoring) for each person, infection status, and the percentage of contacts to treated clusters. Calculation of time-to-event for each survey year is described in detail in the Supplementary Material. One irregularity in data collection is noteworthy: during Year 2 of the study, household surveillance was not performed during a strike of field workers that lasted from Jan 3, 2011 through Feb 18, 2011. This could introduce bias since the rate of reporting infections during household visits (as opposed to health posts) was higher in treated than control villages (87.5% and 83%, respectively). To prevent such bias, we analyze a shorter time interval for Year 2 by censoring observations at the start of the strike. The full Year 2 estimates are included as a secondary analysis.

3.2 Calculation of treatment exposure estimates

The treatment exposure value for village j, denoted m_j, is the proportion of contacts that susceptible people in village j made with people in treated clusters. For control villages, this variable is the percentage of contacts to treated villages (the contamination estimate itself). For treated villages, however, the treatment exposure value is one minus the percentage of contacts to people in control villages (i.e., one minus the contamination estimate).

We analyze a single contact survey per participant and restricted analysis to 3,758 contact surveys that were submitted between August 1, 2009 to February 1, 2010 because this subset had been previously cleaned and analyzed extensively (Potter et al., 2019). Contact surveys were not excluded for those with negative influenza tests. The contact survey was given to a convenience sample: rather than randomly sampling susceptible people, those who reported influenza-like-illness during weekly surveillance visits or at the health post were surveyed

regarding contact patterns. By estimating proportions of contacts to treated clusters in this group, we are assuming that the contact patterns reported by these respondents are similar to those of susceptible people. We analyze only data collected in the morning two days before the survey date. We chose this reporting day because it this time point includes more reports from asymptomatic people. A total of 924 (24.6%) were asymptomatic two days before the survey while only 51 (1.4%) were asymptomatic on the day of the survey. Additionally, a social network analysis of these data found no difference in numbers of contacts recorded the day before the survey vs. two days prior – so there is no evidence that the earlier time point is subject to recall bias (Potter et al., 2019). The survey did not elicit how many of the morning contacts were repeated in the afternoon/evening. We analyze contacts reported in the morning as treatment exposure rates were similar between morning and afternoon contacts (Supplementary Material Table 1). As noted above, the treatment exposure value for village j is denoted m_j, and we will denote our estimator for it by $\hat{m}_j$. The formulas given below for $\hat{m}_j$ are expressed in terms of our survey respondents, but also apply to samples that are taken of susceptible people or are representative of the population of susceptible people.

Our treatment exposure estimates take into account the percentage of contacts reported while the respondent was visiting treated villages (Section 3.2.1) and the percentage of contacts reported in the respondent's own home (compound) that occurred to visitors from treated villages (Section 3.2.2).

3.2.1. Percentage of contacts in treated villages—For each village, we calculate the percentage of contacts reported while respondents from that village were located in treated villages. The denominator is the sum of contacts reported by village residents; the numerator is the sum of those contacts whose reported location was a treated village. Contacts reported to villages that are not in the trial are included in the denominator and are treated the same as contacts to control villages. The numerator includes contacts reported in the respondent's own compound if the respondent was a resident of a treated village. For participants who moved mid-study, the village of residence is the reported village of residence at the time of the contact survey.

We initially calculated treatment exposure rates using reports by asymptomatic people only, assuming that this would be more representative of behavior when uninfected and that the symptomatic people would travel less. We compared these to the estimates based on reports by symptomatic people and (counterintuitively) found that symptomatic reports included slightly higher rates of contacts to clusters of the opposite treatment assignment (Supplementary Material Tables 2 and 3). This is likely because cross-cluster contact rates are fairly low overall and because less data is available for asymptomatic reports, so the small amount of data from asymptomatic respondents includes fewer non-zero counts. Therefore we combined data from both asymptomatic and symptomatic people to estimate the treatment exposure variable more precisely.

3.2.2. Incorporating treatment exposure from visitors to the respondent's compound—The above approach assumes that the location of a contact reported by the respondent indicates the residence of the person contacted. As such it does not account

for visitors to one's compound from a cluster of the opposite treatment assignment, so may underestimate cross-cluster exposure. To incorporate exposure from visitors into the estimate, we will define some notation and first consider the estimates for people living in control clusters. Suppose there are n_j people living in cluster j, and let D_i denote the number of contacts reported by person i who lives in cluster j. Let T_i denote the number of contacts person i has made in a location in a treated cluster. Our estimate for the proportion of contacts in cluster j that susceptibles made to people from treated clusters is

$$\hat{m}_j = \frac{\sum_{i=1}^{n_j} T_i}{\sum_{i=1}^{n_j} D_i}$$

We need to update the numerator to include contacts occurring within the respondent's own compound to visitors from other clusters. We can use estimates reported by these visitors, rather than by respondents in cluster j, to obtain this information. Let $V_{T,j}$ denote the total number of contacts reported by people in any treated cluster during their visits to compounds in cluster j. While these contacts contribute to the denominator in the above estimator, they do not contribute to the numerator (because they occurred within the respondent's assigned cluster), but should. Therefore, when j is a control cluster, our updated estimate incorporating this exposure is:

$$\hat{m}_j = \frac{\sum_{i=1}^{n_j} T_i + V_{T,j}}{\sum_{i=1}^{n_j} D_i} = \frac{\sum_{i=1}^{n_j} T_i}{\sum_{i=1}^{n_j} D_i} + \frac{V_{T,j}}{\sum_{i=1}^{n_j} D_i}$$

The rationale for this adjustment is explained in detail in Potter et al. (2019), and an explanation tailored to this setting is provided in the Supplementary Material.

An analogous update is needed for residents of treated clusters. For these respondents we need to account for visits from members of control clusters. Letting $V_{C,j}$ denote the total number of contacts reported by people in any control cluster during their visits to compounds in cluster j. When j is a treated cluster, our updated estimate incorporating this exposure is:

$$\hat{m}_j = \frac{\sum_{i=1}^{n_j} T_i - V_{C,j}}{\sum_{i=1}^{n_j} D_i} = \frac{\sum_{i=1}^{n_j} T_i}{\sum_{i=1}^{n_j} D_i} - \frac{V_{C,j}}{\sum_{i=1}^{n_j} D_i}$$

3.3. Multiple Imputation for Missing Contact Data

The submitted contact surveys had a large number of missing fields, which, if not modelled appropriately, could create bias in the estimates of cross-cluster exposure. For locations visited outside the home two days before the survey, 24% are missing time of day, 59% are missing the number of people contacted, and 32% do not have a village number recorded. Missing data was slightly less on the day before the survey (55% missing the number of

people contacted and 29% missing the village number), but contamination estimates were similar, generally ranging from 0–3% per village. We used data from two days before due to the higher number of reports from asymptomatic people at that time point (24.6% vs. 1.4%). The survey design elicited at-home contacts differently than those that occurred outside the home: the numbers contacted at home in the morning and in the afternoon/evening were recorded, so village and time point were not collected as separate variables. Furthermore, in 60% of analyzed surveys, the number contacted at home in the morning was missing.

