Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 17(2): 1889–1900.
DOI: 10.3934/mbe.2020099
Received: 09 October 2019
Accepted: 04 December 2019
Published: 19 December 2019

Research article

The spread of influenza-like-illness within the household in Shanghai, China

Meili Li[1], Hong Wang[1], Baojun Song[2] and Junling Ma[3,*]

[1] College of Science, Donghua University, Shanghai 201620, China

[2] Department of Applied Mathematics and Statistics, Montclair State University, Montclair, NJ 07043, USA

[3] Department of Mathematics and Statistics, University of Victoria, Victoria, BC V8W 2Y2, Canada

* **Correspondence:** Email: junlingm@uvic.ca.

Abstract: High-density urban habitats provide a hotbed for the rapid spread of infectious diseases. School children densely aggregate in classrooms. So schools are high incidence area of infectious diseases. This paper aims at investigating the transmission of influenza-like-illness within households with a school child using a survey study of fourth grade elementary school students in Shanghai, China. We found that the pairwise transmission probability within a household is only 0.172, which implies that the average number of infections caused by a single infectious individual in a household in Shanghai is only 0.304. Thus, the majority of transmission must occur outside of a household for a disease to cause an outbreak.

Keywords: household transmission; pairwise transmission probability; influenza-like-illness; Markov Chain

1. Introduction

With the acceleration of global urbanization, high-density urban human habitats provided accelerated spread of infectious diseases in these highly densely populated areas. This leads to new challenges for public health management. We all have the experience that children were infected with an influenza-like-illness (ILI) in school, and then spread the disease to their parents, who then spread it among their coworkers, demonstrating a clear hierarchy of transmission, and rendering the commonly assumed randomly mixed population unrealistic.

In the study of infectious diseases of humans, the household was historically chosen as one of the pivotal units not only because of intimate contacts within the household members but also because of household members sharing living arrangements. In 1952, Simpson [1] took households as the unit to study the infectiousness of chickenpox, mumps, measles. Epidemiological data on the infectiousness in

relation to size of family was obtained. The role of household size is same as the role of community size in determining whether an infectious disease can take off or can become endemic. It was confirmed that the spread and establishment of infectious diseases in a community not only depends on connectivity of the community but on its size as well. For instance, McGrath estimated that a social network of 180 to 440 persons is required to achieve the stable host pathogen relationship necessary for tuberculosis infection to become endemic in a community [2].

To study household transmissions, we need to use a mathematical model that describes household transmissions. Due to the small size of households, many household models are stochastic in nature [3–12]. Deterministic models are used to study the transmission dynamics of tuberculosis on generalized households [13, 14]. Chao et al. [15] developed a random simulation platform FLU based on individual behavior, which also considers the spread of population contact during the day and the spread of household contact at night. Nichols et al. [16] found that large household size is important for the varicella-zoster virus transmission in Guinea Bissau. Households are also incorporated into contact network models [17, 18].

For household transmissions of influenza, Longini et al. [12] first provided a maximum likelihood estimation for probability of being infected in a household using data from a study in Seattle, WA, USA for the 1975/76, 1977/78 and 1978/79 seasons. They estimated that, for influenza A/H3N2 or B seasonal epidemics, the probability that a household member is infected in a household is about 17% for families with children and 13% for families without children. Their study, however, assumed that probability of being infected in a household is independent of the household size. Cauchemez et al. [6] applied MCMC approach to a longitudinal study for the 1999/2000 season in France in order to estimate the transmission rate and the mean infectious period within a household. They made the same assumption that the transmission rate within the household is independent of its size. By assuming that the transmission rate between any pair is constant, Cauchemez et al. [19] studied transmission of influenza A/H1N1. It was found that the observed household transmissions are best explained by a model assuming a transmission rate being proportional to family size. The review paper on the household transmission studies by Tsang et al. [20] is very informative.

In this paper, we aim to study household transmissions in Shanghai, China, using a survey study among fourth-grade elementary school children, and investigate whether household transmission is a crucial factor of ILI transmissions in large cities like Shanghai. We use the findings of Cauchemez et al. [19] and assume that the transmission probability from any susceptible individual to any susceptible individual is the same and independent to any other pairs.

