Statistical Methods for Infectious Diseases
Household Based Studies I
Lecture 7B

M. Elizabeth Halloran

Fred Hutchinson Cancer Research Center and
University of Washington
Seattle, WA, USA

January 27, 2009

Concepts

- Households within a community
- Discrete-time model
- Continuous-time model

Final-value data

- Data structure
- Model
- Likelihood Estimation and Inference
- Bayesian Inference

Concepts
- Households within a community
- Discrete-time model
- Continuous-time model

Final-value data
- Data structure
- Model
- Likelihood Estimation and Inference
- Bayesian Inference

Households within a Community

- Analyses that assume that the households or other transmission units are nested in a community.
- Community-acquired infection serves as a source of initial infection within households as well as possible further cases in the household.
- Infected household members can infect others in the household.

Two general types of parameters

- One for infection from the community, the other for transmission from an infective to a susceptible within the household.
- The first is an unconditional parameter, that is, it does not condition on exposure to infection, the second a conditional parameter.
- discrete time or continuous time.
- For some data structures, such as contact data on sexual contacts, contacts can be substituted for time.

- One parameter describes the rate of community-acquired infection, the other the rate of transmission from an infective to a susceptible within a household.
- Both continuous and discrete time parameters can be transformed into the probability of acquiring infection from the community over the period of time of interest.
- CPI: the community probability of infection.
- SAR: the secondary attack rate within the household.
- If an estimate of prevalence of infection in the population is available, then the transmission probability in the community at large can also be estimated

Discrete-time model: CPI

- Consider a study from time period 0 to time period T. Let a be the probability a susceptible household member becomes infected from the community in one time unit.
- Let $b = 1 - a$ be the corresponding escape probability.
- Then the probability of escaping infection from the community over the T time periods is $B = b^T = (1 - a)^T$.
- The community probability of infection:

$$\mathrm{CPI} = 1 - B = 1 - b^T = 1 - (1 - a)^T. \tag{1}$$

Discrete-time model: SAR

- Let $q = 1 - p$ be the probability of escaping infectious contact in a household within one time unit.
- Then if a person is infectious for T_I time units, the probability of escaping infection from an infective within a household is $Q = q^{T_I} = (1 - p)^{T_I}$.
- The secondary attack rate:

$$\mathrm{SAR} = 1 - Q = 1 - (1 - p)^{T_I}. \tag{2}$$

Continuous-time model

- α denotes the instantaneous risk of infection from the community
- β denotes the instantaneous risk of infection from an infective in the household.
- In the simplest form, if the study duration is from time 0 to time T and the duration of infectiousness is T_I, then

$$\begin{aligned} \text{CPI} &= 1 - \exp(-\alpha T), \\ \text{SAR} &= 1 - \exp(-\beta T_I). \end{aligned} \tag{3}$$

Vaccine effects and other covariates

- $\theta =$ relative per contact susceptibility of a vaccinated compared to an unvaccinated person, $\mathsf{VE}_S = 1 - \theta$.
- $\phi =$ relative infectiousness per contact of a vaccinated compared with an unvaccinated person, $\mathsf{VE}_I = 1 - \phi$.
- $p =$ per time unit (or per contact) transmission probability between two unvaccinated people in a household,
- $\theta p =$ per time unit (or per contact) transmission probability to a vaccinated susceptible from an unvaccinated infective.

Vaccine effects and other covariates

- The secondary attack from an unvaccinated infective individual to a vaccinate susceptible is $\mathrm{SAR}_{01} = 1 - (1 - \theta p)^{T_I}$.
- If only the infective person is vaccinated, then $\mathrm{SAR}_{10} = 1 - (1 - \phi p)^{T_I}$.
- If both people are vaccinated, then $\mathrm{SAR}_{11} = 1 - (1 - \theta\phi p)^{T_I}$. (independent and multiplicative).
- Alternatively, $\psi =$ vaccine's effect on the transmission probability if both the infective and the susceptible in the contact. are vaccinated, so that $\mathrm{SAR}_{11} = 1 - (1 - \psi p)^{T_I}$.
- The vaccine parameters enter similarly into the continuous time models. For example, $\mathrm{SAR}_{11} = 1 - \exp(-\theta\phi\beta T_I)$.