We used multiple imputation, expanding on the procedure used in another analysis of this data set (Potter et al., 2019) to adjust for missing contact data. For outside-home locations, up to four variables may be missing: the response to "Was this location visited?", the time of day (AM or PM) the location was visited, the number of people contacted at that location, and the village where the location is located. The responses to whether the location was visited were imputed based on a log binomial regression model with location type, symptom status, and age category as predictors, stratified on day relative to the survey day. Missing times were imputed by sampling from the distribution of non-missing times for that location type. To impute missing numbers of contacts for non-home locations, we fit a negative binomial distribution to the reported contact numbers, predicting the number contacted by the location, symptom status, time of day, and age category. For at-home contacts, we predicted number contacted based on symptom status, time of day, day relative to survey day, age category, and gender. Missing villages for out-of-home contacts were sampled from the observed distribution of visited villages for the respondent's village of residence, combining data from both survey days. As such we are assuming the data are missing at random; in other words, the predictors in our imputation model are sufficient to explain the distribution of unobserved values (Rubin, 1976).

We created twenty imputed data sets, calculated percentages of contacts to treated clusters for each village in each of these imputed data sets, and combined the percentages using standard rules for combining multiply imputed data (Rubin, 1987).

4. Results

Table 1 displays the treatment exposure estimates for each village enrolled in the trial based on the multiply imputed data. For each village, we display the percentage of contacts reported when the respondent visited treated villages, the estimated percentages of contacts from visitors from villages of the opposite treatment assignment, as well as the overall percent of contacts to treated villages, which was used as a covariate in the contamination-adjusted model. The overall percentages are generally close to zero for control villages and close to 100 for treated villages, with a few exceptions.

Our estimated time-varying treatment effects (both unadjusted and contamination-adjusted) are displayed in Figure 2, Panel A for Year 1 of the study. Since the graph displays the integral of the time-varying coefficients, the slopes of the curves represent the coefficients themselves - the estimated difference in hazard rates between vaccine and control villages at each point in time. Both models indicate that the influenza vaccination program reduces influenza through September. Then it is estimated to be ineffective until February (since no

influenza was circulating), after which the program is associated with an increase in the hazard of influenza for a month. This latter time period coincides with the appearance of the A (H1N1) (2009) pandemic strain of influenza (A/H1N1pdm09) in the community, which first appeared in late January 2010. Panel B of Figure 2 displays numbers of infections by influenza type and week. Panel C presents the two estimators excluding cases of A/H1N1pdm09 influenza from the analysis. Its slope represents the instantaneous effect of the influenza vaccination program on the hazard of infection for non-pandemic strains only.

Figure 3 presents results for the second year of the study. As mentioned previously, the formulation of the vaccine provided during this year included the A/H1N1pdm09 strain, unlike the formulation provided in Year 1. Panel B of the graph shows that substantially fewer infections were detected this year. We expect reports to be lower during the strike (Jan. 1 - Feb. 18, 2011) since household surveillance was not conducted during that time, but frequencies prior to the strike were also much lower than in Year 1. Figure 3 shows that after a delay of approximately two months with little effect, the two estimators both indicate that influenza vaccination reduced incidence in Year 2. The delay is likely due to the relative sparsity of cases in the first weeks of the year. The start of the strike mentioned in Section 3 is shown as a vertical line.

Table 3 displays the estimated difference in cumulative hazard of lab-confirmed symptomatic influenza infection due to the influenza vaccination program. These are simply the values of the curves in Figures 2 and 3 for the last day of follow-up, and the confidence intervals correspond to those in the figures. Because the cumulative hazard is low, the difference in cumulative hazard is approximately equal to the difference in cumulative incidence due to treatment.

The overall incidence rates are displayed in Table 2 for comparison purposes. Since the overall incidence in the control group was 6.13%, our estimated additive effect of −0.68% indicates the vaccination program prevented about 11% of influenza infections.

Our two estimators and confidence intervals are similar, but the no-contamination estimators are slightly attenuated because they assume no mixing between clusters of opposite treatment assignments. The confidence intervals for the contamination-adjusted estimator are slightly wider, reflecting the loss of information caused by contamination, but again, are similar. For Year 1 both effects are not statistically significant when all infections are included but achieve significance (barely) when A/H1N1pdm09 infections are excluded. The Year 2 estimates are statistically significant. The Year 2 estimates are interpreted differently as they cover different time intervals; a higher difference in cumulative incidence is expected for the longer interval if vaccine performance stays the same. While bias from the strike starting Jan 1, 2011 does not impact the Year 2 estimate censored at that date, the uncensored one could be biased. The rates of reporting infections during health post visits (as opposed to household visits) were 12.5% in treated villages and 17% in control villages, so the vaccine effect could be overestimated by including a time interval with only health post visits. Because the rates are similar, and because the strike lasted 49 days of a 320-day follow-up period, the bias is likely low.

4.1. Simulation study

We also perform a simulation study to demonstrate the potential impact of using the contamination-adjusted estimator in settings with higher rates of contamination across communities. The simulation study is similar to the one conducted in Carnegie et al. (2016), with some adaptations to reflect the influenza setting of interest in this paper. We simulate interacting pairs of clusters with the percentage of contacts of members of the treated cluster that are with the untreated cluster fixed at values ranging from 5 to 30 percent. Each simulation has ten pairs of clusters, for a total of 20. Influenza spreads over the network following a stochastic SEIR model parameterized based on the model described in Chao et al. (2010), with infected individuals spending an average of 2 days in the exposed phase and 5 in the infectious phase, and average infectiousness per face-to-face contact calculated to produce an R_0 of 2.4 in an unvaccinated population. Each village in each simulation has a small random pertubation added to the infectiousness parameter to encourage stochastic variation across clusters and simulations. Individuals have an average of 16.5 face-to-face contacts per day because a previous analysis of the network data collected in this study found a lower bound of 16.5 face-to-face contacts per day for asymptomatic individuals and 15 for symptomatic individuals (Potter et al., 2019). For simplicity, we assume that all individuals in treated clusters were vaccinated and that vaccination reduces infectiousness by 50%. Although this set-up is simpler than our trial's design (which includes both vaccinated and unvaccinated individuals in treated clusters), and simplifies the vaccination effect (which typically also impacts susceptibility), it is a straightforward way to induce a reduction in cumulative incidence and demonstrate the properties of this method.