In Section 2, we derive the probabilities for a given number of individuals to escape a household outbreak, and the average number of infections occur in a household with a single initially infectious individual. In Section 3, we use the derived probabilities to estimate the pairwise transmission probability in a household. In Section 4, we estimate the average number of household transmissions in a random household in Shanghai using Shanghai Census data. In Section 5, we give some concluding remarks.

2. Model of within-household transmission

We consider a stochastic SIR epidemiological model for a household of n homogeneously mixed individuals; and one individual is initially infected. We only keep track of the numbers of susceptible

and infected individuals of the household.

The transmission rate between a pair of an infectious individual and a susceptible individual is β, and the recovery rate of an infectious is γ. The states of the system are labeled as (s, i), where s is the number of susceptible individuals of the household and i is the number of infectious individuals of the household. The set of all possible states is $\{(s, i) : 0 \le s + i \le n\}$. Obviously, $(s, 0)$ $(s \le n - 1)$ are all absorbing states. We systemically arrange states by the following order (i.e., in decreasing orders of $s + i$ and then s):

$$(n-1,1),(n-2,2),(n-3,3),\ldots,(0,n),(n-2,1),(n-3,2),\ldots,(0,n-1),(n-3,1),$$
$$(n-4,2),\ldots,(0,n-2),\ldots,(1,1),(0,2),(0,1),(n-1,0),(n-2,0),\ldots,(1,0),(0,0).$$

Markov chain From a transient state (s, i) (i.e., $i > 0$), the system can transfer either to $(s - 1, i + 1)$ because of an infection event occurring at the rate βsi, or to $(s, i - 1)$ because of a recovery event occurring at the rate γi. We consider the embedded discrete time Markov chain, i.e., each time step corresponds to a state change. Then the system transfers from state (s, i) to $(s, i-1)$ with the probability

$$T_{(s,i)\to(s,i-1)} = \frac{\gamma i}{\beta si + \gamma i} = \frac{\gamma}{\beta s + \gamma},$$

and to $(s - 1, i + 1)$ with the probability

$$T_{(s,i)\to(s-1,i+1)} = \frac{\beta si}{\beta si + \gamma i} = \frac{\beta s}{\beta s + \gamma}.$$

All transient states precede all the absorbing states, thus, the transition matrix of the embedded discrete time Markov chain is

$$T = \begin{bmatrix} Q_n & U_n & 0 & & R_n \\ & Q_{n-1} & U_{n-1} & & R_{n-1} \\ & & \ddots & & \vdots \\ & & & Q_1 & R_1 \\ & & & & I \end{bmatrix}, \tag{2.1}$$

where I is the $n \times n$ identity matrix,

$$Q_j = \begin{bmatrix} 0 & \frac{(j-1)\beta}{(j-1)\beta+\gamma} & & & \\ & 0 & \frac{(j-2)\beta}{(j-2)\beta+\gamma} & & \\ & & \ddots & & \\ & & & 0 & \frac{\beta}{\beta+\gamma} \\ & & & & 0 \end{bmatrix}, U_j = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{\gamma}{(j-2)\beta+\gamma} & & & \\ & \frac{\gamma}{(j-3)\beta+\gamma} & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

and R_j is a $j \times n$ $(j = 1, 2, \ldots, n)$ matrix which u, v entry is

$$(R_j)_{(u,v)} = \begin{cases} \frac{\gamma}{(j-1)\beta+\gamma}, & u = 1, v = n - j + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Pairwise transmission probability Our model has two parameters, the transmission rate β and the recovery rate γ. However, the transition matrix T is determined by a single parameter

$$q = \frac{\beta}{\beta + \gamma}, \tag{2.2}$$

which is the pairwise transmission probability between an infectious individual and a susceptible one (i.e., transmission between the pair occurs before the infectious individual recovers, ignoring other infectious individuals in the family). This probability plays a fundamental role in this study.

