BIOMETRICS 38, 115–126
March 1982

Household and Community Transmission Parameters from Final Distributions of Infections in Households

Ira M. Longini, Jr and James S. Koopman
Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

SUMMARY

A model is devised for the distribution of the total number of cases in households from a homogeneous community. In the model, community-acquired infection serves as a source of initial infection within households as well as of possible further cases. In addition, infected household members can infect others in the household. Maximum likelihood procedures for the model parameters are given. The model is fitted to symptom data on influenza and the common cold. Influenza seems to spread more easily in the community than within the household, while the opposite may be the case for the common cold. The model, which does not require specification of the time of onset of infection for individuals, can be fitted to serological data; this would provide a more accurate measure of household infection than the symptom data used.

1. Introduction

A tool for measuring the degree or the relative importance of transmission of infectious agents in households and in the community is needed to help describe the dynamics of disease transmission and to answer practical questions about disease control. The model described in this paper has been devised to provide estimates of separate parameters describing household and community transmission of disease from infection data. These data need not specify the time of onset of infection for individuals, nor is it necessary to identify chains of household infection. Several data sets have been selected from the literature to illustrate use of the model and to show how one can answer questions about the spread of influenza and the common cold. Since infection data are not presently available, symptom (i.e. illness) data have been used and all family members are assumed to be initially susceptible. Those individuals who show symptoms will be referred to as 'cases'.

2. Final-Size Distribution of Household Infections

The final-size distribution for the number of household cases during an epidemic period is now derived. We consider only those presumably-infectious diseases that confer immunity (for some period of time) following infection. Assume that sources of infection from the community are distributed homogeneously throughout the community. In addition, suppose that household members mix at random within the household and can infect one another. Thus, each household member can be infected either from within the household or from the community.

Key words: Case; Common cold; Final-size distribution; Infection from the community; Infection within the household; Influenza; Maximum likelihood; Reed–Frost model.

2.1 *Infection from the Community*

Assume that observations are made of infections in a community, starting in time-period $t = 0$ and ending in time-period $t = T$. This period of observation could correspond to an epidemic season or some other period of epidemiological interest. Define a_t as the probability that a susceptible household member becomes infected from the community in time-period t, and $b_t = 1 - a_t$ as the probability that he escapes infection from the community in time-period t. Now define B as the probability that a susceptible individual is not infected from the community during the period of observation. A general expression for B is given by

$$B = \prod_{t=0}^{T} f(b_t), \tag{1}$$

where $f(\cdot)$ is a bounded function describing the infection rates in the community. Since B is estimated directly from the data, $f(\cdot)$ can take any acceptable form. A simple form for $f(\cdot)$ is $f(b_t) = b_t$. Another form, $f(b_t) = \exp(-a/t)$, is used in the Appendix [see (A1)].

2.2 *Infection among Household Members*

We now consider the effect of secondary spread within the household following introductions from the community. An individual who is infected in time-period t_0 will pass through a series of stages at time-periods $t_1, t_2, \ldots$, until he becomes immune. Define p_t as the probability that an infective who was infected in time-period $t = t_0$ will make infectious contact in the household with another individual in time-period t. Then, $\{p_t\}$ describes the pattern of infectiousness over time. The structure of $\{p_t\}$ is

$$\begin{aligned} &p_t = 0 \quad \text{when} \quad t_0 \leq t \leq t_l, && \text{the latent period,} \\ &p_t > 0 \quad \text{when} \quad t_{l+1} \leq t \leq t_m, && \text{the infectious period,} \\ &p_t = 0 \quad \text{when} \quad t_{m+1} \leq t < \infty, && \text{the immune period.} \end{aligned} \tag{2}$$

Let $q_t = 1 - p_t$ be defined as the probability of escaping infectious contact. Then if there is an infected individual in the household who became infected at time $t = t_0$, we define Q_{t_r} as the probability that a susceptible individual has escaped infection within the household at time t_r, where $t_0 \leq t_r < t_{m+1}$. It follows that $Q_{t_r} = \prod_{t'=t_0}^{t_r} q_{t'}$. The probability Q that the susceptible individual escapes infectious contact from the infective during his entire period of infectiousness is

$$Q = \prod_{t=t_0}^{t_m} q_t = Q_{t_m}. \tag{3}$$

Note that the pattern of infection $\{q_t\}$ does not influence the magnitude of Q. As with B, the value of Q is estimated directly from the data and other forms for (3) can be used [see Ludwig (1975) for the analogous continuous form].