CPI

- CPI involves only the susceptibles directly.
- One could introduce another parameter θ_c denoting a different effect of vaccination on reducing susceptibility to infection from the community.
- Often, just one parameter θ is estimated.

$VE_{S,p}$ *versus* $VE_{S,SAR}$

- Note that if we define $\mathrm{VE}_{S,p} = 1 - \theta$, this will not necessarily equal

$$VE_{S,SAR} = 1 - \frac{SAR_{01}}{SAR_{00}} = 1 - \frac{1 - (1 - \theta p)^{T_I}}{1 - (1 - p)^{T_I}}. \tag{4}$$

Concepts

- Households within a community
- Discrete-time model
- Continuous-time model

Final-value data

- Data structure
- Model
- Likelihood Estimation and Inference
- Bayesian Inference

Households within a Community

- A model for the distribution of the total number of cases in households from a homogeneous community.
- Community-acquired infection serves as source of initial infection within households as well as possible further cases.
- Infected household members can infect others in the household.
- Ludwig (1975); Longini and Koopman (1982)
- Becker (1989), pp. 182–193.

Households within communities

- Estimate per day transmission probabilities and SAR's within households
- Estimate per-day and overall probability of being infected by the community: CPI
- Can add covariates to model, to estimate covariate effects, including vaccine efficacy

Data structure

- Assume observations on infections in a community over a time period from $t = 0$ to $t = T$
- Could correspond to an epidemic season or some other period of epidemiologic interest
- Main criterion is that all, or nearly all, of the outbreaks in the sample of households should essentially have run their course with $[0, T]$

Data structure, cont'd

- Final value data on infection on n households.
- requires biologic confirmation of susceptibility before and infection after.
- often not available, so use symptom data, and assume that everyone was susceptible at the beginning.

Final-size distribution of household infections

- Assume sources of infection from community distributed homogeneously.
- Assume household members mix at random with household.
- Assume each household member can be infected either from within household or from community

Influenza A(H3N2) infections in 1977-1978 and 1980-1981

Observed distribution combined epidemics in Tecumseh, Michigan.

No.	No. of susceptibles per household[a]				
infected	1	2	3	4	5
0	110	149	72	60	13
1	23	27	23	20	9
2		13	6	16	5
3			7	8	2
4				2	1
5					1
Total	133	189	108	106	31

[a] The criterion for classifying individuals as susceptible is a preseason hemagglutination inhibition test detecting no antibody in a dilution of 1 in 128 or less. Households with more than five susceptibles are deleted from all analyses. from Addy et al (1991).

Notation and data structure

- a_{jk} = observed number of households with

  k original susceptibles of which

  j become infected,

  $k = 1, 2, \ldots, K$ and $j = 0, 1, \ldots, k$,

- where $\sum_k \sum_j a_{jk} = n$.

Binomial Model of Probability of Infection

- Basic idea: exposure to infection occurs in discrete contacts
- Generally assume that each contact and exposure is independent of another.
- $p =$ probability of infection upon contact
- $q = 1 - p =$ escape probability
- $q^n = (1 - p)^n =$ probability of escaping infection from n contacts

Infection from the community

- Define a_t = probability a susceptible household member becomes infected from the community in time period t.
- Let $b_t = 1 - a_t$ be the corresponding escape probability
- Define B = probability a susceptible individual is not infected from community during period of observation.
- Expression for B is

$$B = \prod_{t=0}^{T} b_t, \tag{5}$$

Infection from household members

- Now consider effect of secondary spread within household following introductions from the community.
- An individual infected in time period t_0 will pass through a series of stages in time periods $t_1, t_2, \dots,$ until becoming immune.