Note that, for any k,

$$\frac{k\beta}{k\beta + \gamma} = \frac{kq}{(k-1)q+1}, \ \frac{\gamma}{k\beta + \gamma} = \frac{1-q}{(k-1)q+1}.$$

Thus, Q_j, U_j and R_j can be rewritten as

$$Q_j = \begin{bmatrix} 0 & \frac{(j-1)q}{(j-2)q+1} & & & \\ & 0 & \frac{(j-2)q}{(j-3)q+1} & & \\ & & \ddots & & \\ & & & 0 & q \\ & & & & 0 \end{bmatrix}, \tag{2.3}$$

$$U_j = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{1-q}{(j-3)q+1} & & & \\ & \frac{1-q}{(j-4)q+1} & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, \tag{2.4}$$

$$(R_j)_{(u,v)} = \begin{cases} \frac{1-q}{(j-2)q+1}, & u = 1, v = n-j+1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

Number of individuals escaped the epidemic The states $(s, 0)$ for $s = 0, 1, \ldots, n-1$, that s individuals escaped all household infections, is absorbing. We want to compute the absorbing probability of these states. For a finite state discrete time Markov chain, where the states can be divided into transient states and absorbing states, let

$$T = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} \tag{2.6}$$

be its transition matrix, where Q is the transition probabilities within transient states, and R is the transition probabilities from the transient states to the absorbing states. At time step k, the probability that the system stays in state j starting from the state i, is the (i, j) entry of

$$T^k = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}^k = \begin{bmatrix} Q^k & \sum_{i=0}^{k-1} Q^i R \\ 0 & I \end{bmatrix}.$$

Thus,

$$\lim_{k\to\infty} T^k = \begin{bmatrix} 0 & (I-Q)^{-1}R \\ 0 & I \end{bmatrix}. \tag{2.7}$$

Thus, starting from a transient state i, the system eventually is absorbed in a state j with a probability which is the (i, j) entry of $(I - Q)^{-1}R$.

To apply this formula to our model, we let $D_j = I_j - Q_j$ $(j = 1, 2, \ldots, n)$. Then,

$$P = (I-Q)^{-1}R = \begin{bmatrix} D_n & -U_n & & 0 \\ & D_{n-1} & -U_{n-1} & 0 \\ & & \ddots & \vdots \\ & & & D_1 \end{bmatrix}^{-1} \begin{bmatrix} R_n \\ R_{n-1} \\ \vdots \\ R_1 \end{bmatrix}.$$

Starting from a single initially infectious individual, the probability that $s = 0, \ldots, n-1$ individuals escaped infection is the entry in the row corresponding to the initial state $(n-1, 1)$ (in our setup this is row 1) and the column corresponding to the state $(s, 0)$:

$$D_n^{-1} \sum_{k=1}^{n} \left[\prod_{j=n-k+1}^{n-1} U_{2n-k-j+1} D_{2n-k-j}^{-1} \right] R_{n+1-k}. \tag{2.8}$$

Detailed calculations to obtain (2.8) can be found in Appendix A. Since the number of infected individuals is a random variable, we are interested in its expected value. Let Z_n be the expected number of people infected by a single infectious individual in a household of size n. Then

$$Z_n = \sum_{s=0}^{n-1} (n-s-1) P_{(n-1,1)\to(s,0)}(q), \tag{2.9}$$

where the probability $P_{(n-1,1)\to(s,0)}(q)$ $(s = 0, 1, \ldots, n-1)$ is the corresponding entry in P, meaning that the system stops at an absorbing state $(s, 0)$ starting from a transient state $(n-1, 1)$. For various household sizes, the dependence of Z_n on the transmission probability q are shown in Figure 1. One can observe that Z_n is an increasing function of q for a given n; and it is an increasing function of n for any given q.

3. Maximum likelihood estimation

In this section, we will estimate the transmission possibility q from a survey study of the ILI within the households.

Data We conducted a survey among fourth graders of Shanghai Xuejia Elementary School in Shanghai, China, in April 2018. From this survey, we use the results of three questions that are related to ILI among households of the students are

(a) How many people (including yourself) live in your household?
(b) Have you caught an influenza-like-illness since January 1, 2018?
(c) If you answered "Yes" in (b), how many other people (including yourself) have caught the same illness?

