Mathematical Biosciences 235 (2012) 1–7

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Mixing in age-structured population models of infectious diseases ☆

John Glasser [a,*], Zhilan Feng [b], Andrew Moylan [c], Sara Del Valle [d], Carlos Castillo-Chavez [e,1]

[a] Centers for Disease Control and Prevention, Atlanta, GA, USA
[b] Purdue University, West Lafayette, IN, USA
[c] Wolfram Research, Champaign, IL, USA
[d] Los Alamos National Laboratory, Los Alamos, NM, USA
[e] Arizona State University, Tempe, AZ, USA

ARTICLE INFO

Article history:
Received 7 May 2011
Received in revised form 12 September 2011
Accepted 3 October 2011
Available online 20 October 2011

Keywords:
Interpersonal contacts
Preferential mixing
Transmission modeling
Intervention assessment
Indirect effects

ABSTRACT

Infectious diseases are controlled by reducing pathogen replication within or transmission between hosts. Models can reliably evaluate alternative strategies for curtailing transmission, but only if interpersonal mixing is represented realistically. Compartmental modelers commonly use convex combinations of contacts within and among groups of similarly aged individuals, respectively termed preferential and proportionate mixing. Recently published face-to-face conversation and time-use studies suggest that parents and children and co-workers also mix preferentially. As indirect effects arise from the off-diagonal elements of mixing matrices, these observations are exceedingly important. Accordingly, we refined the formula published by Jacquez et al. [19] to account for these newly-observed patterns and estimated age-specific fractions of contacts with each preferred group. As the ages of contemporaries need not be identical nor those of parents and children to differ by exactly the generation time, we also estimated the variances of the Gaussian distributions with which we replaced the Kronecker delta commonly used in theoretical studies. Our formulae reproduce observed patterns and can be used, given contacts, to estimate probabilities of infection on contact, infection rates, and reproduction numbers. As examples, we illustrate these calculations for influenza based on "attack rates" from a prospective household study during the 1957 pandemic and for varicella based on cumulative incidence estimated from a cross-sectional serological survey conducted from 1988–94, together with contact rates from the several face-to-face conversation and time-use studies. Susceptibility to infection on contact generally declines with age, but may be elevated among adolescents and adults with young children.

Published by Elsevier Inc.

1. Introduction

While pathogens spread via interpersonal contacts, transmission may be modeled within and between groups of similar individuals. Appropriate levels of aggregation depend on questions of interest and observations available. Given suitable expressions for heterogeneous mixing, this mean field approach yields dynamic networks whose nodes are ever changing sub-populations defined by age, location, or other strata. Recently, there has been an explosion of models in which network structure defines social contacts among individuals (see, e.g., [24]). Epidemic-control measures have been evaluated using both approaches.

Individual- and population-based models have strengths and weaknesses. Individual-based models capture the chance nature of interpersonal contacts and permit concurrent membership in multiple risk groups (e.g., households and schools or workplaces). Results are presented as frequency distributions from multiple realizations of stochastic processes, allowing policymakers to determine the risk of outcomes more extreme than desired under particular conditions. In contrast, the systems of differential equations comprising population-level models can often be analyzed for general insights. Moreover, their fewer parameters can be more easily estimated from observations. And deficiencies are easier to remedy by comparing predictions to observations and determining the cause of any discrepancies.

While existing formulae represent contacts within subpopulations (e.g., age classes) and between each such group and all others, recently published empirical studies of encounters by which respiratory diseases might be transmitted indicate that parents and children and co-workers also mix preferentially. We generalize the model of Jacquez et al. [19] to include these contacts explicitly, permitting more realistic assessments of the risks

☆ The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or other institutions with which they are affiliated.

* Corresponding author. Address: 1600 Clifton Road, NE, Mail Stop A-34, Atlanta, GA 30333, USA. Tel.: +1 (404) 639 8780; fax: +1 (404) 639 1307.
E-mail address: jglasser@cdc.gov (J. Glasser).
[1] Also Santa Fe Institute, Santa Fe, NM, USA and Cornell University, Ithaca, NY, USA.

0025-5564/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.mbs.2011.10.001

associated with particular outcomes of interest to policymakers. Discrete event/time approximations of systems of differential equations can be simulated (see, e.g., [26]), matching the stochasticity of individual-based models without losing the analytical potential of population-based ones.