Infection from household members

- Define p_t as probability that an infective who was infected at time $t = t_0$ will make infectious contact in the household with another individual in time period t.
- Then $\{p_t\}$ describes pattern of infectiousness over time. The structure of $\{p_t\}$ is
  $p_t = 0$ when $t_0 \leq t \leq t_l$, the latent period
  $p_t > 0$ when $t_{l+1} \leq t \leq t_m$, the infectious period
  $p_t = 0$ when $t_{m+1} \leq t < t_\infty$, the immune period.

Infection from household members, cont'd

- Let $q_t = 1 - p_t$ be the daily probability of escaping infectious contact.
- Then if there is an infected individual in the household who became infected at time $t = t_0$,
- Let Q_{t_r} be the probability that a susceptible individual has escaped infection within the household at time t_r, $t_0 \leq t_r < t_{m+1}$.

Infection from household members, cont'd

- The probability Q that the susceptible individual escapes infectious contact from the infective during his entire period of infectiousness is

$$Q = \prod_{t=t_0}^{t_m} q_t = Q_{t_m}$$

- The pattern of infection $\{q_t\}$ does not influence the magnitude of Q.

Relation of Q and B to other parameters

- $SAR = 1 - Q$, the household secondary attack rate.
- $CPI = 1 - B$, the community probability of infection.
- If household structure is ignored, then CPI is the incidence proportion, or attack rate.

Derive final-size distribution

- Assume all households under consideration free of infected members at beginning and end of period of observation.
- Let $\Pr(j|k)$ be probability that j of k initial susceptibles within a household are infected during course of the epidemic.
- Write $m_{jk} = \Pr(j|k)$ to simplify notation.
- When $k = 1$, it follows that $m_{01} = B$ and $m_{11} = 1 - B$.

Final size distribution, cont'd

- When $k = 2$, $m_{02} = B^2$.
- For m_{12}, two possible ways it can occur
  $\longrightarrow$ either first susceptible individual becomes infected with probability $1 - B$, and second escapes infection from both community and household infected.
  $\longrightarrow$ or first escapes infection and second does not.
- Thus

$$m_{12} = 2(1 - B)BQ = 2m_{11}BQ.$$

- For m_{22}, similarly

$$\begin{aligned} m_{22} &= 2(1 - B)(1 - Q)B + (1 - B)^2 \\ &= 1 - m_{02} - m_{12}, \end{aligned}$$

since the probabilities must sum to one.

Final size distribution, cont'd

- In general, there are $\begin{pmatrix} k \\ j \end{pmatrix}$ ways to get j infected individuals from k originally susceptible ones.
- The $k - j$ susceptible individuals who escape infection must avoid haing infectious contact with the j infective individuals in their household and from the community
- The general expression for m_{jk}

$$\begin{aligned} m_{jk} &= \begin{pmatrix} k \\ j \end{pmatrix} m_{jj} B^{k-j} Q^{j(k-j)}, \quad j < k, \\ m_{kk} &= 1 - \sum_{j=0}^{k-1} m_{jk}. \end{aligned} \tag{6}$$

Final size distribution, cont'd

- If it is assumed that there is spread **only within** households, and there are *i* initial infectives within the household,
- then equation (6) reduces to a form of the final-size distribution for the Reed-Frost model,
- and also describes a more general process in terms of the latent and infectious periods.

When $Q = 1$, no spread in households

- Then (6) reduces to the binomial distribution:

$$m_{jk} = \binom{k}{j} (1 - B)^j B^{k-j}, \quad j \leq k. \tag{7}$$

- This is the distribution of infection in households one would expect if there were no spread, and one would analyze final attack rate data in the community.