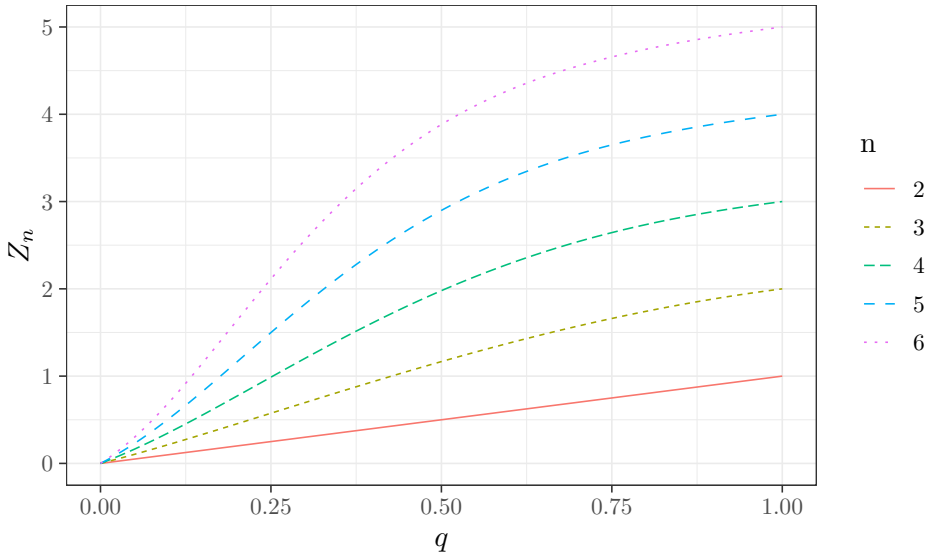

Figure 1. The average number of household infections as a function of the pairwise transmission probability q, for various values of the household size n.

We distributed 100 questionnaires in two classes, and collected 78 answers, among which 74 contained answers for Question (a). The household size distribution is plotted in Figure 2. Not all answers are valid for our research. Specifically, some do not contain a valid answer (either an answer is blank or the answer to Question (c) is great than the household size), and some answered "greater than or equal to 7" for Question (c) but specific size is not available. There are 56 valid answers, which are listed in Table 1. Among valid answers, 9 students answered "no" to Question (b) (who also subsequently answered 0 to Question (c)), and thus cannot be be used to study household-transmissions. The distribution of the number of ILI infections (i.e., answer to Question (c)) is plotted in Figure 3.

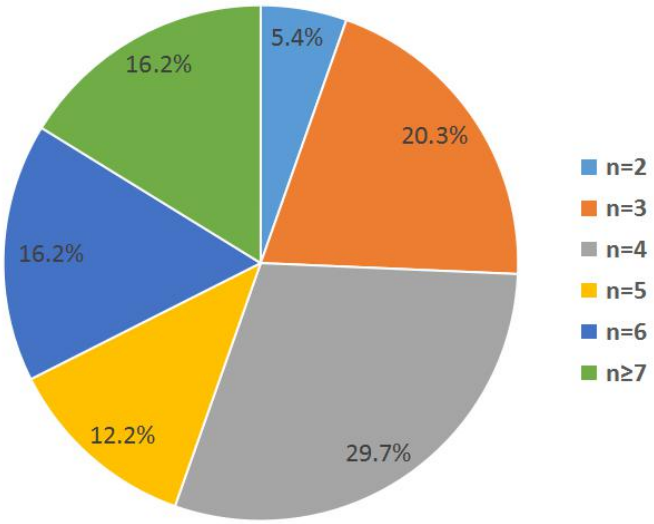

Figure 2. The household size distribution.

Table 1. Valid questionnaire responses, i.e., all questions are answered, and the answer to (c) is less than or equal to that of to (b). Note that some answers indicate that no ILI occurred in that family, and cannot be used in this study.

