Design and Evaluation of Prophylactic Interventions Using Infectious Disease Incidence Data from Close Contact Groups

Yang Yang

Department of Biostatistics, The Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, U.S.A.

Ira M. Longini, Jr.

Department of Biostatistics, The Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, U.S.A.

M. Elizabeth Halloran

Department of Biostatistics, The Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, U.S.A.

Summary. Prophylaxis of contacts of infectious cases such as household members and treatment of infectious cases are methods to prevent spread of infectious diseases. We develop a method based on maximum likelihood to estimate the efficacy of such interventions and the transmission probabilities. We consider both the design with prospective follow-up of close contact groups and the design with ascertainment of close contact groups by an index case as well as randomization by groups and by individuals. We compare the designs using simulations. We estimate the efficacy of the influenza antiviral agent oseltamivir in reducing susceptibility and infectiousness in two case-ascertained household trials.

Keywords: Antiviral agent; Community trial; Infectious disease; Intervention efficacy; Left truncation

1. Introduction

Transmission of many infectious diseases takes place mainly through close contacts in mixing groups such as households, daycare centers, schools, and the workplace, and to a lesser extent through casual contacts in the community at large. Data from studies based on close contact groups offer a basis for estimating person-to-person and community-to-person transmission probabilities (Longini and Koopman, 1982; O'Neill *et al.*, 2000), and more importantly, for evaluating the effectiveness of prophylactic interventions such as vaccination (Halloran *et al.*, 1997) and antiviral agents. Key study design elements include randomization scheme and ascertainment method. Often in infectious disease studies, the intervention product and placebo are randomized so that all participants within a group receive either the intervention product or placebo. However, individual randomization within a group could improve precision of the parameter estimates (Datta *et al.*, 1999) as well as identifiability of the parameters of interest. In some studies, households are ascertained once a case is identified (Welliver *et al.*, 2001; Hayden *et al.*, 2004), while in the prospective design, contact groups free of disease are enrolled at the beginning of an epidemic season and then followed to some predetermined end point (Monto *et al.*, 2002; Hayden *et al.*, 2000). The case-ascertained trial size could be much smaller than that of a prospective follow-up study with the same number of cases, but at the price of potential bias due to left truncation of the infection status of the non-index cases. These latter cases could already be infected at the time of ascertainment of the close contact group, but not yet showing symptoms. If this left truncation can be dealt with, then the case-ascertained trials may be preferable to the larger prospective trials.

Table 1. Two randomized multi-center trials conducted in North America and Europe in winter seasons for evaluating the efficacy of oseltamivir, an influenza antiviral agent.

	Trial I (Welliver *et al.*, 2001)	Trial II (Hayden *et al.*, 2004)
Time of trial	1998-1999	2000-2001
Households	372	277
Population	1329	1110
Treatment for illness	None	Oseltamivir
Duration of medication		
Illness treatment	N/A	5 days
Prophylaxis	7 days	10 days
Follow up (symptom diary)	14 days	30 days
Infected/Total(index cases)†	165/372	179/298
Infected/Total(susceptible)†		
Control‡	38/464	45/392
Oseltamivir	4/493	14/420

Numbers may slightly differ from references due to different criteria of data inclusion for analysis.
† Numerator is the number of infections confirmed by lab test, and denominator is the total number of subjects in the corresponding group. Some households had > 1 index cases.
‡ Participants in the control group received placebo in trial I and no treatment in trial II (see text for further explanation).

Two randomized, controlled multi-center Phase III efficacy trials of oseltamivir, an orally administered influenza neuraminidase inhibitor, are examples of studies with case-ascertained follow-up and household-level randomization. These two trials were conducted in North America and Europe during the winter influenza seasons of 1998 - 1999 (Welliver *et al.*, 2001)and 2000 - 2001 (Hayden *et al.*, 2004). Details of the trials are given in Table 1.

Many methods are available for analyzing clinical trial data of acute infectious diseases based on close contact groups (Becker, 1989). Longini and Koopman (1982), Longini *et al.*(1988), Addy *et al.*(1991) and Magder and Brookmeyer (1993) developed methods that use only final infection status of individuals within each close contact group. Rampey *et al.*(1992) developed a method for time of onset data for prospective trials, but not for case-ascertained trials.

In this paper, we develop an estimation procedure based on maximum likelihood for both prospective and case-ascertained clinical trials in close contact groups. We develop a method to deal with the left truncation in the case-ascertained study design. Individual- and group-level randomization schemes as well as prospective and case-ascertained designs are compared using simulations. The approaches are generalized to stratified populations that include discrete covariates, allowing for different transmission probabilities. We analyze the data from the two trials of oseltamivir using these methods.

2. Methods

Suppose, without loss of generality, that influenza is the infectious disease of interest and that the close contact groups are households. In addition, the intervention of interest is the prophylactic use of influenza antiviral agents. We define two types of potentially infectious contact: (1) being in the household with another infected person, and (2) making contact with possibly infected people outside of the household. We define p as the transmission probability per daily contact within the household between a susceptible person and an infective person if both have not received antiviral agents. Similarly, define b as the daily probability that a susceptible and untreated person is infected by a source of infection from the community.