2.3 *Final-Size Distribution*

Assume that household members mix randomly among themselves and that the probability that a household member is infected in the community is not affected by the number of infected members in his household. Also, assume that all households under consideration are initially free of infected members at the beginning and end of the period of observation.

To derive the final-size distribution, let $\text{pr}(j \mid k)$ be the probability that j of k initial susceptibles within a household are infected during the course of the epidemic. We write $m_{jk} = \text{pr}(j \mid k)$ to simplify notation. Using similar ideas to those of Ludwig (1975) concerning final-value distributions, the values of m_{jk} are derived in the following way.

When $k = 1$, it follows from the above assumptions that $m_{01} = B$ and $m_{11} = 1 - B$. When $k = 2$, we have, since there is random mixing, $m_{02} = B^2$. As regards m_{12}, there are two ways in which this event can occur. Either the first susceptible individual becomes infected with probability B, and the second escapes infection (from both the infective in the household, with probability Q, and in the community at large), or the first susceptible individual escapes and the second does not. It follows that

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

Derivation of m_{22} follows a similar argument, giving

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

as expected, since the probabilities must sum to one.

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

$$m_{jk} = \binom{k}{j} m_{jj} B^{k-j} Q^{j(k-1)}, \qquad j < k,$$

and (4)

$$m_{kk} = 1 - \sum_{j=0}^{k-1} m_{jk}.$$

The density function (4) can be shown to represent the general expression for specific models considered by others. If it is assumed that there is spread only within the household, and there are initially i infectives within the household, then (4) becomes

$$m_{jk} = \binom{k}{j} m_{ij} Q^{(i+j)(k-j)}, \qquad j < k,$$

and $i+j$ is the final number of infectives in the household. This equation is equivalent to (1.13) of Ludwig (1975) and (14.7) of Bailey (1975). Although (4) provides the final-size distribution for the Reed–Frost model, it also describes a more general process in terms of length of latent and infectious periods. Strict adherence to the Reed–Frost assumptions would require that $t_m - t_{l+1} < \varepsilon$, where ε is small compared to $t_l - t_0$. When considering infection from the community, (4) gives the final-size distribution for the modified Reed–Frost model of Sugiyama (1960).

When $Q = 1$, there is no spread of infection among family members and the disease in question is presumably not 'infectious'. Then (4) reduces to the binomial distribution:

$$m_{jk} = \binom{k}{j} (1-B)^j B^{k-j}, \qquad j \leqslant k. \tag{5}$$

If B is allowed to vary according to the beta distribution, then (5) describes the beta-binomial distribution. Griffiths (1973) used this distribution with data for influenza and the common cold.

3. Estimation

The parameters of interest, Q and B, can be estimated by maximum likelihood (ML). Assume that there are n households and define a_{jk} ($k = 1, 2, \ldots, K$ and $j = 0, 1, \ldots, k$) as the observed frequencies of households with j infectives from k susceptibles, where $\sum_k \sum_j a_{jk} = n$. Then the likelihood function is

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

The explicit form of the log likelihood function from (4) is

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

The ML estimators $\hat{Q}$ and $\hat{B}$ are solutions of

$$0 = \frac{\partial \ln L}{\partial Q}\bigg|_{\hat{Q},\hat{B}} = \sum_k \sum_j a_{jk}\left\{\frac{1}{m_{jj}}\left(\frac{\partial m_{jj}}{\partial Q}\right) + \frac{j(k-j)}{Q}\right\}, \tag{6}$$

$$0 = \frac{\partial \ln L}{\partial B}\bigg|_{\hat{Q},\hat{B}} = \sum_k \sum_j a_{jk}\left\{\frac{1}{m_{jj}}\left(\frac{\partial m_{jj}}{\partial B}\right) + \frac{(k-j)}{B}\right\}. \tag{7}$$