The outcome measure is the cumulative incidence of influenza after 60 days. The contamination-adjusted estimator attempts to recover the difference in cumulative incidence that would be observed between the two arms of the study if there were truly no contamination across clusters. The "true value" for this estimand is found via simulation, based on 2000 replications of the epidemic process in a population with two fully distinct clusters. This gave a reduction in cumulative incidence of 8.3 percentage points in the treated cluster relative to the control. For simulations with interacting clusters, we perform 250 replications of the epidemic process and compute both the contamination-adjusted estimator and the no-contamination estimator (which wrongly assumes there was no contamination between clusters).

The simulation study results in Figure 4 show that the no-contamination estimator is steadily more attenuated relative to the difference in cumulative incidence as the rate of contamination increases. The contamination-adjusted estimator, on the other hand, recovers the true difference in cumulative incidence in the absence of contamination well, though that estimate becomes increasingly variable as the rate of contamination increases. In Panel B of the figure, we see that the root mean squared error (RMSE) of the contamination-adjusted estimator is still consistently lower than observed for the no-contamination estimator, indicating that the bias reduction outweighs the increased variability when assessing accuracy of the estimate.

5. Discussion

We have applied novel statistical methodology to estimate the overall effect of a trivalent influenza vaccine program in Niakhar, Senegal. This method incorporates social contact data together with treatment and infection data to reduce the bias in this estimate caused by interference between clusters. Ours is the first study we know of applying this novel method to contact and infection data collected jointly in a clinical trial setting. We produce the first estimates of contact rates between clusters of opposite treatment assignments for this trial and the first, to our knowledge, in Senegal. Our results provide insight into the extent to which the standard assumption of partial interference is violated in a trial of this structure and of the impact of this violation on estimates.

Our time-varying effect estimates show that in Year 1 of the study, the treatment program – vaccination of children – reduced lab-confirmed symptomatic infection with seasonal influenza in the community. Our estimates found the treatment program to be associated with a small (though statistically insignificant) increase in infections with A/H1N1pdm09 influenza. While other studies have found evidence for this relationship (Cowling et al., 2010; Skowronski et al., 2010), others have found evidence that trivalent influenza vaccination protects against A/H1N1pdm09 infection. A meta-analysis of 17 studies, including the two just mentioned, found that the overall evidence points to a protective effect, but the authors cautioned against drawing a solid conclusion because most of the studies reviewed were observational (Yin et al., 2012). Two subsequent randomized trials also found evidence for a protective effect (Cowling et al., 2012; Mcbride et al., 2016).

The extent of contamination measured in our data resulted in little difference between the cumulative incidence for the estimator adjusting for contamination and the one assuming no contamination. The latter was smaller because, as has been found in other studies, contacts to members of clusters of the opposite assignment attenuate the estimate of the overall effect from what it would have been with no contamination (Carnegie et al., 2016; Tiono et al., 2013; Wang et al., 2014). The model we implement explicitly adjusts for contamination, correcting this under-estimation. In addition, the standard errors associated with this adjusted estimator were larger than those for the no-contamination estimator because information available to estimate the effect of the treatment program decreases as mixing increases – so these intervals accurately reflect the decrease in information from zero mixing to the small level of mixing we observed. Our simulation study shows a stronger difference between the two methods when contamination is higher. For example, when the true difference in cumulative incidence is 8.3 percentage points, if contamination is 15%, the contamination-adjusted estimator removes about four percentage points of bias caused by assuming no contamination. The impact on bias and variance is illustrated by our simulation study, which shows the expected attentuation of the effect estimate for values of contamination ranging from zero to 35%. As noted in Carnegie et al. (2016), the approach we have used to estimate the overall effect fails when 50% of contacts occur to clusters of the opposite treatment assignment. This is because our method uses the contact rates between clusters to differentiate treatment status, so no information distinguishing clusters is available for our approach when mixing is at 50%.

The level of contamination in the data was fairly small: the percent of contacts to clusters of the opposite treatment was between 0% and 3% for most villages, although there were some outliers, with 14% being the largest observed value. To our knowledge, these are the first data-based contamination estimates of this type for Senegal. Our finding that this amount of contamination has a negligible impact on the effect estimate may be encouraging for researchers who carefully define cluster selection to minimize contamination, as was done in this study. As contact and travel patterns can vary substantially between cultures and contexts, our estimates may not generalize to other geographic areas, so further measurement of contamination is recommended. Figure 1 shows little separation between the villages in this trial, but they were separated by physical boundaries such as bodies of water and roads, and their definition as cultural/political entities also has an impact on social contact behavior.

Our study has several limitations. First, convenience sampling was used in collecting contact and travel data. Instead of random sampling, participants with ILI were surveyed during household surveillance visits, and their responses were used to estimate the percentage of contacts that susceptible individuals made to treated clusters. Information on contact patterns prior to symptom onset suggest that contact patterns while symptomatic vs. asymptomatic do not differ substantially. However, in future surveys, random sampling of susceptible individuals is recommended to ensure a representative sample.