Our essay is structured as follows: we begin by deriving extant mixing formulae from first principles, and then describe recently published observations and introduce new formulae capable of reproducing them. Next, we illustrate the utility of such calculations by estimating age-specific probabilities of infection on contact, given "attack rates" or risks of infection. And finally, we estimate the corresponding infection rates, next generation matrix, and reproduction numbers. In our discussion, we highlight the importance of representing inter-group mixing realistically in models designed to evaluate possible mitigation strategies. We believe that these refinements increase the range of applications for which population modeling is appropriate.

2. Methods

2.1. Theoretical studies

Busenberg and Castillo-Chavez [3] define c_{ij} as proportions of contacts that members of group i have with group j, given that i has contacts. Their criteria that mixing models should meet are:

(1) $c_{ij} \geqslant 0$,
(2) $\sum_{j=1}^{k} c_{ij} = 1,\ j = 1,\ldots,k,$ and
(3) $a_i N_i c_{ij} = a_j N_j c_{ji}$,

where the N_i are group sizes and a_i are the average per capita contact rates of groups $i = 1,\ldots,k$, called activities. Common formulae derivable from these conditions follow.

2.2. Proportionate mixing

If we write $c_{ij} = f_i g_j$, where $f_i > 0$ and $g_j > 0$, then

$$1 = \sum_{j=1}^{k} c_{ij} = f_i \sum_{j=1}^{k} g_j, \quad \text{and} \quad f_i = 1 \Big/ \sum_{j=1}^{k} g_j, \quad \forall i = 1,\ldots,k.$$

This implies that f_i is a constant, say L, whereupon $c_{ij} = L g_j = \bar{c}_j$. Substituting, $a_i N_i \bar{c}_j = a_j N_j \bar{c}_i$ and $\sum_{j=1}^{k} a_i N_i \bar{c}_j = \sum_{j=1}^{k} a_j N_j \bar{c}_i$. Rearranging, $a_i N_i \sum_{j=1}^{k} \bar{c}_j = \bar{c}_i \sum_{j=1}^{k} a_j N_j$. As $\sum_{j=1}^{k} \bar{c}_j = 1,$ evidently $\bar{c}_i = \frac{a_i N_i}{\sum_{j=1}^{k} a_j N_j}$.

2.3. Preferential mixing

If a proportion ε_i of i-group contacts is reserved for others in group i, called preferences, and the complement $(1 - \varepsilon_i)$ is distributed among all groups, including i, via the proportionate mixing formula above,

$$c_{ij} = \varepsilon_i \delta_{ij} + (1 - \varepsilon_i) \frac{(1 - \varepsilon_j) a_j N_j}{\sum_k (1 - \varepsilon_k) a_k N_k},$$

where δ_{ij} is the Kronecker delta (i.e., $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$). Because $\frac{a_i N_i}{\sum_j a_j N_j} = \bar{c}_i$, we could write $c_{ij} = \varepsilon_i \delta_{ij} + (1 - \varepsilon_i) \frac{(1-\varepsilon_j)\bar{c}_j}{\sum_k (1-\varepsilon_k)\bar{c}_k}$. Jacquez et al. [19] obtained the first of these preferential mixing expressions by allowing the fraction of within-group contacts, ε, to vary between groups in Nold's [25] preferred mixing model. Hethcote's [16] equation (4.14) is the same as hers with epsilon and its complement reversed. Similar ideas are evident in Barbour's [1] modeling of schistosomiasis or the extensive HIV modeling at the beginning of the pandemic (see, e.g., [6] and references therein). Castillo-Chavez et al. [7], and Blythe and Castillo-Chavez [2] use the log normal distribution and an arbitrary continuous function, respectively, for their main diagonals.

Whatever its merits, the formula of Roberts and Tobias [27], $c_{ij} = \delta_{ij} + \frac{\varepsilon}{a_i}\sqrt{a_i a_j}$ with $\varepsilon < 1$ in our notation, does not satisfy Busenberg's and Castillo-Chavez' [3] second condition. Consider their four-group example (but omitting the reference for generality) when $i = 2$,

$$\sum_j c_{2j} = \frac{1}{a_2}\left[\varepsilon\sqrt{a_2 a_1} + a_2 + \varepsilon\sqrt{a_2 a_3} + \varepsilon\sqrt{a_2 a_4}\right] = \varepsilon\frac{\sqrt{a_1}}{\sqrt{a_2}} + 1 + \varepsilon\frac{\sqrt{a_3}}{\sqrt{a_2}} + \varepsilon\frac{\sqrt{a_4}}{\sqrt{a_2}},$$

which clearly exceeds 1 unless $\varepsilon = 0$, whereupon there would be no off-diagonal contacts.