These equations may be solved iteratively using the method of scoring. The elements of the information matrix are given by the expected values, with respect to a_{jk}, of the second partial derivatives, which are

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial Q^2}\right) = \sum_k \sum_j n_k m_{jk}\left[\frac{1}{m_{jj}}\left\{\frac{1}{m_{jj}}\left(\frac{\partial m_{jj}}{\partial Q}\right)^2 - \frac{\partial^2 m_{jj}}{\partial Q^2}\right\} + \frac{j(k-j)}{Q^2}\right], \tag{8}$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial Q\,\partial B}\right) = \sum_k \sum_j n_k \left(\frac{m_{jk}}{m_{jj}}\right)\left\{\frac{1}{m_{jj}}\left(\frac{\partial m_{jj}}{\partial Q}\right)\left(\frac{\partial m_{jj}}{\partial B}\right) - \left(\frac{\partial^2 m_{jj}}{\partial Q\,\partial B}\right)\right\}, \tag{9}$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial B^2}\right) = \sum_k \sum_j n_k m_{jk}\left[\frac{1}{m_{jj}}\left\{\frac{1}{m_{jj}}\left(\frac{\partial m_{jj}}{\partial B}\right)^2 - \left(\frac{\partial^2 m_{jj}}{\partial B^2}\right)\right\} + \frac{k-j}{B^2}\right], \tag{10}$$

where $n_k = \sum_j a_{jk}$, the total number of households with k initial susceptibles.

3.1 *Starting-Point*

An estimate, B_k, of B is available from (4) and the frequencies of the zero class, a_{0k}, $k = 1, 2, \ldots, K$. We have

$$\frac{a_{0k}}{n_k} = m_{0k} = \hat{B}_k^k$$

and

$$\hat{B}_k = \left(\frac{a_{0k}}{n_k}\right)^{1/k}. \tag{11}$$

The estimate over all k is

$$\hat{B}_0 = \frac{1}{n}\sum_k n_k \left(\frac{a_{0k}}{n_k}\right)^{1/k}. \tag{12}$$

Estimation of B from (12) is quite efficient since most of the information about B is contained in a_{0k}. In general, the efficiency of the estimate increases as a_{0k}/n_k increases. An estimate of Q can be derived in a similar manner, where

$$\frac{a_{1k}}{n_k} = m_{1k} = k(1-\hat{B}_0)\hat{B}_0^{k-1}\hat{Q}_0^{k-1}$$

and

$$\hat{Q}_0 = \left\{\frac{a_{1k}}{kn_k(1-\hat{B}_0)\hat{B}_0^{k-1}}\right\}^{1/(k-1)}. \tag{13}$$

However, this estimate is not very efficient since it ignores the frequencies a_{jk}, $j = 2, \ldots, K$, which contain much of the information about Q. A more efficient estimator should use all the available information. Towards this end, an estimator for Q is derived from an approximating set of differential equations for (4), given in the Appendix. From (A5) in the Appendix,

$$\hat{Q}_1 \doteq \left(\frac{1-\bar{\theta}}{\hat{B}_0}\right)^{1/\bar{\phi}}, \tag{14}$$

where

$$\bar{\phi} = \frac{\sum_k \sum_j j a_{jk}}{n} \quad \text{and} \quad \bar{\theta} = \frac{\sum_k \sum_j (j a_{jk}/k)}{n}.$$

The value $\bar{\phi}$ is the average number of individuals infected per household, while the value $\bar{\theta}$ is the average fraction of individuals infected per household, which is sometimes called the 'household attack rate'.

The estimators given by (12) and (14) are used to obtain starting-values for the fully efficient ML estimators given by (6) and (7).