Second, the extent of missing data in the contact survey is substantial. As noted previously, for locations visited outside the home two days before the survey, 24% are missing time of day, 59% are missing the number of people contacted, and 32% do not have a village number recorded. We used multiple imputation to adjust for missing data. Simulations have shown that multiple imputation can yield unbiased results even when the proportion of missing data is as high as 90%, as long as the imputation model is correctly specified and the data are Missing At Random (MAR) (Madley-Dowd et al., 2019). However, bias is still a risk if these conditions do not hold. For example, if numbers of people contacted in villages of the opposite assignment were higher for participants who did not respond to this question than for those who responded (and who have similar values for covariates included the multiple imputation model), then the true contamination values may be higher than our predicted values. This would mean that the magnitude of the true overall effect is larger than our estimate. If, on the other hand, we have overestimated contamination, then the true effect may be closer to our no-contamination estimate (closer to −0.65 than −0.68). Implementation of similar surveys in the future may be improved by a diary-based approach, in which participants fill out a paper diary as they go about their day (Mossong et al., 2008; Béraud et al., 2015; Melegaro et al., 2017; Johnstone-Robertson et al., 2011; Horby et al., 2011; Fu et al., 2012; Read et al., 2014). In addition we would recommend consideration of procedures employed by Kiti et al. (2014), including conducting a pilot study, providing wristwatches with pre-programmed alarms to remind participants to fill out their diary, and by assigning “shadow” respondents to fill out the diary for illiterate participants. Alternately and potentially more accurate would be an approach using remote wireless sensors to detect when two participants are located within 1.5 meters of each other - a distance at which infection may be transmitted (Kiti et al., 2016; Fournet and Barrat, 2014; Barclay et al., 2014; Génois et al., 2015). While the latter may be prohibitively expensive at the scale of

this study, it could be employed for studies with smaller sized clusters (e.g. households or compounds).

A second limitation of the contact survey is that contacts were reported separately for morning and afternoon time intervals without recording the extent of overlap. Because morning and afternoon contamination estimates were similar, either is likely a reasonable approximation to the percent of contacts to clusters of the opposite assignment during a full day. However, it would be preferable to record numbers of contacts throughout the entire day in future studies. We also note that contacts recorded on the day of the survey did not contribute to analysis since truncation bias arose from the fact that most surveys were conducted in the morning. A diary-based approach would avoid this problem, or if interviews are conducted, they should focus on days before the survey day. The literacy level of the population of interest should be considered in choosing the optimal approach to collect contact data.

Finally, the type of contacts recorded in our study emphasize transmission via large droplets (in close proximity) rather than by aerosol droplets which have a longer range. While many studies have investigated the importance of fomite transmission, physical contacts, small droplets, and aerosol droplets for transmission, their relative importance is not well understood (Weber and Stilianakis, 2008; Cowling et al., 2013; Teunis et al., 2010; Wei and Li, 2016; Kutter et al., 2018). Although the contact survey had limitations, it seems unlikely that the true contamination levels are higher enough than our estimated ones to substantially impact the efficacy estimates. Therefore we believe that our conclusion that contamination was low and had only a small impact on efficacy estimates is valid. However, careful design of the contact survey would improve data precision if a similar approach is applied when clusters are smaller and closer. We would recommend such studies as future research. For example, a compound-based randomization scheme had been considered for this trial design instead of village-based, and in fact, the protocol allowed for both possibilities. The level of contamination for such a design, which would likely be higher than that for villages, could be estimated with our social network data in order to understand its potential impact on estimation. Although our method adjusts for the contamination, higher contamination decreases the information available to detect an effect. Since our approach removes the dilution from the effect estimate while simultaneously increasing standard errors, the lost power from contamination is not regained via our adjustment. Rather, the estimate and standard error estimates are both more accurate than unadjusted estimates. We expect this relationship to hold for other adjustment approaches which have been proposed but, to our knowledge, not yet applied or tested (e.g., Reiner Jr. et al. (2016)).

We also recommend collection and estimation of cross-cluster contamination for different types of contacts (e.g., physical contacts, sexual contacts), for various definitions of clusters in various settings. These estimates can be used to inform future trial designs, choose whether the method we have applied would be better than one which does not adjust for contamination, and ultimately improve the accuracy of vaccine effectiveness and standard error estimates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to all the families who participated in this trial and to the full research teams at Institut de Recherche pour le Développement and Institut Pasteur de Dakar in Senegal. We appreciate helpful comments from two anonymous reviewers on this manuscript. We would like to acknowledge funding received from NIH/NIAID R01 AI085073 (PI Michael Hudgens), NIH/NIAID R37 AI032042 (PI M. Elizabeth Halloran)), NIH/NIGMS U01-GM070749 (PI M. Elizabeth Halloran), and 1U01IP000174 (CoAg between CDC and PATH, PI John Victor) funded by National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention.

References

Aalen OO, 1989. A linear regression model for the analysis of life times. Statistics in Medicine 8, 907–925. [PubMed: 2678347]

Ainslie KE, Haber M, Orenstein WA, 2019. Challenges in estimating influenza vaccine effectiveness. Expert review of vaccines 18 (6), 615–628. [PubMed: 31116070]

Aronow PM, Samii C, et al. , 2017. Estimating average causal effects under general interference, with application to a social network experiment. The Annals of Applied Statistics 11 (4), 1912–1947.

Barclay VC, Smieszek T, He J, Cao G, Rainey JJ, Gao H, Uzicanin A, Salathé M, 2014. Positive network assortativity of influenza vaccination at a high school: implications for outbreak risk and herd immunity. PLoS One 9 (2), e87042. [PubMed: 24505274]

Béraud G, Kazmercziak S, Beutels P, Levy-Bruhl D, Lenne X, Mielcarek N, Yazdanpanah Y, Boëlle P-Y, Hens N, Dervaux B, 2015. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. PLoS One 10 (7), e0133203. [PubMed: 26176549]

Carnegie NB, Wang R, De Gruttola V, 2016. Estimation of the overall treatment effect in the presence of interference in cluster-randomized trials of infectious disease prevention. Epidemiologic Methods 5 (1), 57–68.

Chao DL, Halloran ME, Obenchain VJ, Longini Ira M. J., 2010. FluTE, a publicly available stochastic influenza epidemic simulation model. PLoS Computational Biology 6 (1), e1000656. [PubMed: 20126529]

Coburn BJ, Wagner BG, Blower S, 2009. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). BMC medicine 7 (1), 30. [PubMed: 19545404]

Cowling BJ, Ip DK, Fang VJ, Suntarattiwong P, Olsen SJ, Levy J, Uyeki TM, Leung GM, Peiris JM, Chotpitayasunondh T, et al. , 2013. Aerosol transmission is an important mode of influenza A virus spread. Nature communications 4, 1935.