Define AVE_S as the antiviral efficacy for susceptibility to infection and illness. Then $\text{AVE}_S = 1 - \theta$, where θp is the reduced transmission probability resulting in illness if the susceptible person is taking an antiviral agent and exposed to an untreated infected person in the household. For simplicity, we assume that efficacy is the same for contacts outside the household, i.e., the reduced transmission probability resulting in illness for a person taking an antiviral agent is θb. With a rich data set, one could estimate two separate

parameters. Define AVE_I as the antiviral efficacy for infectiousness. Then, $\text{AVE}_I = 1 - \phi$, where ϕp is the reduced transmission probability if the infective person is treated. We can also estimate the AVE_T, the total effect on transmission when both people in a transmission pair are treated, similar to the VE_T in Halloran *et al.*(1997), also discussed in Farrington (2003). We consider two types of parameterization for AVE_T. In the first model, we assume independence and multiplicativity of θ and ϕ, so that the transmission probability reduces to $\theta\phi p$ and the total effect on the transmission is $\text{AVE}_T = 1 - \theta\phi$. In the second model, we drop the assumption of independence, and the transmission probability reduces to ψp, where ψ is estimated as a separate parameter. In this case, $\text{AVE}_T = 1 - \psi$. For identifiability, we make the following assumptions about influenza: (1) The latent period (i.e., time from infection to being infectious) is the same as the incubation period (i.e., time from infection to the onset of illness symptoms). (2) The probability distributions of the lengths of the latent and the infectious periods are known and independent of treatments.

2.1. Maximum likelihood estimation

2.1.1. Prospective follow-up study design

We start with the model for the prospective follow-up study. Without loss of generality, let the trial start on day 1 and end on day T. The simplest data for each participant are (1) the first date with symptoms if infected, (2) assigned treatment, and (3) treatment period. Other time-dependent or -invariant information could be included. Let $\tilde{t}_i$ denote the day of illness onset for an infected person i. We let $r_i(t) = (0$ untreated, 1 treated) indicate the treatment status of person i on day t. Under the assumption of independence between θ and ϕ, the probability that a susceptible person i escapes infection by an infective family member j on day t is given by

$$q_{ij}(t) = 1 - \theta^{r_i(t)}\phi^{r_j(t)}pf(t|\tilde{t}_j), \tag{1}$$

where the function $f(t|\tilde{t}_j)$ is the probability that person j is infectious on day t given the day of illness onset $\tilde{t}_j$. When the assumption of independence is dropped, the escape probability is

$$q_{ij}(t) = 1 - \theta^{r_i(t)(1-r_j(t))}\phi^{r_j(t)(1-r_i(t))}\psi^{r_i(t)r_j(t)}pf(t|\tilde{t}_j). \tag{2}$$

It follows that

$$e_i(t) = (1 - \theta^{r_i(t)}b)\prod_{j\in D_i} q_{ij}(t) \tag{3}$$

and

$$Q_i(t) = \prod_{\tau=1}^{t} e_i(\tau) \tag{4}$$

are the escape probabilities for person i on day t and up to day t, respectively, where D_i is the set of people in the same household with person i. The probability that person i is infected on day t is given by

$$Z_i(t) = Q_i(t-1)(1 - e_i(t)).$$

However, we do not observe the exact infection time but just the onset time of illness. Let l_{max} and l_{min} be the maximum and minimum duration of the latent period. Define $\underline{t_i} = \tilde{t}_i - l_{max}$ and $\overline{t_i} = \tilde{t}_i - l_{min}$ as the earliest and latest potential infection day for person i. Further denote the likelihood contributed by person i by $L_i = L_i(b, p, \theta, \phi, \psi|\tilde{t}_j, j \in D_i)$. Then

$$L_i = \begin{cases} Q_i(T), & \text{if individual } i \text{ is not infected,} \\ \sum_{t=\underline{t_i}}^{\overline{t_i}} g(\tilde{t}_i|t)Z_i(t), & \text{otherwise,} \end{cases} \tag{5}$$

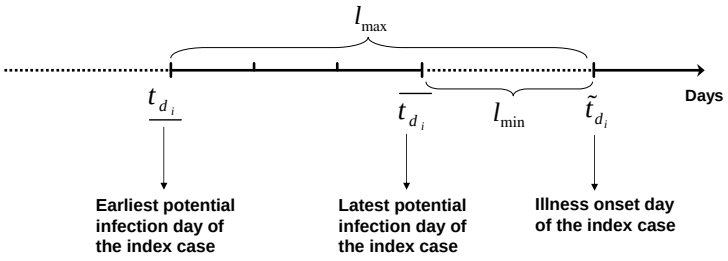

Fig. 1. Time sequence of the earliest and latest potential infection days and the illness onset day for an index case as determined by the minimum and maximum duration of the latent period. Members other than the index case in the household must have escaped infection up to day $\underline{t_{d_i}}$. If infected after day $\underline{t_{d_i}}$, a non-index case shows no illness symptoms up to day $\tilde{t}_{d_i}$ with a positive probability. This probability is 1 if the infection happens after day $\overline{t_{d_i}}$.

where $g(\tilde{t}_i|t)$ is the probability of illness onset on day $\tilde{t}_i$, given infection on day t, or equivalently, the probability that the latent period lasts for $\tilde{t}_i - t$ days (Rampey *et al.*, 1992). The Newton-Raphson algorithm can be applied to obtain the MLEs.