3.2 *The Truncated Case*

In some cases, the zero class a_{0k} is not present, since households are surveyed only after an initial infective has appeared. In this case, the zero-truncated distribution is used. The probability density function is then

$$m_{jk} = \binom{k}{j} m_{jj} B^{k-j} Q^{j(k-j)}/(1-B^k). \tag{15}$$

Application of the ML procedure to (15) is somewhat tedious but straightforward. The absence of the zero class makes it difficult to obtain a starting-point for B [see (12)]. Moreover, an estimate of a_{0k} would be useful for epidemiological reasons. A method for estimating a_{0k} is described by Irwin (1963), although the derivation was first proposed by McKendrick (1926). Griffiths (1973) used the method in fitting the truncated beta-binomial distribution of household data.

For application in this paper, an initial guess $\hat{a}_{0k}^{(0)}$ is made and the corresponding ML estimates $\hat{B}^{(0)}$ and $\hat{Q}^{(0)}$ are found from the nontruncated distribution (4). A new estimate $a_{0k}^{(1)}$ is found using the proportional allocation based on (4), such that

$$\hat{a}_{0k}^{(1)} = n_k'[(\hat{B}^{(0)})^k/\{1-(\hat{B}^{(0)})^k\}],$$

where $n_k' = \sum_{j=1} a_{jk}$. Then the next estimates, $\hat{B}^{(1)}$ and $\hat{Q}^{(1)}$, are found from $\hat{a}_{0k}^{(1)}$ and $\hat{B}^{(1)}$, after which $\hat{a}_{0k}^{(2)}$ can be calculated. The iterations are continued until

$$|\hat{a}_{0k}^{(r)} - \hat{a}_{0k}^{(r-1)}| < \varepsilon \tag{16}$$

for some small value of ε. Hartley (1958) has shown that the estimates $\hat{B}$ and $\hat{Q}$ obtained in this way are the ML estimates. He also gave methods for accelerating convergence.

If the disease in question is rare, or does not spread readily in the community, then the estimate for a_{0k} may approach infinity and the estimate for B will approach one. When this is the case, some upper limit should be placed on a_{0k}. This limit would logically be the number of households in the community from which the data is drawn, as suggested by Griffiths (1973).

3.3 *Computation*

In our experience, when the zero class is present, convergence of $\hat{Q}$ and $\hat{B}$ to an accuracy of 10^{-4} is quite rapid. For the data given in the next section, with starting-points given by (12) and (14), convergence usually occurred in three or four iterations. The asymptotic variances and covariance were readily obtained by inversion of the information matrix. However, when the zero class was absent, convergence on iterates of $\hat{a}_{0k}$ was somewhat slow. This problem was alleviated by forming geometric projections on $\hat{a}_{0k}$, as suggested by Hartley (1958). When the ML estimates for Q and B are found by iterating on $\hat{a}_{0k}$, estimates for the variances and covariance are not directly available from the information matrix. In order to get true variance estimates, the variation of $\hat{Q}$ and $\hat{B}$ due to iteration on $\hat{a}_{0k}$ must be taken into account.

4. Application to Data

4.1 *Influenza Epidemics*

The model was fitted to the Asian influenza epidemic household data previously examined by Sugiyama (1960). The data set is the only one considered for which the zero class is present. The distribution (4) (which is the final-size distribution for Sugiyama's model) fits the data quite well—as expected, since Sugiyama got an equally good fit. Results are given in Table 1. Since $\hat{Q} = .834$, the estimated probability of a susceptible individual being infected by a case during the course of his infectious period is .166. The average length of the latent and infectious periods for influenza are about two and four days, respectively (see Kilbourne, 1975). Therefore, in (2), $l = 2$ and $m = 6$. If it is assumed that $p_t = p$ for

Table 1
Observed and expected distributions of Asian influenza data of Sugiyama. Households of size three (zero class present)

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
$\hat{Q} = .834$, $\text{var}(\hat{Q}) = .0063$, $\text{cov}(\hat{Q}, \hat{B}) = .0004$	$\hat{B} = .886$ $\text{var}(\hat{B}) = .0009$	
$\chi^2(1\text{ df}) = 1.222$,	$(.25 < P < .50)$	