2.4. Empirical studies

While the formula of Jacquez et al. [19] captures the most striking feature of age-specific mixing, the activity and preference for contemporaries of older children and adolescents relative to others (results not shown), it does not capture two patterns apparent in recent studies of face-to-face conversations [28,22] and periods in proximity with others [9,31]: (1) sub- and super-diagonals representing contacts between parents and children and vice versa, and (2) a rectangular area bounded by the working ages within which contacts are independent of age.

2.5. Refined formulae

Accordingly, we extend Jacquez et al.'s formula to include contacts between parents and children (the sub- and super-diagonals in Fig. 1a) and among co-workers (the square in Fig. 1b) as well as contemporaries (the main diagonal in Fig. 1a):

$$c_{ij} = \phi_{ij} + \left(1 - \sum_{l=1}^{4} \varepsilon_{li}\right) f_j, \quad f_j = \frac{\left(1 - \sum_{l=1}^{4} \varepsilon_{lj}\right) a_j N_j}{\sum_{k=1}^{n} \left(1 - \sum_{l=1}^{4} \varepsilon_{lk}\right) a_k N_k}.$$

When $\vec{\varepsilon}_2 = \vec{\varepsilon}_3 = \vec{\varepsilon}_4 = 0$ (where $\vec{\varepsilon}_l$ represents the vector whose components are ε_{li}), this expression reduces to the formula of Jacquez et al. [19]. Fig. 1a illustrates contacts between parents and children as well as among contemporaries (i.e., $\vec{\varepsilon}_4 = 0$). Fig. 1b illustrates only contacts among co-workers (i.e., $\vec{\varepsilon}_1 = \vec{\varepsilon}_2 = \vec{\varepsilon}_3 = 0$).

Because the sub- and super-diagonals extend over ages $i \geqslant G$ and $i \leqslant L–G$, respectively, where G is the generation time (i.e., average age at which women bear daughters), L is longevity (i.e., average expectation of life at birth), and $L > G$, we define ϕ_{ij} as:

$$\phi_{ij} = \begin{cases} \delta_{ij}\varepsilon_{1i} + \delta_{i(j+G)}\varepsilon_{2i} + I_W(i,j)\frac{\varepsilon_{4i}}{W_{\max} - W_{min}}, & i \geqslant G, \\ \delta_{ij}\varepsilon_{1i} + \delta_{i(j-G)}\varepsilon_{3i} + I_W(i,j)\frac{\varepsilon_{4i}}{W_{\max} - W_{min}}, & i \leqslant L–G. \end{cases}$$

Only people whose ages equal or exceed G can have children, and only those whose ages equal or are less than $L–G$ can have parents, but people aged at least G but not more than $L–G$ can have both children and parents. In these inequalities, we mix indices and real numbers, but if age classes are 0–4, 5–9, … and $G = 25$ years, for example, by $i > G$ we mean $i >$ class 5. Other new variables are W_{min} and W_{max}, average ages at entry to and exit from the workforce (Fig. 1b), ε_{1i}–ε_{4i}, fractions of contacts reserved for contemporaries, children $(j - G)$, parents $(j + G)$, and co-workers (if $W_{min} \geqslant i$, $j \leqslant W_{max}$), respectively, and the corresponding delta or indicator function. That is,

$$\delta_{i(j\pm G)} = \begin{cases} 1 & \text{if } i = j \pm G \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad I_W(i,j) = \begin{cases} 1 & \text{if } i,j \in W \\ 0 & \text{otherwise,} \end{cases}$$