Household-level randomization leads to the same treatment status for all members of the same household, if they are treated during the same period. Thus, the escape probability (1) simplifies to

$$q_{ij}(t) = 1 - \psi^{r_i(t)} pf(t|\tilde{t}_j), \tag{6}$$

where $1 - \psi$ is the total effect. Only ψ is identifiable from within-household infections, and θ can only be identified from contacts within the community. Under the assumption of independence between θ and ϕ, ϕ is estimated as ψ/θ. If the assumption of independence it not made, ϕ can no longer be identified.

2.1.2. Case-ascertained follow-up study design

In a case-ascertained follow-up study, a household is enrolled only if an index case (i.e., illness) is identified, giving rise to left truncation. The full likelihood in (5) should be conditioned on the infection status of household members on the illness onset day of the index case. Possible selection bias due to left truncation is then eliminated because the index case does not contribute to the conditional likelihood. To see this, assume that person i is a non-index-case household member. Let d_i be the identifier for the index case in the household of person i. Also, let $\tilde{t}_{d_i}$ be the illness onset day of d_i. Similarly, $\underline{t_{d_i}} = \tilde{t}_{d_i} - l_{max}$ and $\overline{t_{d_i}} = \tilde{t}_{d_i} - l_{min}$ are the earliest and latest potential infection day of the index case (see Fig. 1).

The marginal probability that i does not show illness symptoms up to day $\tilde{t}_{d_i}$ is

$$\Pr(\tilde{t}_i > \tilde{t}_{d_i}) = \left\{ \sum_{t=1}^{T} Z_i(t) \ \Pr(\tilde{t}_i > \tilde{t}_{d_i}|t) \right\} + Q_i(T), \tag{7}$$

where $\Pr(\tilde{t}_i > \tilde{t}_{d_i}|t) = \sum_{\tau > \tilde{t}_{d_i}} g(\tau|t)$ is the probability that the latent period is longer than $\tilde{t}_{d_i} - t$, given person i was infected on day t. Since the latent period is bounded between l_{min} and l_{max}, $\Pr(\tilde{t}_i > \tilde{t}_{d_i}|t)$ is 0 for $\{t : \tilde{t}_{d_i} - t \geq l_{max}\}$ and 1 for $\{t : \tilde{t}_{d_i} - t < l_{min}\}$, or equivalently 0 for $\{t : t \leq \underline{t_{d_i}}\}$ and 1 for $\{t : t > \overline{t_{d_i}}\}$. Also since $\left\{ \sum_{t > \overline{t_{d_i}}}^{T} Z_i(t) \right\} + Q_i(T) = Q_i(\overline{t_{d_i}})$, the marginal probability reduces to

$$\Pr(\tilde{t}_i > \tilde{t}_{d_i}) = \left\{ \sum_{t=\underline{t_{d_i}}+1}^{\overline{t_{d_i}}} Z_i(t) \ \Pr(\tilde{t}_i > \tilde{t}_{d_i}|t) \right\} + Q_i(\overline{t_{d_i}}). \tag{8}$$

By conditioning the full likelihood L_i in (5) on the marginal likelihood $L_i^m = \Pr(\tilde{t}_i > \tilde{t}_{d_i})$, the solution to the left truncation problem naturally results from a decomposition of L_i and L_i^m as follows. The marginal likelihood can be written as $L_i^m = Q_i(\underline{t_{d_i}})\mathcal{A}_i$, where

$$\mathcal{A}_i = \sum_{t=\underline{t_{d_i}}+1}^{\overline{t_{d_i}}} \Big\{ \Big(\prod_{\tau=\underline{t_{d_i}}+1}^{t-1} e_i(\tau)\Big)\big(1-e_i(t)\big)\ \Pr(\tilde{t}_i > \tilde{t}_{d_i}|t) \Big\} + \prod_{t=\underline{t_{d_i}}+1}^{\overline{t_{d_i}}} e_i(t). \tag{9}$$

Similarly, the full likelihood can be broken into $L_i = Q_i(\underline{t_{d_i}})\mathcal{B}_i$, where

$$\mathcal{B}_i = \begin{cases} \prod_{\underline{t_{d_i}}+1}^{T} e_i(t), & \text{if } i \text{ is not infected,} \\ \sum_{t=\underline{t_i}}^{\overline{t_i}} g(\tilde{t}_i|t)\big(\prod_{\tau=\underline{t_{d_i}}+1}^{t-1} e_i(\tau)\big)\big(1-e_i(t)\big), & \text{otherwise,} \end{cases}$$

is the full likelihood after $\underline{t_{d_i}}$. The conditional likelihood contribution of i is given by

$$L_i^c = L_i/L_i^m = \mathcal{B}_i/\mathcal{A}_i, \tag{10}$$

and the joint conditional likelihood $\prod_i L_i^c$ for the whole population will be maximized to obtain the MLEs. The expression of L_i^c in (10) suggests that the likelihood history up to day $\underline{t_{d_i}}$ can be cancelled in the conditional likelihood for case-ascertained follow-up studies. This is not surprising since $\underline{t_{d_i}} + 1$ is the first day with uncertainty about infection status of non-index-case members in the household. For an index case, both the full and the marginal likelihood contributions are equal to the probability of illness onset on day $\tilde{t}_{d_i}$, hence the likelihood contribution of the index case is cancelled by conditioning.