Student	Household size (a)	Infections (c)	Student	Household size (a)	Infections (c)
1	3	1	29	4	4
2	6	3	30	3	2
3	6	1	31	4	3
4	3	3	32	5	3
5	4	1	33	4	1
6	3	2	34	4	1
7	6	2	35	4	0
8	4	1	36	3	1
9	6	2	37	4	1
10	3	1	38	3	2
11	4	1	39	5	1
12	2	1	40	3	2
13	5	2	41	3	1
14	4	1	42	5	1
15	4	4	43	6	0
16	6	5	44	5	2
17	3	2	45	6	0
18	5	1	46	4	0
19	2	1	47	4	1
20	4	0	48	4	2
21	5	2	49	4	1
22	5	1	50	4	1
23	6	1	51	4	3
24	6	0	52	6	1
25	3	2	53	6	3
26	3	0	54	4	0
27	3	0	55	4	3
28	3	1	56	4	1

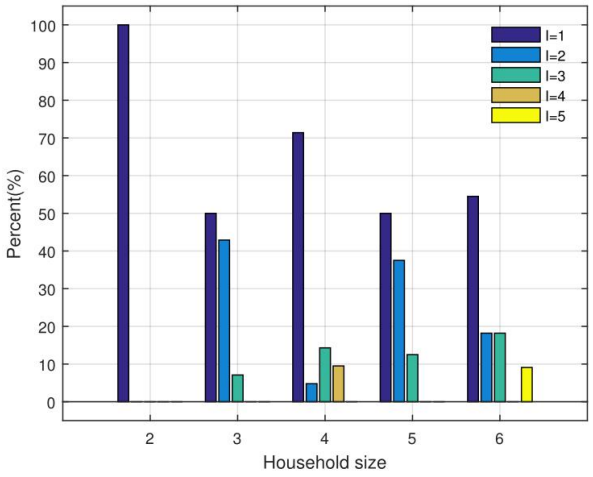

Figure 3. The distribution of infections within a household (including the index case) for each household size $n = 2, 3, \cdots, 6$.

Estimation We use the maximum likelihood estimation method to estimate q. For the ith observation with a household size N_i and I_i infections in the household, the probability of this observation is $P_{(N_i-1,1)\to(N_i-I_i,0)}(q)$. Thus, we construct the following likelihood function

$$L(q) = \prod_{i=1}^{47} P_{(N_i-1,1)\to(N_i-I_i,0)}(q). \tag{3.1}$$

We find q that maximizes $L(q)$. Equivalently, it also maximizes

$$\ell(q) = \ln L(q) = \sum_{i=1}^{47} \ln P_{(N_i-1,1)\to(N_i-I_i,0)}(q). \tag{3.2}$$

We then use the likelihood ratio test [21] to estimate the 95% confidence interval. The estimated transmission probability is

$$q = 0.172,\ 95\% \text{ confidence interval } (0.120, 0.237). \tag{3.3}$$

4. Estimating the household transmission of ILI in Shanghai, China

With the estimation of the pairwise transmission probability q, and the household size distribution in Shanghai, China, we can then estimate the average number of household ILI infections Z caused by a single infectious individual in Shanghai. We obtain the household size distribution in Shanghai from the 2010 Shanghai Census data [22], which is summarized in Table 2 and Figure 4. Specifically,

$$Z = \sum_{n=1}^{\infty} p_n Z_n\,, \tag{4.1}$$

where Z_n is defined in (2.9), and p_n is the fraction of size-n households in Shanghai. With the estimated q in (3.3),

$$Z = 0.304,\ 95\% \text{ confidence interval } (0.205, 0.431). \tag{4.2}$$

Table 2. The household size distribution in Shanghai, China. Data was obtained from "Tabulation on the 2010 Population Census of Shanghai Municipality" [22].

Household size (n)	Number of households
1	1641920
2	2805535
3	2549205
4	732700
5	424518
6	72296
7	17988
8	6091
9	1739
≥10	1265
Total	8253257

5. Conclusion

We use a stochastic household transmission model to estimate the pairwise transmission probability in a household in Shanghai, China. The estimated probability is $q = 0.172$ with a 95% confidence interval $(0.120, 0.237)$. This is not a large probability. Using the household distribution obtained from the 2010 Shanghai Census data, we estimated that the average number of people infection in the same household of a random infectious individual in Shanghai is 0.304 with a 95% confidence interval $(0.205, 0.431)$. This means that, if an ILI is able to cause an epidemic (with a basic reproduction number at least 1) in a randomly mixed population with households (i.e., outside-household contacts are random), each infectious individual must infected more than 0.696 individuals outside of his/her household at the beginning of the epidemic. This means that the majority of infections occur outside of the household. However, from our study, only $17/76$ school children did not catch an ILI, meaning that the majority of the children were infected. This suggests that ILI are driven by the infections in school.