Cowling BJ, Ng S, Ma ES, Fang VJ, So HC, Wai W, Cheng CK, Wong JY, Chan K-H, Ip DK, et al. , 2012. Protective efficacy against pandemic influenza of seasonal influenza vaccination in children in hong kong: a randomized controlled trial. Clinical Infectious Diseases 55 (5), 695–702. [PubMed: 22670050]

Cowling BJ, Ng S, Ma ESK, Cheng CKY, Wai W, Fang VJ, Chan K-H, Ip DKM, Chiu SS, Peiris JSM, Leung GM, 2010. Protective efficacy of seasonal influenza vaccination against seasonal and pandemic influenza virus infection during 2009 in hong kong. Clinical Infectious Diseases 51 (12), 1370–1379. URL 10.1086/657311 [PubMed: 21067351]

Delaunay V, Douillot L, Diallo A, Dione D, Trape J-F, Medianikov O, Raoult D, Sokhna C, 2013. Profile: the Niakhar Health and Demographic Surveillance System. International journal of epidemiology 42 (4), 1002–1011. [PubMed: 24062286]

Delaunay V, Marra A, Levi P, Etard J-F, 2002. Niakhar DSS, Senegal. INDEPTH Network. Populations and health in developing countries 1, 279–285.

Diallo A, Diop OM, Diop D, Niang MN, Sugimoto JD, Ortiz JR, Diarra B, Goudiaby D, Lewis KD, Emery SL, et al. , 2019. Effectiveness of seasonal influenza vaccination in children in Senegal during a year of vaccine mismatch: A cluster-randomized trial. Clinical Infectious Diseases.

Eckles D, Karrer B, Ugander J, 2016. Design and analysis of experiments in networks: Reducing bias from interference. Journal of Causal Inference 5 (1), 2193–3685.

Fournet J, Barrat A, 2014. Contact patterns among high school students. PLoS One 9 (9), e107878. [PubMed: 25226026]

Fu Y. c., Wang D-W, Chuang J-H, 2012. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. PLoS One 7 (10), e45113. [PubMed: 23056193]

Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A, 2015. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science 3 (3), 326–347.

Halloran ME, Haber M, Longini IM Jr, Struchiner CJ, 1991. Direct and indirect effects in vaccine efficacy and effectiveness. American journal of epidemiology 133 (4), 323–331. [PubMed: 1899778]

Halloran ME, Hudgens MG, 2016. Dependent happenings: a recent methodological review. Current epidemiology reports 3 (4), 297–305. [PubMed: 28133589]

Halloran ME, Struchiner CJ, 1991. Study designs for dependent happenings. Epidemiology 2 (5), 331–338. [PubMed: 1742381]

Horby P, Thai PQ, Hens N, Yen NTT, Thoang DD, Linh NM, Huong NT, Alexander N, Edmunds WJ, Duong TN, et al. , 2011. Social contact patterns in vietnam and implications for the control of infectious diseases. PLoS One 6 (2), e16965. [PubMed: 21347264]

Hudgens MG, Halloran ME, 2008. Toward causal inference with interference. Journal of the American Statistical Association 103 (482), 832–842. [PubMed: 19081744]

Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LD, Bekker L-G, Wood R, 2011. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. American journal of epidemiology 174 (11), 1246–1255. [PubMed: 22071585]

Keeling MJ, Rohani P, 2008. Modeling infectious diseases in humans and animals. Princeton University Press.

Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ, 2014. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. PLoS One 9 (8), e104786. [PubMed: 25127257]

Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, Cappa L, Panisson A, Barrat A, Cattuto C, et al. , 2016. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. EPJ data science 5 (1), 21. [PubMed: 27471661]

Kutter JS, Spronken MI, Fraaij PL, Fouchier RA, Herfst S, 2018. Transmission routes of respiratory viruses among humans. Current opinion in virology 28, 142–151. [PubMed: 29452994]

Liu L, Hudgens MG, 2014. Large sample randomization inference of causal effects in the presence of interference. Journal of the American Statistical Association 109 (505), 288–301. [PubMed: 24659836]

Madhi SA, Cutland CL, Kuwanda L, Weinberg A, Hugo A, Jones S, Adrian PV, Van Niekerk N, Treurnicht F, Ortiz JR, et al. , 2014. Influenza vaccination of pregnant women and protection of their infants. New England Journal of Medicine 371 (10), 918–931. [PubMed: 25184864]

Madley-Dowd P, Hughes R, Tilling K, Heron J, 2019. The proportion of missing data should not be used to guide decisions on multiple imputation. Journal of clinical epidemiology 110, 63–73. [PubMed: 30878639]

Mcbride WJ, Abhayaratna WP, Barr I, Booy R, Carapetis J, Carson S, De Looze F, Ellis-Pegler R, Heron L, Karrasch J, et al. , 2016. Efficacy of a trivalent influenza vaccine against seasonal strains and against 2009 pandemic h1n1: A randomized, placebo-controlled trial. Vaccine 34 (41), 4991–4997. [PubMed: 27595443]

Melegaro A, Del Fava E, Poletti P, Merler S, Nyamukapa C, Williams J, Gregson S, Manfredi P, 2017. Social contact structures and time use patterns in the Manicaland Province of Zimbabwe. PLoS One 12 (1), e0170459. [PubMed: 28099479]

Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, et al. , 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine 5 (3), e74. [PubMed: 18366252]

Niang MN, Sugimoto JD, Diallo A, Diarra B, Ortiz JR, Lewis KD, Lafond KE, Halloran ME, Widdowson M-A, Neuzil KM, et al. , 2020. Estimates of inactivated influenza vaccine effectiveness among children in senegal: Results from 2 consecutive cluster-randomized controlled trials in 2010 and 2011. Clinical Infectious Diseases.

Potter GE, Wong J, Sugimoto J, Diallo A, Victor JC, Neuzil K, Halloran ME, 2019. Networks of face-to-face social contacts in Niakhar, Senegal. PLoS One 14 (8), e0220443. [PubMed: 31386686]

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, Guan Y, Jiang CQ, Cummings DA, 2014. Social mixing patterns in rural and urban areas of southern China. Proceedings of the Royal Society of London B: Biological Sciences 281 (1785), 20140268.

Reiner RC Jr., Achee N, Barrera R, Burkot TR, Chadee DD, Devine GJ, Endy T, Gubler D, Hombach J, Kleinschmidt I, et al. , 2016. Quantifying the epidemiological impact of vector control on dengue. PLoS neglected tropical diseases 10 (5), e0004588. [PubMed: 27227829]

Rubin DB, 1976. Inference and missing data. Biometrika 63 (3), 581–592.

Rubin DB, 1987. Multiple Imputation for Nonresponse in Surveys. J. Wiley & Sons, New York.

Sävje F, Aronow PM, Hudgens MG, 2020. Average treatment effects in the presence of unknown interference. Annals of Statistics, in press.