2.1.3. Allowing for covariate-specific transmission probabilities

It is straightforward to extend the likelihood model to allow for covariate-specific transmission probabilities. For example, it is possible to allow for age-group specific transmission probabilities between children, p_{cc}, adults p_{aa}, adults to children p_{ac}, and children to adults p_{ca}, as well as age-group specific community probabilities of infection, b_a and b_c, where subscript a denotes adults and c denotes children (Addy *et al.*, 1991).

2.2. Assessing Fit

We use a frequency grouping method motivated by the development of goodness-of-fit tests for logistic regression models (Hosmer and Lemeshow, 1980). Since the only observable binary outcome is daily illness onset status (yes/no) for each person, we track the exposure status of each participant and then calculate the probability of illness onset for each day based on the fitted model. This probability serves two purposes: first, it is the model-predicted frequency of illness onset times per person-day that will be compared with the observed onset times; and second, it measures the risk level of exposure to infection that is used to determine the risk categories.

The probability of illness onset on day t for person i is

$$\pi_i(t) = \sum_{\tau=t-l_{max}}^{t-l_{min}} \Big\{\big(1-e_i(\tau)\big) \prod_{s=t-l_{max}}^{\tau-1} e_i(s)\Big\} g(t|\tau).$$

Suppose the population of $\pi_i(t)$'s has been grouped into m risk levels according to the percentiles with cut points $0 = c_0 < c_1 < \ldots < c_m = 1$. Then $\hat{n}_k = \sum_{c_{k-1}<\pi_i<c_k} \pi_i$ is the fitted number of illness onsets at level

Table 2. Empirical cumulative distributions of the latent period and the infectious period for influenza (Elveback et al., 1976).

Latent Period		Infectious Period	
Duration (days)	Cumulative Probability	Duration (days)	Cumulative Probability
1	0.2	3	0.3
2	0.8	4	0.7
3	1.0	5	0.9
		6	1.0

k. Let N_k be the total person-days and n_k be the observed number of illness onsets at level k, then

$$\sum_{j=1}^{m} \frac{N_k(n_k - \hat{n}_k)^2}{\hat{n}_k(N_k - \hat{n}_k)} \sim \chi^2_{m-2},$$

which simplifies to $\sum_{j=1}^{m} \frac{(n_k - \hat{n}_k)^2}{\hat{n}_k}$, if $\hat{n}_k \ll N_k$ for all k.

3. Simulation study

We created a discrete event stochastic simulation model to assess how well the estimation procedures perform and to investigate what the best intervention trial design would be. We created a simulation community composed of 749 households of size two and larger with 2000 people roughly based on the distributions of age and household sizes from the US Census 2000. Since households with a single member do not provide information on ϕ, we do not include single-member households in the simulation. The profile of the simulated household sizes is $\{2:67\%, 3:13\%, 4:10\%, 5:7\%, 6:2\%, 7:1\%\}$. We stopped the simulated trials on day 100 which is generally the maximum length of clinical trials of influenza antiviral agents and vaccines. We set the values of parameters as $b = 0.004, p = 0.1, \theta = 0.7, \phi = 0.2$, and assume independence between θ and ϕ. The empirical latent and infectious period distributions were based on past experience with influenza (Elveback *et al.*, 1976), and are given in Table 2, from which $f(t|\tilde{t}_i)$ and $g(\tilde{t}_i|t)$ were derived. One thousand stochastic replications were carried out for each scenario investigated.

We are interested in which intervention trial designs give the most efficient estimates of the AVE_S and the AVE_I. Table 3 compares estimates, Monte Carlo standard errors, and coverage rates of 95% confidence intervals (C.I.) based on estimated standard errors between randomization schemes (columns) and between ascertainment methods (rows) for θ and ϕ. In both ascertainment methods, both θ and ϕ are biased slightly upward under household-level randomization. Most striking is that individual-level randomization is much more efficient than household-level randomization regardless of the ascertainment scheme. In addition, household-level randomization yields lower coverage probabilities of 95% C.I.'s. Of all infections that occurred, only 5% on average occurred during prophylaxis under household-level randomization, much lower than 11% under individual-level randomization. The low proportion of infections during prophylaxis leads to the larger standard errors of $\hat{\theta}$ and $\hat{\phi}$. However, the overall attack rates are similar, 43% for individual-level randomization and 44% for household-level randomization. Table 3 also shows that estimates of θ and ϕ in the case-ascertained study are almost as precise as in the prospective study, given the same randomization scheme. Point estimates of the transmission probabilities p and b do not differ by trial design, but the prospective design gives a much smaller Monte Carlo standard error for $\hat{b}$ due to its richer information about community-to-person transmissions (results not shown). In our simulation setting with $b = 0.004$, the case-ascertained design involves on average only 62% of the households included in the prospective design. If we change b to 0.001 and 0.01, this proportion decreases to 23% and increases to 90%, respectively.

Table 3. Comparison of MLEs by randomization schemes and household follow-up schemes. Results are based on 1000 simulations.

Parameter†		Estimate		MC s.e.‡		95% C.I. coverage(%)$	
		I⋆	H⋆	I	H	I	H
θ							
	Prospective	0.70	0.71	0.083	0.25	95.3	93.8
	Case-ascertained	0.70	0.71	0.083	0.26	96.1	94.3
ϕ							
	Prospective	0.20	0.24	0.045	0.16	94.6	91.5
	Case-ascertained	0.20	0.24	0.044	0.15	95.3	91.3

† True efficacy-related parameters are set to θ=0.70, ϕ=0.20.
‡ Monte Carlo standard errors.
⋆ I: individual-level randomization, H: household-level randomization.
$ 95% C.I. is obtained as $\exp\{\log(\hat{\lambda}) \pm 1.96 \times s.e.(\log(\hat{\lambda}))\}$, $\lambda = \theta, \phi$.