Table 2
Observed and expected distributions of influenza data of Hope-Simpson and Sutherland (zero class absent)

Number of cases	Households of size four*		Households of size five*	
	Observed	Expected	Observed	Expected
0	—	8.50†	—	2.47†
1	3	2.95	1	1.27
2	3	3.18	2	1.34
3	6	5.81	2	2.18
4	12	12.06	4	4.51
5	—	—	9	8.73
Total	24	24.00	18	18.03
	$\hat{Q}=.601$, $\hat{B}=.715$ $\chi^2(1\text{ df})=.018$ $(.75<P<.90)$		$\hat{Q}=.664$, $\hat{B}=.655$ $\chi^2(2\text{ df})=.463$ $(.75<P<.90)$	

*For the pooled data (i.e. households of size four and five together) the expected values are not given here. The fit was good, with $\chi^2(5\text{ df})=.918$ $(.950<P<.975)$; $\hat{Q}=.644$, $\hat{B}=.656$; and $\hat{a}_{04}=5.43$, $\hat{a}_{05}=2.48$.
†Not included in the total.

$t=3,\ldots,6$, then application of (3) yields $\hat{p}=.044$ as the estimate of the daily probability that an infectious individual will infect a susceptible family member within the household.

For infection from the community, $\hat{B}=.856$; the estimated probability that a susceptible individual will be infected from the community during the course of the epidemic is thus .144. The approximate percentage of cases due to community exposure can be calculated by setting Q to 1, which causes all cases to originate from the community. From (5), the expected number of cases would be

$$nk(1-\hat{B})=14.4,$$

while the number of cases allowing spread within the household (i.e. $\hat{Q}=.834$) is 19. Hence, 75% of the total cases were due to infection from the community. This percentage may be somewhat overestimated, however, since there will be some reduction in susceptibles due to spread within households (i.e. the number of susceptibles is actually less than nk). Kemper (1980) discusses the errors in calculations concerning secondary spread when the apparent number of susceptible individuals is less than the actual number.

The model was also fitted to the influenza data of Hope-Simpson and Sutherland (1954). In this case, the zero classes are missing and the method of §3.2 is used. The fit (shown in Table 2) is excellent for households of size four and five, both separately and pooled. Hope-Simpson and Sutherland got a good fit using the classical Reed–Frost model; however, there were unanswered questions as to the extent of community involvement in the epidemic. The estimates of a_{0k} were quite small, indicating that probably most houses were invaded during the epidemic, if the samples of households given were indeed representative of the community. Using the pooled estimates, it is calculated that, at maximum, 52% of the total cases were due to infection from the community.

In general, community-acquired infection has been shown to play an important role in the spread of influenza. This was particularly true of Asian influenza, where the schools served as foci of infection. It was estimated by Elveback *et al.* (1976) that almost all the apparent secondary spread among school children within the household was due to infection acquired in the schools. In addition, about half the apparent secondary spread among adults in the household was due to mixing in the community.

With regard to spread within the household, $\hat{p} = .097$ for the influenza epidemic studied by Hope-Simpson and Sutherland, while $\hat{p} = .044$ for the influenza epidemic studied by Sugiyama. These results indicate that the agent of the former epidemic was more infectious in the household than that of the latter epidemic. Aside from agent characteristics, cultural and population differences could account for variations in household infectiousness.

4.2 *The Common Cold*

The data on outbreaks of the common cold were collected over a two-year period by Brimblecombe *et al.* (1958). Households of size five were partitioned into three levels of domestic crowding, depending on the number of rooms occupied by the family. Table 3 shows that the fit is acceptable at all three levels of domestic crowding. Note that the estimates of the zero classes approach the total size of the community except in the case of uncrowded households where the limit may be somewhat less. This would indicate that community-acquired cases serve only as index cases and that subsequent spread is largely confined to the household. However, homogeneity and ascertainment bias could also account for this result (see the last part of §5).