We assumed that all infections except the first one in the households are caused by the first infected individual in the household. This is reasonable as household members have much closer contacts than others. If this is not true, then our estimated pairwise transmission probability q is an over-estimate, because we attributed external infections to household members. Thus, our conclusion that household average transmission is small still holds.

We also assumed that the transmission probability is independent to age, and that the household members are randomly mixed, so that the transmission probability between any pair of individuals is the same. However, as suggested by Longini et al. [12], this may not be true, but the difference in age-specific transmission probabilities may be small.

We used the survey data from Shanghai. However, we believe that the data is typical for household transmission in all large Chinese cities. It would be interesting to conduct such surveys in other cities world wide, and compare household transmissions in different cultures.

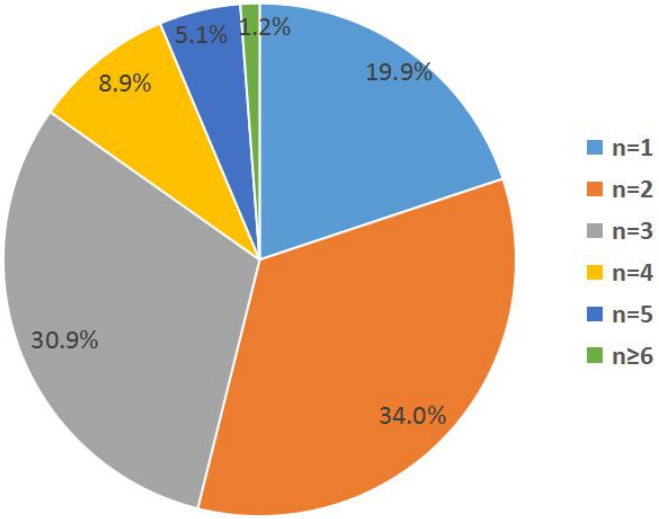

Figure 4. The household size distribution in Shanghai, China.

Acknowledgments

We thank the anonymous reviewers for their constructive comments. This research was partically supported by National Natural Science Foundation of China (No.11771075) (ML), State Scholarship Fund of China (CSC No.201906635011) (ML), and Natural Sciences and Engineering Research Council Canada discovery grant (JM). ML's work was carried out when she was a visiting scholar at Montclair State University.

Conflict of interest

The authors declare no conflicts of interest.

References

1. R. E. H. Simpson, Infectiousness of communicable diseases in the household, *Lancet.*, **2** (1952), 549–554.
2. J. W. McGrath, Social network of disease spread in the lower Illinios valley: a simulation approach, *Am. J. Phys. Anthropol.*, **77** (1988), 483–496.
3. H. Andersson, T. Britton, Stochastic epidemic models and their statistical analysis, Springer Science & Business Media, 2012.
4. F. Ball, Stochastic and deterministic models for SIS epidemics among a population partitioned into households, *Math. Biosci.*, **156** (1999), 41–67.
5. T. Britton, Stochastic epidemic models: A survey, *Math. Biosci.*, **225** (2010), 24–35.
6. S. Cauchemez, F. Carrat, C. Viboud, A. J. Valleron, P. Y. Boelle, A Bayesian MCMC approach to study transmission of infuenza: application to household longitudinal data, *Stat. Med.*, **23** (2004), 3469–3487.

7. P. J. Dodd, N. M. Ferguson, Approximate disease dynamics in household-structured populations, *J. R. Soc. Interface.*, **4** (2007), 1103–1106.

8. J. T. Wu, S. Riley, C. Fraser, G. M Leung, Reducing the impact of the next influenza pandemic using household-based public health interventions, *PLoS Med.*, **3** (2006), 1532–1540.

9. T. House, M. J. Keeling, Household structure and infectious disease transmission, *Epidemiol. Infect.*, **137** (2009), 654–661.