4. Data analysis

The two trials of oseltamivir in Table 1 fall into the category of case-ascertained follow-up with household-level randomization. Only models (1) or (2) and not model (6) can be applied since neither trial had families in which all members received treatments during the same period. In the first trial (trial I, Welliver *et al.*(2001)), index cases were not treated with either oseltamivir or placebo but eligible exposed household members (aged 12+ years) were given oseltamivir for prophylaxis or placebo within 48 hours of the onset of illness symptoms in the index case. In the second trial (trial II, Hayden *et al.*(2004)), all index cases were treated with oseltamivir after ascertainment. Exposed household members (aged 1+ years) were randomized to groups with or without oseltamivir for prophylaxis and, in the case of illness, were treated with oseltamivir. The AVE_S can be estimated solely from trial I, but most of the information about the AVE_I comes from pooling data from the two trials. We used only laboratory-confirmed illness in our analysis. We divided the population into two age groups: children (1-17) and adults (18+).

In an initial analysis, we assumed the two trials share the same p, θ, ϕ, and ψ, but have different probabilities of infection from the community, i.e., b_1 for trial I and b_2 for trial II. Hence, the pooled population was partially stratified by trial. The MLEs of b_1 and b_2 for the two trials are about the same, 0.0012 (s.e. 0.00036) and 0.0011 (s.e. 0.00026), indicating that each untreated person had about 1/1000 chance of being infected from outside of the household each day.

Table 4 shows the results based on MLEs when we allow for age-group specific transmission probabilities. Under the assumption of independence between θ and ϕ (model (1)), prophylactic use of oseltamivir is shown to be significantly protective against infection with illness by $\widehat{\text{AVE}}_S = 0.85$ (95% C.I.=(0.52, 0.95)). Thus, if a person uses oseltamivir prophylactically, it reduces his chance of being infected with illness by 85% per daily exposure to an untreated infected person. In addition, oseltamivir reduces the infectiousness of infected people who take the drug therapeutically or prophylactically, with $\widehat{\text{AVE}}_I = 0.66$ (95% C.I.=(-0.10, 0.89)), but the effect is not statistically significant. Thus, if an infected person uses oseltamivir therapeutically or prophylactically, it reduces his chance of transmitting influenza to another person not using the drug by 66% per daily household exposure. With AVE_T=0.95 (95% C.I.=(0.77, 0.99)), when both people in a transmission pair are treated, oseltamivir reduces the chance of transmission by 95%. Welliver *et al.*(2001) reported an efficacy in preventing clinical influenza of 0.89 (95%C.I.=(0.71, 0.96)) and an efficacy in inhibiting viral shedding of 0.84 (95%C.I.=(0.57, 0.95)) for trial I, and Hayden *et al.*(2004) reported a conditional efficacy in preventing lab-confirmed influenza of 0.68 (95%C.I.=(0.35, 0.84)) when index cases were treated. These results are comparable to our efficacy estimates. If we drop the assumption of independence between θ and ϕ (model (2)), the point estimates of AVE_S and AVE_I are raised to 0.93 (95% C.I.=(0.50, 0.99)) and 0.78 (95% C.I.=(-0.27, 0.96)) respectively, whereas AVE_T decreases to 0.87 (95% C.I.=(0.41, 0.97)). To compare these two models, a likelihood ratio test is performed, and the p-value is 0.16 based on a χ^2_1 distribution,

Table 4. Maximum likelihood estimates by age (1-17 vs 18+) for pooled oseltamivir trials conducted in 1998-1999 and 2000-2001, North America and Europe.

With Assumption $\psi = \theta\phi$	Parameter	MLE	95% C.I.
Yes	b_c †	0.0023	(0.0015, 0.0035)
	b_a	0.00055	(0.0003, 0.001)
	p_{cc}	0.038	(0.023, 0.063)
	p_{ca}	0.012	(0.007, 0.021)
	p_{ac}	0.018	(0.008, 0.040)
	p_{aa}	0.022	(0.014, 0.034)
	AVE_S	0.85	(0.52, 0.95)
	AVE_I	0.66	(-0.10, 0.89)
	AVE_T	0.95	(0.77, 0.99)
No	AVE_S	0.93	(0.50, 0.99)
	AVE_I	0.78	(-0.27, 0.96)
	AVE_T	0.87	(0.41, 0.97)
	SAR_{cc} ‡	0.15	(0.074, 0.21)
	SAR_{ca}	0.049	(0.021, 0.075)
	SAR_{ac}	0.071	(0.014, 0.13)
	SAR_{aa}	0.086	(0.047, 0.12)

†, ‡ Subscription c denotes child (1-17), a denotes adult (18+), and ca denotes child-to-adult transmission.
‡ SAR_{vu} is based on the average 4.1 days of infectious period, i.e., $\text{SAR}_{vu} = 1 - (1 - p_{vu})^{4.1}$.

indicating that the assumption of independence and multiplicativity between θ and ϕ is reasonable for these data.