Scheike TH, Zhang M-J, 2011. Analyzing competing risk data using the R timereg package. Journal of Statistical Software 38 (2), 1–15. URL http://www.jstatsoft.org/v38/i02/

Skowronski DM, De Serres G, Crowcroft NS, Janjua NZ, Boulianne N, Hottes TS, Rosella LC, Dickinson JA, Gilca R, Sethi P, et al. , 2010. Association between the 2008–09 seasonal influenza vaccine and pandemic H1N1 illness during spring–summer 2009: four observational studies from Canada. PLoS Medicine 7 (4), e1000258. [PubMed: 20386731]

Sobel ME, 2006. What do randomized studies of housing mobility demonstrate?: causal inference in the face of interference. Journal of the American Statistical Association 101 (476), 1398–1407.

Teunis PF, Brienen N, Kretzschmar ME, 2010. High infectivity and pathogenicity of influenza A virus via aerosol and droplet transmission. Epidemics 2 (4), 215–222. [PubMed: 21352792]

Tiono AB, Ouédraogo A, Ogutu B, Diarra A, Coulibaly S, Gansané A, Sirima SB, O'Neil G, Mukhopadhyay A, Hamed K, 2013. A controlled, parallel, cluster-randomized trial of community-wide screening and treatment of asymptomatic carriers of *Plasmodium falciparum* in Burkina Faso. Malaria Journal 12, 79. [PubMed: 23442748]

Toulis P, Kao E, 2013. Estimation of causal peer influence effects. In: International conference on machine learning. pp. 1489–1497.

Ugander J, Karrer B, Backstrom L, Kleinberg J, 2013. Graph cluster randomization: Network exposure to multiple universes. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 329–337.

Wang R, Goyal R, Lei Q, Essex M, De Gruttola V, 2014. Sample size considerations in the design of cluster randomized trials of combination HIV prevention. Clinical Trials 11, 309–318. [PubMed: 24651566]

Wang X, Li Y, O'Brien KL, Madhi SA, Widdowson M-A, Byass P, Omer SB, Abbas Q, Ali A, Amu A, et al. , 2020. Global burden of respiratory infections associated with seasonal influenza in children under 5 years in 2018: a systematic review and modelling study. The Lancet Global Health 8 (4), e497–e510. [PubMed: 32087815]

Weber TP, Stilianakis NI, 2008. Inactivation of influenza A viruses in the environment and modes of transmission: a critical review. Journal of Infection 57 (5), 361–373. [PubMed: 18848358]

Wei J, Li Y, 2016. Airborne spread of infectious agents in the indoor environment. American Journal of Infection Control 44 (9), S102–S108. [PubMed: 27590694]

Yin JK, Chow MYK, Khandaker G, King C, Richmond P, Heron L, Booy R, 2012. Impacts on influenza A (H1N1) pdm09 infection from cross-protection of seasonal trivalent influenza vaccines and A (H1N1) pdm09 vaccines: systematic review and meta-analyses. Vaccine 30 (21), 3209–3222. [PubMed: 22387221]

Zimmerman RK, Nowalk MP, Chung J, Jackson ML, Jackson LA, Petrie JG, Monto AS, McLean HQ, Belongia EA, Gaglani M, et al. , 2016. 2014–2015 influenza vaccine effectiveness in the United States by vaccine type. Clinical Infectious Diseases, ciw635.

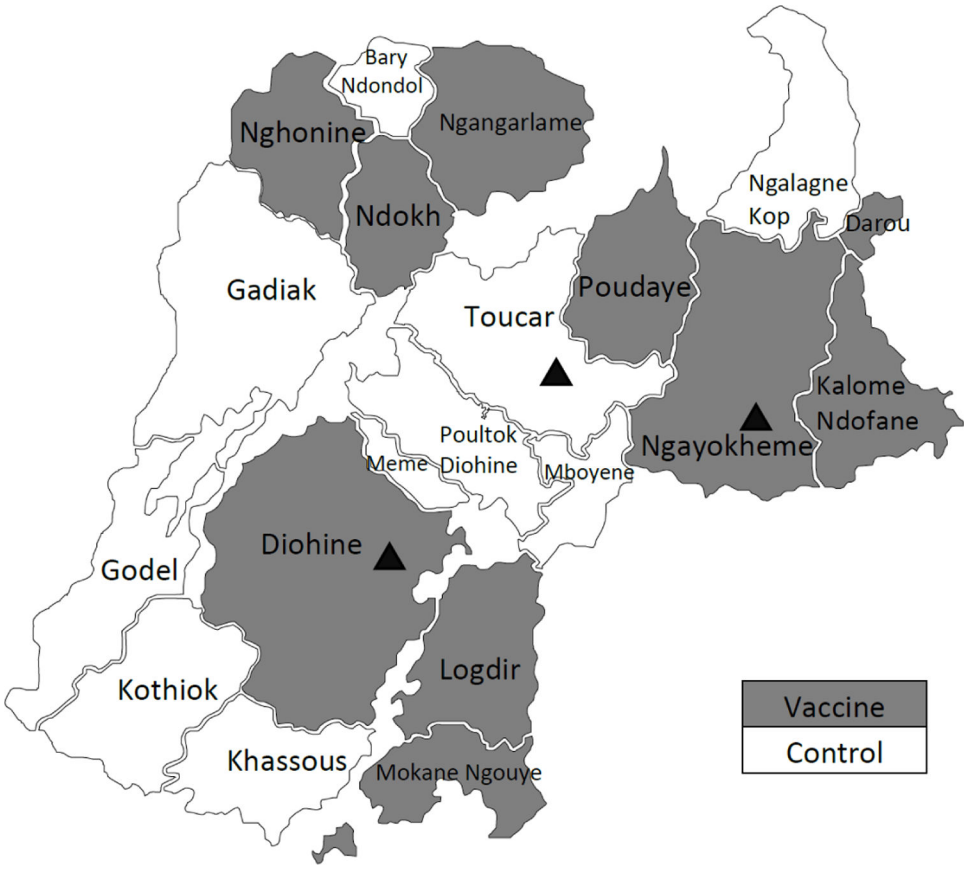

Fig. 1.
Map of the twenty villages included as clusters in the influenza vaccine trial.