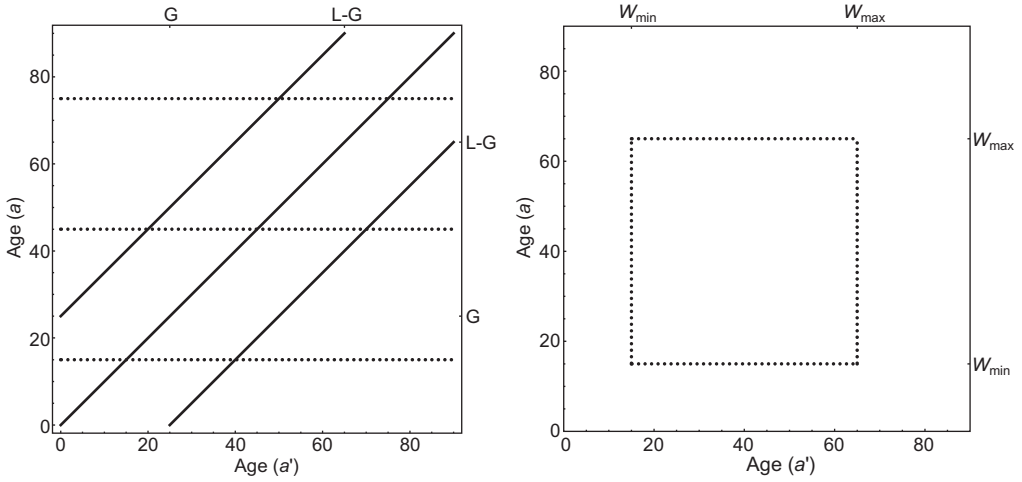

Fig. 1. Schematic contact matrices illustrating, on the left, the main and off-diagonals representing contacts among contemporaries, between children and parents, and vice versa, and on the right, the age range within which age-independent contacts among co-workers occur. In this figure and the next, $G = 25$ and $L = 90$ years.

where $W = [W_{min}, W_{max}]$. Notice that the non-zero elements of $\vec{\varepsilon}_2$ and $\vec{\varepsilon}_3$ are related. If G = 25 years, for example, then $a_i \times N_i \times \varepsilon_{2i} = a_j \times N_j \times \varepsilon_{3j}$, for $i = 6, 7, \ldots$, $j = i - 5$. Accordingly, we estimate ε_{3i} by assuming that $\varepsilon_{2i} = \frac{a_j \times N_j \times \varepsilon_{3j}}{a_i \times N_i}$. Notice also that $0 \leqslant \sum_{l=1}^{4} \varepsilon_{li} < 1$ and that mixing among co-workers does not depend on age provided that $i \geqslant W_{min}$ and $j \leqslant W_{max}$.

While delta formulations are undeniably heuristic, contemporaries need not be exactly the same age [16], nor need the ages of parents and children to differ by exactly the generation time. Accordingly, we reformulate ϕ_{ij} to incorporate this more realistic feature. Let α and α' denote the ages of susceptible and infected individuals, respectively. Further, let $a(\alpha)$ denote the average number of contacts per person aged α per unit of time and $N(\alpha)$ denote the number of people aged α. The continuous analogue of c_{ij} can be formulated as:

$$c(\alpha, \alpha') = \phi(\alpha, \alpha') + \left[1 - \sum_{l=1}^{4} \varepsilon_l(\alpha)\right] f(\alpha'),$$

$$f(\alpha') = \frac{\left[1 - \sum_{l=1}^{4} \varepsilon_l(\alpha')\right] a(\alpha') N(\alpha')}{\int_0^\infty \left[1 - \sum_{l=1}^{4} \varepsilon_l(u)\right] a(u) N(u) du},$$

where

$$\phi(\alpha, \alpha') = \begin{cases} g_1(\alpha, \alpha')\varepsilon_1(\alpha) + g_2(\alpha, \alpha')\varepsilon_2(\alpha) + I_W(\alpha, \alpha')\frac{\varepsilon_4(\alpha)}{W_{\max} - W_{min}}, & \alpha \geqslant G, \\ g_1(\alpha, \alpha')\varepsilon_1(\alpha) + g_3(\alpha, \alpha')\varepsilon_3(\alpha) + I_W(\alpha, \alpha')\frac{\varepsilon_4(\alpha)}{W_{\max} - W_{min}}, & \alpha \leqslant L - G, \end{cases}$$

with

$$g_1(\alpha, \alpha') = \frac{1}{\sqrt{2\pi}\sigma_1(\alpha)} e^{-\frac{(\alpha' - \alpha)^2}{2[\sigma_1(\alpha)]^2}}, \quad g_2(\alpha, \alpha') = \frac{1}{\sqrt{2\pi}\sigma_2(\alpha)} e^{-\frac{[\alpha' - (\alpha - G)]^2}{2[\sigma_2(\alpha)]^2}},$$

and $g_3(\alpha, \alpha') = \frac{1}{\sqrt{2\pi}\sigma_3(\alpha)} e^{-\frac{[\alpha' - (\alpha + G)]^2}{2[\sigma_3(\alpha)]^2}}$,

where the $g_k(\alpha, \alpha')$ (k = 1, 2, 3) are Gaussian kernels with standard deviations $\sigma_k(\alpha)$ and

$$I_W(\alpha, \alpha') = \begin{cases} 1 & \text{if } \alpha, \alpha' \in W \\ 0 & \text{otherwise,} \end{cases} \text{ where } W = [W_{\min}, W_{\max}].$$