In Table 4, we only present estimates of transmission probabilities for model (1) since the assumption of independence between θ and ϕ does not have a substantial effect on these estimates. The estimated probability of being infected from the community on any given day for children (0.0023) is four times higher than that for adults (0.00055). For secondary spread within the household, the estimated transmission probability between untreated children ($\widehat{p}_{cc} = 0.038$) is nearly twice as high as that between untreated adults ($\widehat{p}_{aa} = 0.022$). The $\widehat{\text{AVE}}_S$ and the $\widehat{\text{AVE}}_I$ are about the same as those found in the unstratified analysis (results not shown). Another way to assess secondary spread in the household is via the household secondary attack rate (SAR). We define SAR_{vu} as the probability that an untreated infected person with covariate value v infects an untreated household member with covariate value u throughout the former person's infectious period. Since we assume the average infectious period for influenza to be 4.1 days, we define $\text{SAR}_{vu} = 1 - (1 - p_{vu})^{4.1}$. The estimated SARs are give in Table 4. The estimated SAR between children is almost twice that between adults.

We assessed how well model (1) stratified by age-group fits the data. We categorized the probabilities of illness onset per person-day into 10 risk levels in Table 5 according to the clustering pattern of the probabilities. For the most part, there is good agreement between the observed and predicted illness onset counts. To calculate an approximate χ-square statistic, we collapsed the first three levels into one level, levels 4 and 5 into one level, and levels 6 and 7 into one level. This results in $\chi^2_4 = 2.36$ that has a p-value of 0.67, indicating an adequate fit.

To assess sensitivity of the efficacy estimates to the distributions of the latent and infectious periods, we re-analyzed the data at several combinations of different distributions as shown in Table 6. The estimates of AVE_S and AVE_T are robust to the distribution of latent and infectious periods, as compared to AVE_I. The estimates of AVE_S (AVE_T) range from 0.83 to 0.87 (0.92 to 0.97) under model (1) and 0.91 to 0.94 (0.85 to 0.89) under model (2), all 95% confidence intervals excluding 0 (not shown in the table). The point estimates of AVE_I increase as the mean duration of the latent and infectious periods become longer, varying from 0.47 to 0.79 under model (1) and from 0.57 to 0.93 under model (2), and remain statistically not significant in most cases (95% confidence intervals not shown).

Table 5. Assessing goodness-of-fit of the likelihood model† for pooled oseltamivir trials conducted in 1998-1999 and 2000-2001, North America and Europe.

Risk Level	Total Person-days	Observed # of illness onsets	Predicted # of illness onsets
1	2084	0	0
2	1321	0	0
3	15878	8	9
4	1434	1	3
5	8165	19	22
6	933	3	3
7	935	5	4
8	1241	12	9
9	1084	17	18
10	894	25	27

† With assumption of $\psi = \theta\phi$.

Table 6. Sensitivity of Maximum likelihood estimates for efficacies to distributions of the latent period and the infectious period for pooled oseltamivir trials conducted in 1998-1999 and 2000-2001, North America and Europe.

Latent† of Period	With Assumption $\psi = \theta\phi$	Infectious Period‡								
		3.7[d]			4.1[e]			4.8[f]		
		AVE_S	AVE_I	AVE_T	AVE_S	AVE_I	AVE_T	AVE_S	AVE_I	AVE_T
1.5[a]										
	Yes	0.84	0.47	0.92	0.85	0.54	0.93	0.87	0.62	0.95
	s.e.	(0.094)	(0.284)	(0.059)	(0.087)	(0.243)	(0.048)	(0.079)	(0.198)	(0.036)
	No	0.91	0.57	0.85	0.92	0.64	0.86	0.93	0.72	0.89
	s.e.	(0.091)	(0.285)	(0.124)	(0.080)	(0.239)	(0.106)	(0.066)	(0.191)	(0.087)
2.0[b]										
	Yes	0.84	0.60	0.94	0.85	0.66	0.95	0.87	0.72	0.96
	s.e.	(0.096)	(0.239)	(0.048)	(0.089)	(0.204)	(0.039)	(0.080)	(0.167)	(0.029)
	No	0.92	0.74	0.85	0.93	0.78	0.87	0.94	0.84	0.89
	s.e.	(0.079)	(0.232)	(0.119)	(0.070)	(0.195)	(0.102)	(0.059)	(0.156)	(0.083)
2.4[c]										
	Yes	0.83	0.70	0.95	0.84	0.74	0.96	0.86	0.79	0.97
	s.e.	(0.100)	(0.211)	(0.041)	(0.093)	(0.180)	(0.033)	(0.083)	(0.148)	(0.025)
	No	0.93	0.89	0.85	0.93	0.91	0.86	0.94	0.93	0.89
	s.e.	(0.075)	(0.196)	(0.118)	(0.067)	(0.162)	(0.101)	(0.058)	(0.129)	(0.084)

†, ‡ Mean duration (days) of latent and infectious periods.
a Probability mass function is (1:0.6, 2:0.3, 3:0.1).
b Probability mass function is (1:0.2, 2:0.6, 3:0.2).
c Probability mass function is (1:0.1, 2:0.4, 3:0.5).
d Probability mass function is (3:0.6, 4:0.2, 5:0.1, 6:0.1).
e Probability mass function is (3:0.3, 4:0.4, 5:0.2, 6:0.1).
f Probability mass function is (3:0.1, 4:0.2, 5:0.5, 6:0.2).