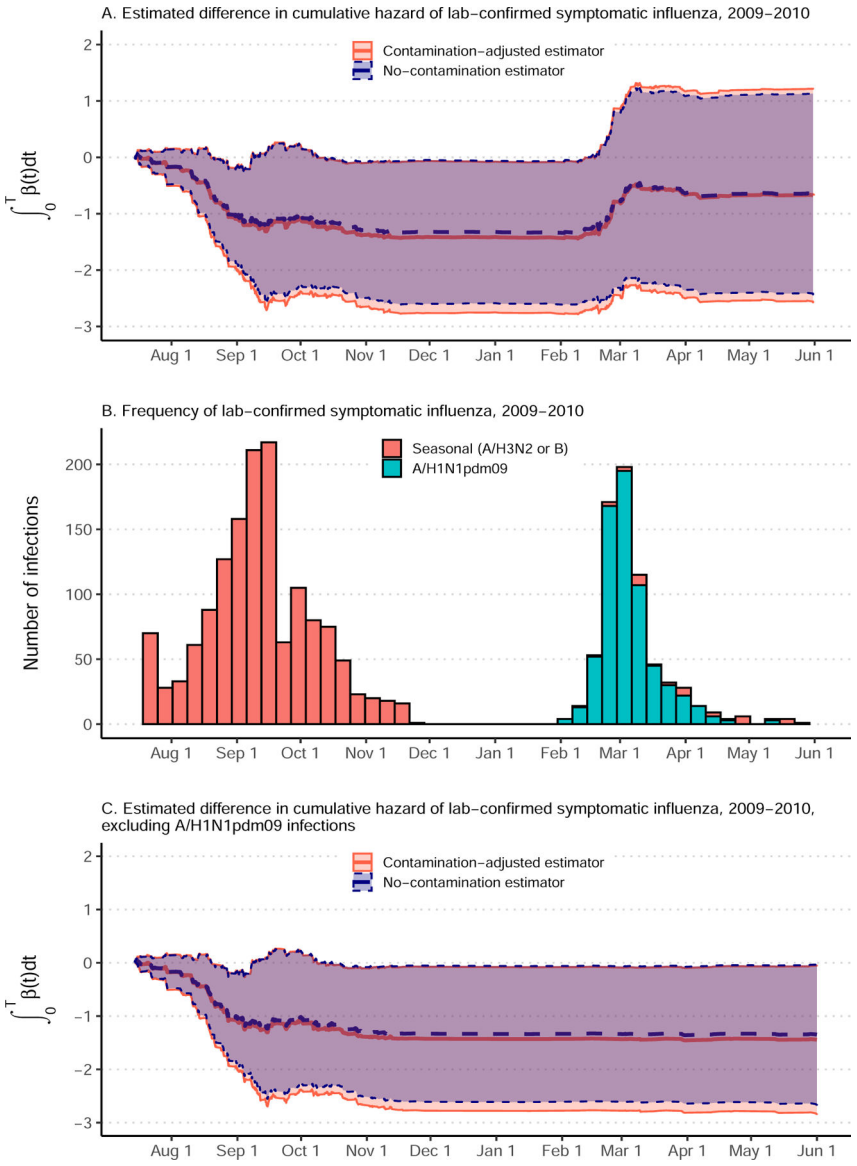

Fig. 2.

Panel A shows the estimated effects of the influenza vaccination program for Year 1 (July 2009 - May 2010) of the study. Shading shows 95% confidence intervals. Panel B shows incidence of influenza infections by time and type. Panel C shows the estimated effects of the influenza vaccination program during Year 1 on symptomatic infection with seasonal influenza strains (A/H3N2 or B).