Besides the above-mentioned relationship between $\varepsilon_2(\alpha)$ and $\varepsilon_3(\alpha)$, for each α, $0 \leqslant \sum_{l=1}^{4} \varepsilon_l(\alpha) < 1$. Fig. 2 corresponds to the horizontal lines on Fig. 1a.

We fit a hybrid of these discrete and continuous formulations of our model (i.e., our discrete formulation with Gaussian kernels instead of deltas) to observations from the four above-mentioned empirical studies, all of which are discrete, using the FindMinimum function in Mathematica™. This amounts to choosing ε_{li} and σ_{ki}, as well as G, L and the Ws, that minimize an objective function, here the mean squared error. With one starting value for each variable, FindMinimum uses BFGS quasi-Newton methods. When there are constraints, FindMinimum uses interior point methods. We found it necessary to constrain $\vec{\varepsilon}_1$ and $\vec{\sigma}_k$ lest the main diagonal dominate, and fixed G, L, and the Ws after convergence. We ensured that solutions were robust by using different initial conditions for simple models and solutions of simpler models (e.g., with off-diagonals identical, without contacts among co-workers, ...) for more complex ones.

2.6. Applications

Here we use "attack rates" and risks of infection, information typically available for transmission modeling, to derive probabilities of infection on contact, infection rates and reproduction numbers for two respiratory diseases. For purposes of illustration, we use the above-mentioned empirical contact matrices. Should their age classes not correspond to or include those in one's transmission model, our mixing model could be used instead.

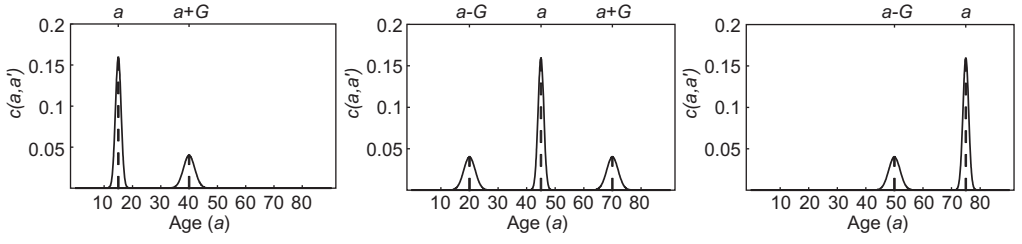

Fig. 2. Left to right, the panels correspond to horizontal lines on Fig. 1a. People of all ages contact contemporaries preferentially, but children aged <G also contact their parents, adults aged >L–G their children, and those in between both. The heights of the Gaussian curves are determined by $\varepsilon(\alpha)$ and their widths by $\sigma(\alpha)$. A similar line through Fig. 1b anywhere between $y = W_{min}$ and $y = W_{max}$ would equal ε_4 between $x = W_{min}$ and $x = W_{max}$, but be 0 otherwise.

Influenza pandemics occur when recombination results in novel strains against which there is little or no population immunity [12], at least among children, so we can calculate the risks of infection $\lambda_i = -\ln(1-y_i)$, where y_i are commonly called "attack rates." We use those from a prospective household transmission study during the 1957 pandemic [5]. The second equation in the classic "Susceptible, Infected, and Removed" model, for example, is $I_i' = \lambda_i S_i - \gamma I_i$, where $\lambda_i = a_i \beta_i \sum_{j=1}^{n} c_{ij} y_j$, and γ is the recovery rate, S_i, I_i, and R_i are the numbers susceptible, infected, and removed, respectively, and $N_i = S_i + I_i + R_i$. We solve for the endemic equilibrium, rearrange and divide both sides by N_i, obtaining $\frac{I_i}{N_i} = \frac{\lambda_i}{\gamma} \times \frac{S_i}{N_i}$. As $S_i/N_i \approx 1$ during pandemics, whereupon $y_i = \lambda_i/\gamma$, we calculate the β_i given the a_i and c_{ij} from the above-mentioned empirical studies (i.e., $a_i = \sum_j C_{ij}$, $c_{ij} = C_{ij}/a_i$).