5. Discussion

We have developed statistical methods for estimating the effectiveness of infectious disease interventions from illness incidence data in close contact groups. We have used these methods to evaluate which study design may be the best for clinical trials in close contact groups. Our simulations indicate that individual-level randomization is more efficient than household-level randomization in terms of producing smaller standard errors for the parameter estimates of interest. Our simulations also reveal that, if correctly adjusted for left truncation, the case-ascertained follow-up study provides estimates of almost the same quality as those provided by the prospective follow-up study, except that the latter estimates b, a nuisance parameter, with better precision.

In this work, we provide the first joint estimates of the efficacies of an influenza antiviral agent in reducing susceptibility to infection with illness (AVE_S), in reducing infectiousness (AVE_I), and the combined effect (AVE_T). We estimate that the prophylactic use of oseltamivir reduces the probability of infection and illness given exposure to infection by 85%. In addition, if oseltamivir is taken therapeutically soon after symptoms appear or prophylactically, then the probability the infected person will transmit to others in the household is reduced by 66%. Longini *et al.*(2004) used values similar to these estimates in a stochastic simulation model of pandemic influenza to show that the targeted use of oseltamivir in close contact groups could slow the transmission of influenza on the community level. Such an intervention would be important because there would be little or no vaccine available against the first wave of pandemic influenza (Longini *et al.*, 1978; Longini *et al.*, 2004). In addition, the targeted use of oseltamivir could be used to possibly contain a potential pandemic strain of influenza at the source (Longini, *et al.*, 2005). Above we make an assumption of independence between AVE_S and AVE_I, and the combined effect AVE_T under this assumption is 95%. Without this assumption, both AVE_S and AVE_I increase to some extent, and AVE_T is lowered to 87%. Our higher estimates of the SAR between children than between adults and from the community to children than to adults, similar to those of Addy *et al.*(1991), confirm that children are both the major introducers and spreaders of influenza in households.

Several issues in this research deserve future investigation. First, unmeasured heterogeneity may be present in both the probability that a household is invaded with influenza and SARs within households. In this case, random effects terms added to the transmission probabilities might be appropriate (Halloran *et al.*, 2003a). However, this would complicate the model and increase the data requirements. We already take the clustering of cases into account through the transmission model that includes secondary transmission within households. A related issue is the possibility of ascertainment bias for the case-ascertainment design since households more prone to influenza invasion have a higher probability of having an index case and, thus, getting into the sample. Dealing with this potential source of bias will require further research. Parameter estimation could be affected by potential misclassification of infection and illness status of study participants. Validation set methods could be developed for study designs that include more sensitive infection detection tests for a subset of study participants (Halloran *et al.*, 2003b). Extension of Bayesian approaches (Cauchemez *et al.*, 2004) could allow estimation of the latent and infectious periods as well.

We have assumed that the susceptibility status of each individual in the trial is known on day 1. However, if some of the susceptibles are only partially susceptible or immune at that time, there would not be a bias in the AVE estimators as long as the trial were randomized. If the trial were not randomized, then the initial susceptibility status of each participant would have to be taken into account, at least partially through an immunity marker such as antibody level.

For small trials or trials with a small numbers of illnesses, the maximum likelihood method that we develop here could have convergence problems. To help deal with this problem and to find starting values for the maximum likelihood algorithm, we developed an alternative approach based on generalized linear models fitted by iteratively re-weighted least squares in combination with the EM algorithm (Yang *et al.*, 2004).

In this work, we assume that the intervention trial would be performed on a sample of several hundred households that would constitute a small fraction of the community. This has been the case in past trials. Thus, intervention in the trial households would not affect transmission in the community nor would the trial provide information on the epidemic curve in the community. For this reason we modeled the daily probability that a susceptible and untreated person is infected by a source of infection from the community, b, as constant in time. If information on the epidemic curve were available from the community, then b could be treated as time-dependent (Longini *et al.*, 1996). In addition, if the trial were done on a community scale, then it would be appropriate to take the entire epidemic process into account in the analysis (Halloran *et al.*, 1999; Longini *et al.*, 2002).

Based on this research, we recommend that future intervention trials in close contact groups for the efficacy of the intervention to prevent infection, illness and transmission be randomized on an individual level and be case-ascertained if there are limited resources for the trial. Use of the individual level randomization in the trials analyzed here would have resulted in a more precise estimate of the AVE_I. The close contact group setting allows the estimation methods to condition on exposure to infection, which, in turn, allows estimation of AVE_S, AVE_I, and AVE_T (Halloran *et al.*, 1997). These estimates can be used in transmission models to assess the potential effect of the intervention on a community level (Longini *et al.*, 2004; Longini *et al.*, 2005).

Acknowledgements

This work was partially supported by National Institute of Allergy and Infectious Diseases grant R01-AI32042 and National Institute of General Medical Sciences MIDAS grant U01-GM070749. The data on the household Oseltamivir prophylaxis studies were provided by Roche Laboratories Inc.

References

Census 2000. (Posted July 25, 2001) United States Census Bureau. *http://www.census.gov/.*

Addy, C. L., Longini, I. M. and Haber, M. J. (1991) A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics*, 47, 961-974.

Becker, N. G. (1989) *Analysis of Infectious Disease Data.* New York, NY: Chapman and Hall.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. and Boëlle, P. Y. (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statist. Med.*, 23, 3469-3487.