Estimation

- The parameters Q and B can be estimated by maximum likelihood. The likelihood function is

$$L(Q,B) = \prod_k \prod_j m_{jk}^{a_{jk}}.$$

- The explicit form of the log likelihood function from (6) is

$$\ln L = c + \sum_k \sum_j a_{jk}\{\ln m_{jj} + (k-j)\ln B + j(k - j\ln Q\}).$$

- Solve the score functions iteratively by method of scoring.
- Use the information matrix to get variances.
- Use estimates from the data to get starting values.

Application to an Influenza Epidemic:

- Households of size 3 (Sugiyama 1960)
- Zero class is present.
- $\hat{B} = 0.856$, var($\hat{B} = 0.0009$).
- $\hat{Q} = 0.834$, var($\hat{Q} = 0.0063$)
- So the estimated probability of a susceptible individual being infected by a case during the course of his infectious period is $1 - Q = SAR = 0.166$.
- Assuming latent period $l = 2$ days and infectious period is 4 days, and $p_t = p$ for $t = 3, \ldots, 6$, then $q_t = 0.834^{0.25} = 0.956$
- Daily probability of infection in a household $p_t = 1 - q_t = 0.044$.

Application to an Influenza Epidemic, cont'd

Number of cases	Observed	Expected
0	29	29.17
1	9	7.87
2	2	3.62
3	2	1.34
Total	42	42.00

- $\chi^2(1df) = 1.222$, $(.25 < P < .50)$

Application to an Influenza Epidemic, con't:

- $\hat{B} = 0.856$,
- So the estimated $\widehat{CPI} = 0.144$.

Influenza A(H3N2) infections in 1977-1978 and 1980-1981

Observed distribution combined epidemics in Tecumseh, Michigan.

No. infected	No. of susceptibles per household[a] 1	2	3	4	5
0	110	149	72	60	13
1	23	27	23	20	9
2		13	6	16	5
3			7	8	2
4				2	1
5					1
Total	133	189	108	106	31

[a] The criterion for classifying individuals as susceptible is a preseason hemagglutination inhibition test detecting no antibody in a dilution of 1 in 128 or less. Households with more than five susceptibles are deleted from all analyses. from Addy et al (1991).

Using Markov chain Monte Carlo Methods for Estimation of the Parameters

- O'Neill et al 2000
- Similar model for the probability of escape from the community and from the household as in Longini and Koopman (1982) where $q_c = B$ (community), $q_h = Q$ (household).
- Metropolis-Hastings algorithm for Bayesian inference
- With uniform prior distributions, the posterior density is equivalent to the likelihood.
- The estimates for the Tecumseh data using maximation on MCMC output were virtually identical to those in Addy et al (1991).
- Here $q_h = 0.8408$ and Addy et al $q_h = 0.8406$
- Here $q_c = 0.8677$, and Addy et al $q_c = 0.8677$

O'Neill et al (2000)

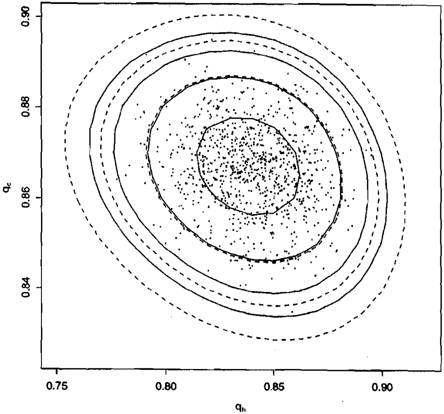

Fig. 1. Influenza data—Reed–Frost model without protection: MCMC sample values (1000 values, at sampling interval 10): ———, contour lines surrounding highest posterior density credible intervals at 50%, 90%, 99% and 99.9% levels; - - - - -, posterior probability density function values of 10%, 1% and 0.1% of its maximum

Extensions to time-of-onset data

- time of infection or disease of each person in each household
- households (partnerships) within communities
- Allows more precise estimation using temporal information
- Allows more flexible distributions of latent, infectious and incubation periods.
- Estimate VE_S, VE_I, VE_T (or AVE)

Thank you!