10. C. Fraser, Estimating individual and household reproduction numbers in an emerging epidemic, *PLoS One.*, **2** (2007), e758.

11. W. Mahikul, L. J. White, K. Poovorawan, N. Soonthornworasiri, P. Sukontamarn, P. Chanthavilay, et al., Modeling household dynamics on respiratory syncytial virus (RSV), *PLoS One.*, **14** (2009), e0219323.

12. I. M. Longini, J. S. Koopman, A. S. Monto, J. P. Fox, Estimating household and community transmission parameters for influenza, *Am. J. Epidemiol.*, **115** (1982), 736–751.

13. J. P. Aparicio, A. F. Capurro, C. Castillo-Chavez, Transmission and dynamics of tuberculosis on generalized households, *J. Theor. Biol.*, **206** (2000), 327–341.

14. B. Song, C. Castillo-Chavez, J. P. Aparicio, Tuberculosis models with fast and slow dynamics: the role of close and casual contacts, *Math. Biosci.*, **180** (2002), 187–205.

15. D. L. Chao, M. E. Halloran, V. J. Obenchain, I. M. Longnini, FluTE, a publicly available stochastic influenza epidemic simulation model, *PLoS Comput. Biol.*, **6** (2010), e1000656.

16. R. A. Nichols, K. T. Averbeck, A. G. Poulsen, M. M. Bassam, F. Cabral, P. Aaby, et al., Household size is critical to varicella-zoster virus transmission in the tropics despite lower viral infectivity, *Epidemics.*, **3** (2011), 12–18.

17. E. M. Volz, J. C. Miller, A. Galvani, L. A. Meyers, Effects of heterogeneous and clustered contact patterns on infectious disease dynamics, *PLoS Comput. Biol.*, **7** (2011), e1002042.

18. J. Ma, P. van den Driessche, F. H. Willeboordse, Effective degree household network disease model, *J. Math. Biol.*, **66** (2012), 75–94.

19. S. Cauchemez, C. A. Donnelly, C. Reed, A. C. Ghani, C. Fraser, C. K. Kent, et al., Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States, *N. Engl. J. Med.*, **361** (2009), 2619–2627.

20. T. K. Tsang, L. L. H. Lau, S. Cauchemez, B. J. Cowling, Household transmission of influenza virus, *Trends Microbiol.*, **24** (2016), 123–133.

21. B. M. Bolker, Ecological models and data in R, Princeton University Press, 2008.

22. 2010 Shanghai Census Information, 2012. Available from: http://www.stats-sh.gov.cn/html/huibian/indexch.htm.

Appendix

A. Calculation of absorbing probabilities

$$
\begin{aligned}
P &= \begin{bmatrix} D_n & -U_n & & 0 \\ & D_{n-1} & -U_{n-1} & 0 \\ & & \ddots & \vdots \\ & & & D_1 \end{bmatrix}^{-1} \begin{bmatrix} R_n \\ R_{n-1} \\ \vdots \\ R_1 \end{bmatrix} \\
&= \begin{bmatrix} I_n & -D_n^{-1}U_n & & \\ & I_{n-1} & -D_{n-1}^{-1}U_{n-1} & \\ & & \ddots & -D_2^{-1}U_2 \\ & & & I_1 \end{bmatrix}^{-1} \begin{bmatrix} D_n & & & \\ & D_{n-1} & & \\ & & \ddots & \\ & & & D_1 \end{bmatrix}^{-1} \begin{bmatrix} R_n \\ R_{n-1} \\ \vdots \\ R_1 \end{bmatrix} \\
&= \begin{bmatrix} I_n & D_n^{-1}U_n & \dots & D_n^{-1}U_n \cdot \ldots \cdot D_2^{-1}U_2 \\ & I_{n-1} & \dots & D_{n-1}^{-1}U_{n-1} \cdot \ldots \cdot D_2^{-1}U_2 \\ & & \ddots & \vdots \\ & & & I_1 \end{bmatrix} \begin{bmatrix} D_n^{-1}R_n \\ D_{n-1}^{-1}R_{n-1} \\ \vdots \\ D_1^{-1}R_1 \end{bmatrix}.
\end{aligned}
$$

Thus, the first row of P is the first row of (2.8).