Varicella-zoster virus is relatively stable, so increasing proportions by age are immune. We fit Farrington's [13] model,

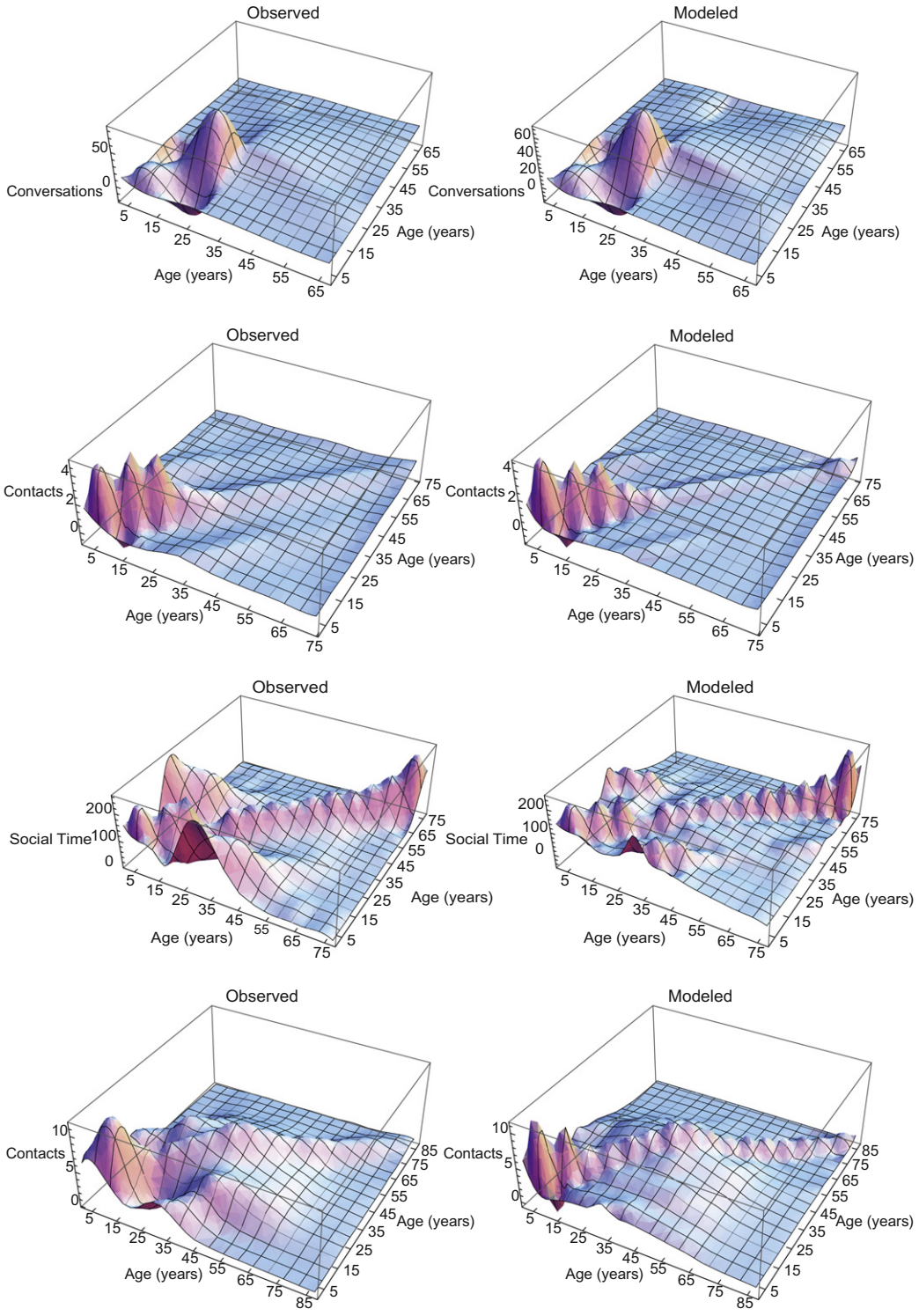

Fig. 3. Interpolating functions fitted to geometric means of corresponding row- and column-elements from the four empirical studies (left) and fits of our model (right). On the left are, top to bottom, the observations of Wallinga et al. [28], Mossong et al. [22], Zagheni et al. [31], and Del Valle et al. [9], respectively.