For the common cold, the average lengths of the latent and infectious periods are about three and seven days, respectively [see Monto (1976) for colds caused by coronavirus]. Assuming that $p_t = p$ for $t = 4, 5, \ldots, 10$, the estimates of p at each level of domestic

Table 3

Observed and expected distributions of common cold data of Brimblecombe et al. Households of size five (zero class absent)

Number of cases	Degree of domestic crowding					
	Uncrowded		Crowded		Overcrowded	
	Observed	Expected	Observed	Expected	Observed	Expected
0	—	6000.00*	—	Upper limit†	—	Upper limit†
1	156	156.44	155	143.28	112	104.74
2	55	52.61	41	53.87	35	40.74
3	19	22.36	24	27.15	17	21.44
4	10	8.55	15	12.79	11	10.64
5	2	2.04	6	3.91	6	3.45
Total	242	242.00	241	241.00	181	181.01
	$\hat{Q} = .900$, $\hat{B} = .992$		$\hat{Q} = .878$, $\hat{B} = .999$		$\hat{Q} = .872$, $\hat{B} = .999$	
	$\chi^2(2\text{ df}) = .864$		$\chi^2(2\text{ df}) = 5.91$		$\chi^2(2\text{ df}) = 4.12$	
	$(.50 < P < .75)$		$(.05 < P < .10)$		$(.10 < P < .25)$	

*Not included in the total.

†The total number of households in the community.

crowding are as follows:

Uncrowded: $\hat{p} = .015$,
Crowded: $\hat{p} = .018$,
Overcrowded: $\hat{p} = .028$.

These results indicate that there is an increase in disease spread within the household with increasing levels of domestic crowding.

5. Discussion

McKendrick (1926) did the first mathematical investigation of household versus community acquisition of infection; he was concerned with epidemics of bubonic plague. Although he was not able to estimate parameters measuring infectious contact, he did estimate the ratio of the probability of household-acquired to that of community-acquired infection. He concluded that this ratio was 200 : 1 for plague. Sugiyama (1960) first parameterized the Reed–Frost model to deal with household and community-wide infection. Kemper (1980) formulated a model which parameterizes the effect of asymptomatic infections as well as household and community-wide infection. He showed how the presence of asymptomatic infectives within the household, and of community-wide infection (beyond the index case), can be a source of error in the calculation of secondary attack rates.

Elveback *et al.* (1976) developed a stochastic simulation model which includes infectious-contact probabilities not only for the household and community, but also for other important mixing groups such as pre-school playgroups, schools and neighborhood clusters. However, their model does not allow for statistical parameter estimation or calculation of variances from data.

The Reed–Frost model has been extensively applied to household data on infectious disease. Bailey (1975, Chapter 14) gives a detailed account of these efforts. Aside from the fact that the classical Reed–Frost model does not recognize community-wide infection, there are other difficulties. The Reed–Frost model assumes that the disease being studied is characterized by a constant latent period and a very short infectious period. Although these assumptions roughly hold for such diseases as measles and mumps, others, such as influenza and the common cold, have longer infectious periods than latent periods. For these diseases, the task of separating out chains of infection is quite difficult since the infectious periods of successive generations of cases will tend to overlap. This problem is compounded when individuals are being infected from outside the household, making specific chains difficult to identify, even when the Reed–Frost model assumptions are satisfied. When it is difficult or impossible to distinguish chains of infection, it is reasonable to use the final-size distribution of cases within households to estimate parameters. An even more important reason to use the final-size distribution has to do with data acquisition. Although symptom data (i.e. cases) were used as examples in this paper, infection data, which also include asymptomatic infections but exclude clinical syndromes not caused by the infection, provide a more accurate description of disease transmission. Serological measures are the most efficient form of infection information for a population. Blood samples taken before and after an epidemic period can be used to determine which household members become infected during the course of the epidemic as well as to establish initial levels of susceptibility. Such information provides the final-size distributions of household infections, but does not provide information on the time of onset or duration of infection for individuals.