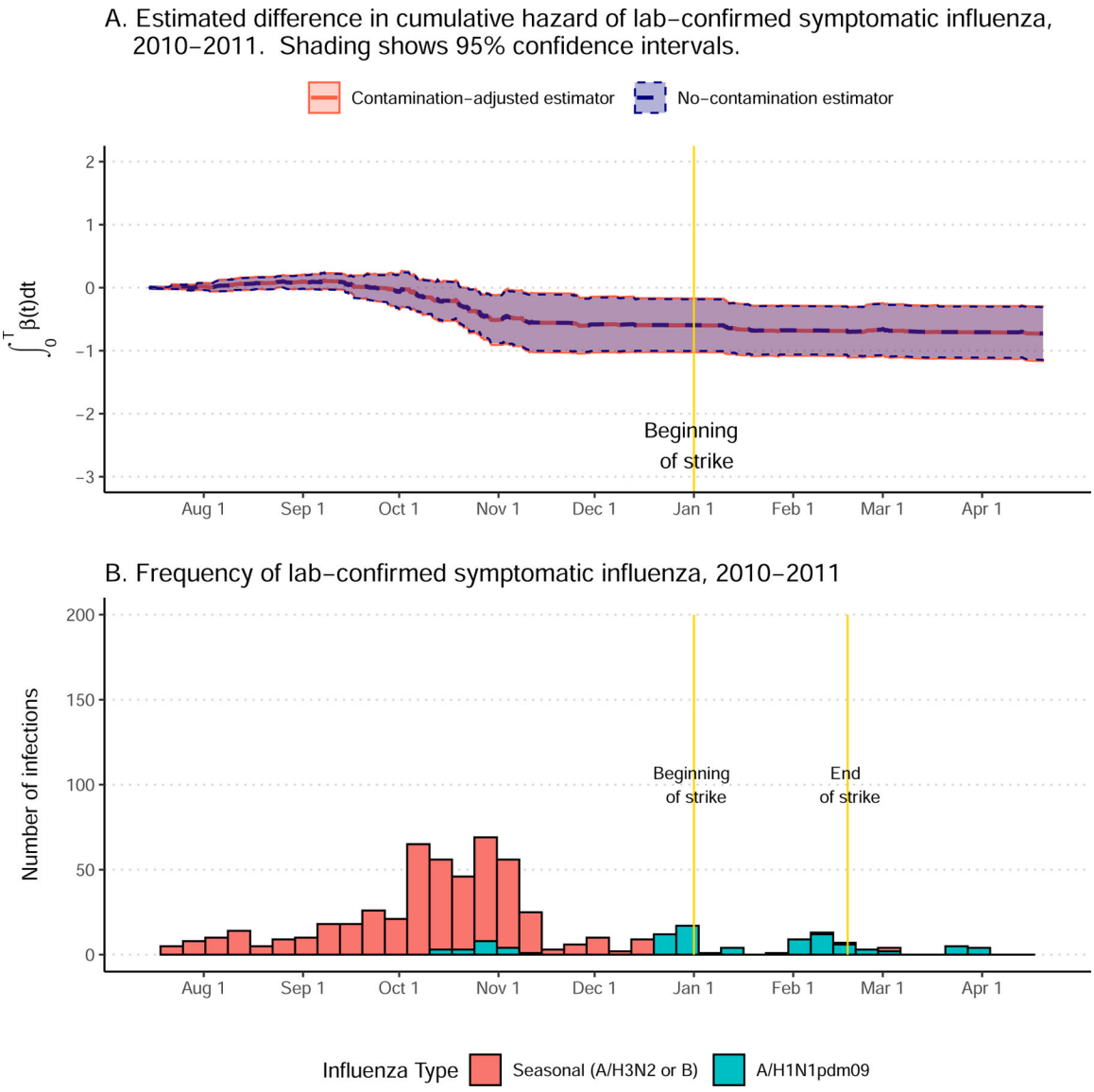

Fig. 3.

Panel A shows the estimated effects of the influenza vaccination program for Year 2 (July 2010 - May 2011) of the study. Panel B shows incidence of influenza infections by time and type.

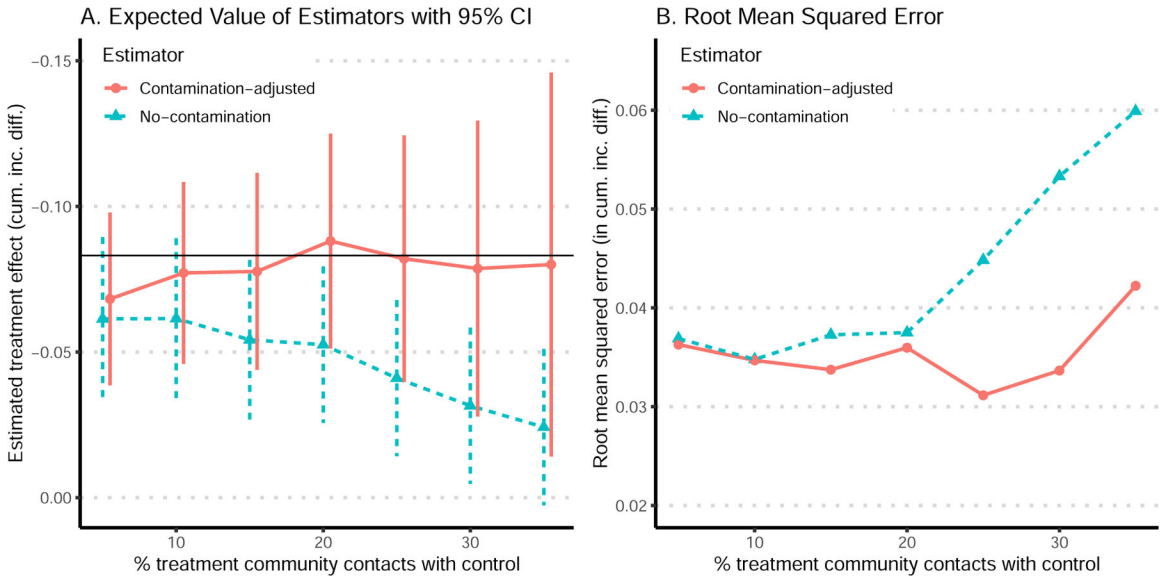

Fig. 4.

Panel A shows the value of the contamination-adjusted (red solid line) and no-contamination (blue dashed line) estimators and associated 95% confidence intervals across values of cross-community contamination. The horizontal line shows the true value of the estimand. Because of the substantial overlap in confidence intervals, the lines are shifted slightly for visibility, but contamination rates were at 5% intervals. Panel B shows the root mean squared error of the estimator (with respect to the true difference in cumulative incidence in the absence of contamination of −0.083).

Table 1:

Percentages of contacts with residents of treated clusters based on (1) contacts reported while located in treated clusters, (2) contacts in the respondent's own compound to visitors from clusters of the opposite treatment assignment, and (3) total percentages of contacts to residents of treated clusters (treatment exposure).

Village	Treatment Assignment	Percent reported in treated clusters $\sum_{i=1}^{n_j} T_i / \sum_{i=1}^{n_j} D_i$	Percent from visitors $V_{C,j} / \sum_{i=1}^{n_j} D_i$	Treatment exposure m_j
Kalome Ndofane	Vaccine	100	0	100
Ngayokheme	Vaccine	99	0	99
Ndokh	Vaccine	99	1	99
Ngangarlame	Vaccine	99	0	99
Diohine	Vaccine	99	0	98
Mokane Ngouye	Vaccine	99	1	98
Nghonine	Vaccine	98	2	96
Logdir	Vaccine	95	2	93
Darou	Vaccine	96	5	90
Poudaye	Vaccine	93	2	90
Ngalagne Kop	Control	0	0	0
Mboyene	Control	0	0	0
Poultok Diohine	Control	0	0	0
Bary Ndondol	Control	0	1	1
Toucar	Control	1	0	1
Gadiak	Control	2	0	2
Godel	Control	2	0	2
Khassous	Control	3	0	3
Kothiok	Control	3	0	3
Meme	Control	14	0	14

Table 2:

Incidence of influenza by treatment group and study year.

Study Year	Treated	Control	All
Year 1, all infections	999/18200 (5.49%)	1076/17550 (6.13%)	2075/35750 (5.8%)
Year 1, excluding A/H1N1pdm09	630/18200 (3.46%)	833/17550 (4.75%)	1463/35750 (4.09%)
Year 2, all infections	224/18547 (1.21%)	341/17815 (1.91%)	565/36362 (1.55%)

Table 3:

Estimated difference in cumulative incidence of influenza (measured in percentage points) due to the influenza vaccination program.

	Contamination-Adjusted		No-Contamination	
Study Year	**Estimate**	**95% C.I.**	**Estimate**	**95% C.I.**
Year 1, all infections	−0.68	[−2.53, 1.18]	−0.65	[−2.40, 1.09]
Year 1, excluding A/H1N1pdm09	−1.45	[−2.81, −0.08]	−1.35	[−2.64, −0.06]
Year 2 (July - Dec 2010)	−0.59	[−1.01, −0.17]	−0.59	[−0.99, −0.19]
Year 2 (July 2010 - May 2011)	−0.73	[−1.16, −0.31]	−0.73	[−1.14, −0.32]