Datta, S., Halloran, M. E. and Longini, I. M. (1999) Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: randomization by individual versus household. *Biometrics*, 55, 792-798.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39, 1-38.

Elveback, L. R., Fox, J. P. and Ackerman, E. (1976) An influenza simulation model for immunization studies. *American Journal of Epidemiology*, 103, 152-165.

Farrington, C. P. (2003) On vaccine efficacy and reproduction numbers. *Mathematical Biosciences*, 185, 89-109.

Halloran, M. E., Struchiner, C. J. and Longini, I. M. (1997) Study designs for different efficacy and effectiveness aspects of vaccination. *American Journal of Epidemiology*, 146, 789-803.

Halloran, M. E., Longini, I. M. and Struchiner, C. J. (1999) Design and interpretation of vaccine field studies. In *Epidemiologic Reviews: Vaccines* (eds A. S. Monto and S. B. Thacker), 21, 73-88.

Halloran, M. E., Longini, I. M., Cowart, D. M. and Nizam, A. (2002) Community trials of vaccination and the epidemic prevention potential. *Vaccine*, 20, 3254-3262.

Halloran, M. E., Préziosi, M-P. and Chu, H. (2003a) Estimating vaccine efficacy from secondary attack rates. *J. Am. Statist. Ass.*, 98, 38-46.

Halloran, M. E., Longini, I. M., Gaglani, M. J., Piedra, P. A. and Chu, H. (2003b) Estimating efficacy of trivalent, cold-adapted, influenza virus vaccine (CAIV-T) against influenza A (H1N1) and B using surveillance cultures. *American Journal of Epidemiology*, 158, 305-311.

Hayden, F. G., Belshe, R., Villanueva, C., Lanno, R., Hughes, C., Small, I., Dutkowski, R., Ward, P. and Carr, J. (2004) Management of influenza in households: a prospective, randomized comparison of Oseltamivir treatment with or without postexposure prophylaxis. *Journal of Infectious Diseases*, 189, 440-449.

Hayden, F. G., Gubareva, L. V., Monto, A. S., Klein, T. C., Elliott, M. J., Hammond, J. M., Sharp, S. J. and Ossi, M. J., Zanamivir Family Study Group (2000) Inhaled zanamivir for the prevention of influenza in families. *New England Journal of Medicine*, 343, 1282-1289.

Hosmer, D. W. and Lemeshow, S. (1980) Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics*, A9(10), 1043-1069.

Longini, I. M., Ackerman, E. and Elveback, L. R. (1978) An optimization model for influenza A epidemics. *Math. Bioscience*, 38, 141-157.

Longini, I. M. and Koopman, J. S. (1982) Household and Community Transmission Parameters from Final Distributions of Infections in Households. *Biometrics*, 38, 115-126.

Longini, I. M., Koopman, J. S., Haber, M. J. and Cotsonis, G. A. (1988) Statistical Inference for Infectious Diseases: Risk-Specified Household and Community Transmission Parameters. *American Journal of Epidemiology*, 128, 845-859.

Longini, I. M. and Halloran, M. E. (1996) A frailty mixture model for estimating vaccine efficacy. *Applied Statistics*, 45, 165-173.

Longini, I. M., Halloran, M. E. and Nizam, A. (2002) Model-based estimation of vaccine effects from community vaccine trials. *Statistics in Medicine*, 21, 481-496.

Longini, I. M., Halloran, M. E., Nizam, A. and Yang, Y. (2004) Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, 159, 623-633.

Longini, I. M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A. T. and Halloran, M. E. (2005) Containing pandemic influenza at the source. *Science*, 309, 1083-1087.

Louis, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc. B*, 44, 226-233.

Magder, L. and Brookmeyer, R. (1993) Analysis of infectious disease data from partner studies with unknown source of infection. *Biometrics*, 49, 1110-1116.

Monto, A. S., Pichichero, M. E., Blanckenberg, S. J., Ruuskanen, O., Cooper, C., Fleming, D. M. and Kerr, C. (2002) Zanamivir prophylaxis: an effective strategy for the prevention of influenza types A and B within households. *Journal Infectious Diseases*, 186, 1582-1588.

O'Neill, P., Balding, D. J., Becker, N. G., Eerola, M. and Mollison, D. (2000) Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Appl. Statist.*, 49, 517-542.

Rampey, A. H., Longini, I. M., Haber, M. J. and Monto, A. S. (1992) A discrete-time model for the statistical analysis of infectious disease incidence data. *Biometrics*, 48, 117-128.

Satten, G. A., Mastro, T. D. and Longini, I. M. (1994) Modelling the female-to-male per-act HIV transmission probability in an emerging epidemic in Asia. *Statist. Med.*, 13, 2097-2106.

Tanner, M. A. (1996) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* New York, NY: Springer-Verlag.

Welliver, R., Monto, A. S., Carewicz, O., Schattemanet, E., Hassman, M., Hedrick, J., Jackson, H. C., Huson, L., Ward, P. and Oxford, J. S. (2001) Effectiveness of oseltamivir in preventing influenza in household contacts: a randomized controlled trial. *Journal of the American Medical Associtation*, 285, 748-754.

Yang, Y., Longini, I. M., Halloran, M. E. (2004) Design and Evaluation of Prophylactic Interventions Using Infectious Disease Incidence Data from Close Contact Groups. *Technical Report #04-09, Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA, USA.*