Although the model of this paper fits the data presented quite well, others have had equal success with different models. Griffiths (1973) was able to get excellent fits to the same data, as well as to other data sets, using the beta-binomial distribution, the assumption being that a good fit indicates that the disease is not infectious. Heasman and Reed (1961) were able to get a good fit to the common cold using the Reed–Frost model. Becker (1980) got an even better fit to the same data using a generalized model of household infection in which the probability of infectious contact varied according to the beta distribution. In neither study was infection from the community beyond the index cases considered, and, indeed, there may be little community spread of infection in the case of the common cold (see §4.2). Spicer, in his discussion on Bailey (1955), raised the question of choice among different mathematical models which all explain the data. He concluded that further independent epidemiological evidence was needed as a basis for choice. Accordingly, the model proposed here is testable by examination of how the community and household parameters vary with measurable environmental factors. Reciprocally, it can be used to identify community and family transmission factors that are correlated with family parameters. For example, a high probability of infectious contact from the community, for certain strains of influenza, has suggested that immunization of school children may be an effective control strategy (see Elveback *et al.*, 1976). This strategy was actually carried out for the Hong Kong strain of inflennza (see Monto *et al.*, 1969) and was shown to be efficacious.

Finally, the data used in this paper have certain limitations. First, not all the households in the common cold study were under surveillance for the same period of time. Therefore, the probability of community acquisition of infection will not necessarly be homogeneous across all households. This heterogeneity could artificially decrease the parameter of infectious contact from the community and increase the parameter of infectious contact within the household. Another problem has to do with ascertainment bias. By inclusion only of households with cases, there may be a tendency to overrepresent those households in which the disease spreads easily. This will affect the parameters in the same direction as heterogeneity. Both of these limitations in the data presented here can be overcome by different field procedures.

Acknowledgements

The authors are grateful to Ken Guire, who did the programming for the computation involved in this work. The research was partially supported by the National Institutes of Health under National Research Service Award A1-6103-01.

Résumé

On propose une modélisation de la distribution du nombre total des cas d'infection dans les familles d'une communauté homogène. Dans le modèle, une infection provenant de la communauté sert de source initiale d'infection à l'intérieur des familles aussi bien que pour des cas d'infection ultérieurs éventuels. De plus des membres infectés d'une famille peuvent infecter d'autres personnes de la famille. On donne des procédures de maximum de vraisemblance pour estimer les paramètres du modèle. Le modèle est ajusté aux données symptotiques de la grippe et du rhume ordinaire. La grippe semble se propager plus facilement dans la communauté qu'à l'intérieur des foyers, tandis que ce pourrait être le contraire pour le rhume ordinaire. Le modèle n'exige pas que soit précisé l'instant de début d'infection pour les individus et peut être ajusté à des données sérologiques. Ceci fournirait une mesure plus précise de l'infection des familles que l'utilisation des données symptotiques.

References

Bailey, N. T. J. (1955). Some problems in the statistical analysis of epidemic data. *Journal of the Royal Statistical Society, Series B* **17,** 35–68.

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Disease and its Applications*, 2nd ed. New York: Hafner.

Becker, N. (1980). An epidemic chain model. *Biometrics* **36,** 249–254.

Brimblecombe, F. S. W., Cruickshank, R., Master, F. L., Reid, D. D. and Stewart, G. T. (1958). Family studies of respiratory infections. *British Medical Journal* **1,** 119–128.

Elvebeck, L. R., Fox, J. P., Ackerman, E., Langworthy, A., Boyd, M. and Gatewood, L. (1976). An influenza simulation model for immunization studies. *American Journal of Epidemology* **103,** 152–165.

Griffiths, D. A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of disease. *Biometrics* **29,** 637–648.

Hartley, H. O. (1958). Maximum likelihood estimates from incomplete data. *Biometrics* **14,** 174–194.

Heasman, M. A. and Reid, D. D. (1961). Theory and observation in family epidemics of the common cold. *British Journal of Preventive and Social Medicine* **15,** 12–16.

Hethcote, H. W. and Waltman, P. (1973). Optimal vaccination schedules in a deterministic epidemic model. *Mathematical Biosciences* **18,** 365–381.

Hope-Simpson, R. E. and Sutherland, I. (1954). Does influenza spread within the household? *Lancet* **1,** 721–726.

Irwin, J. O. (1963). The place of mathematics in medical and biological statistics. *Journal of the Royal Statistical Society, Series A* **126,** 1–44.

Kemper, J. T. (1980). Error sources in the evaluation of secondary attack rates. *American Journal of Epidemiology* **112,** 457–464.

Kilbourne, E. D. (1975). *The Influenza Viruses and Influenza.* New York: Academic Press.

Ludwig, D. (1975). Final size distributions for epidemics. *Mathematical Biosciences* **23,** 33–46.

McKendrick, A. G. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society* **44,** 98–130.

Monto, A. S. (1976). Coronaviruses. In *Viral Infections of Humans*, A. S. Evans (ed.), 127–141. New York: Plenum.

Monto, A. S., Davenport, F. M., Napier, J. A. and Francis, R. (1969). Effect of vaccination of a school age population upon the course of an A2/Hong Kong influenza epidemic. *Bulletin of the World Health Organization* **41,** 537–542.

Sugiyama, H. (1960). Some statistical contributions to the health sciences. *Osaka City Medical Journal* **6,** 141–158.

Received September 1980; *revised December* 1980

Appendix

Approximating Differential Equations

Let $S(t)$, $I(t)$ and $R(t)$ be continuous variables for the number of susceptible, infected and removed (immune) individuals, respectively, within the household at time t. Household members mix homogeneously and make infectious contact with one another at rate p per time unit. They also make infectious contact with the community at rate $a(t)$, at time t. Since the epidemic is of finite duration, we have

$$\int_0^\infty a(w)\,dw = a < \infty. \tag{A1}$$

Also, infected individuals are removed at the proportionality rate γ, where $\gamma^{-1} = \bar{\alpha}$ is the average of the infectious period.

The initial-value problem from the above assumption is

$$\begin{aligned} \frac{dS(t)}{dt} &= -S(t)\{a(t) + pI(t)\}, \\ \frac{dI(t)}{dt} &= \frac{-dS(t)}{dt} - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t), \end{aligned} \tag{A2}$$

with $S(0) = k > 0$, $I(0) = 0$, $R(0) = 0$ and $S(t) + I(t) + R(t) = k$. From (A1) and (A2) it is easy to show [using arguments similar to those of Hethcote and Waltman (1973)] that the limiting values $S(\infty)$, $I(\infty)$ and $R(\infty)$ exist and that $S(\infty) \geq 0$, $I(\infty) = 0$ and $R(\infty) > 0$.

Substitution of the third into the first equation of (A2) yields

$$\frac{dS(t)}{S(t)} = -\{a(t)\,dt + p\bar{\alpha}R(t)\,dt\},$$

which has the solution

$$S(t) = k \exp\left[-\left\{\bar{\alpha}pR(t) + \int_0^t a(w)\,dw\right\}\right]. \tag{A3}$$

The final values are defined as $S(\infty) = k - \phi$, $I(\infty) = 0$ and $R(\infty) = \phi$ (see §3.1). Taking the limit of (A3) as $t \to \infty$ yields

$$1 - \theta = \exp(-\bar{\alpha}p\phi)\exp(-a), \tag{A4}$$

where $\theta = \phi/k$.

The right-hand side of (A4) is composed of two parts. The term $\exp(-a)$ is an approximation for the probability of a susceptible individual escaping infection from the community. The term $\exp(-\bar{\alpha}p\phi)$ is an approximation for the probability that a susceptible individual escapes infection from infected household members. Substituting $\hat{Q}, \hat{B}, \bar{\theta}$ and $\bar{\phi}$ from §3.1 into (A4) yields

$$1 - \bar{\theta} \doteq \hat{Q}\bar{\phi}\hat{B}.$$

Since an independent estimate $\hat{B}_0$ of B is available, Q is estimated by

$$\hat{Q}_1 \doteq \left(\frac{1-\bar{\theta}}{\hat{B}_0}\right)^{1/\bar{\phi}}. \tag{A5}$$