$F(\alpha) = 1 - e^{-\int_0^\alpha \lambda(u)du}$, where $\lambda(\alpha) = (a\alpha - c)e^{-b\alpha} + d$, to proportions of sera with protective antibody titers (see [23]) collected during the third National Health and Nutrition Survey (http://www.cdc.gov/nchs/nhanes/nh3data.htm), 1988–94, using the FindFit function in Mathematica™. FindFit uses singular value decomposition and the Levenberg–Marquardt method for linear and nonlinear least-squares, respectively, and the FindMinimum methods described above otherwise. As the probability of remaining susceptible, $P_S(\alpha) = e^{-\int_0^\alpha \lambda(u)du}$, its negative derivative, $-P'_S(\alpha) = \lambda(\alpha)e^{-\int_0^\alpha \lambda(u)du}$ is the probability density function of first infection or "attack rate." In this case, we calculate the β_i via the definition of λ_i.

Given these β_i, we can write $a_i\beta_i c_{ij} = \beta_{ij}$, the rates of effective contact between members of groups i and j (alternatively, infection of susceptible members of group i by infectious members of group j) that are required for transmission modeling. Noting that the β_{ij}/γ are elements of the next-generation matrix, whose largest eigenvalue is the basic or intrinsic reproduction number [10], we can also calculate $\mathfrak{R}_0$.

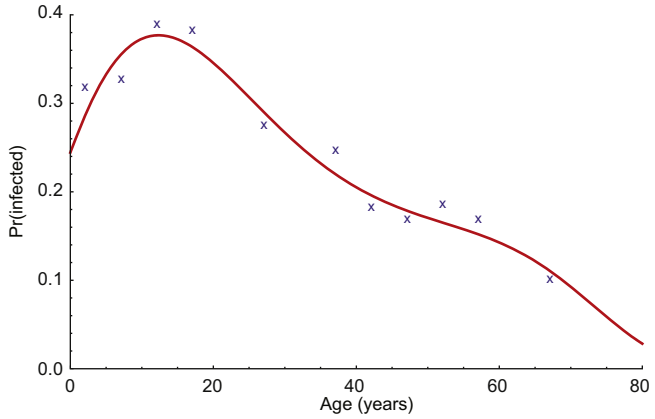

Fig. 4. "Attack rates" from a prospective household study during the 1957 pandemic [5] and the fitted regression equation, $\Pr(\alpha) = \exp[-1.13 + 0.12\alpha - 0.007\alpha^2 + 0.0001\alpha^3 - 8 \times 10^{-07}\alpha^4]/(1 + \exp[-1.13 + 0.12\alpha - 0.007\alpha^2 + 0.0001\alpha^3 - 8 \times 10^{-07}\alpha^4])$, where α = age in years.

3. Results

The observed contact matrices are asymmetric, presumably because persons contacted need not also have participated in these studies (i.e., study populations were not closed). As this violates the third condition of Busenberg and Castillo-Chavez [3], we replaced reported elements by the geometric means of those from the corresponding rows and columns before estimating the ε's and σ's. Hsu Schmitz and Castillo-Chavez [18] describe circumstances under which the contributions of people not surveyed may be estimated.

Fig. 3a–d illustrate observations thus adjusted from these recent empirical studies (on the left) and our fitted models (on the right). The observed $C_{ij} = a_i c_{ij}$, together with influenza "attack rates" (Fig. 4), yield probabilities of infection on contact β_i (Fig. 5a–d). Using proportions protected and risks of infection derived from the cross-sectional survey of antibodies to varicella instead, the calculated β_i are similar.

Together with $1/\gamma$ of 3.8 days [8], these yield $\mathfrak{R}_0 \approx 3.9$ for influenza. We assume a 1-day latent period, but Carrat et al. [4] report viral shedding from healthy volunteers beginning the first day post-challenge and continuing for 4.8 days. People with varicella may be infectious from before lesions appear until scabs form, typically 6–7 days [17], for which $\mathfrak{R}_0 \approx 12.6$.

4. Discussion

Patterns apparent in recently published studies of face-to-face conversations and periods in proximity with others motivated us to elaborate the preferential mixing model of Jacquez et al. [19] to include contacts between parents and children and among co-workers as well as contemporaries. Unlike mixing among contemporaries, that between parents and children and among co-workers

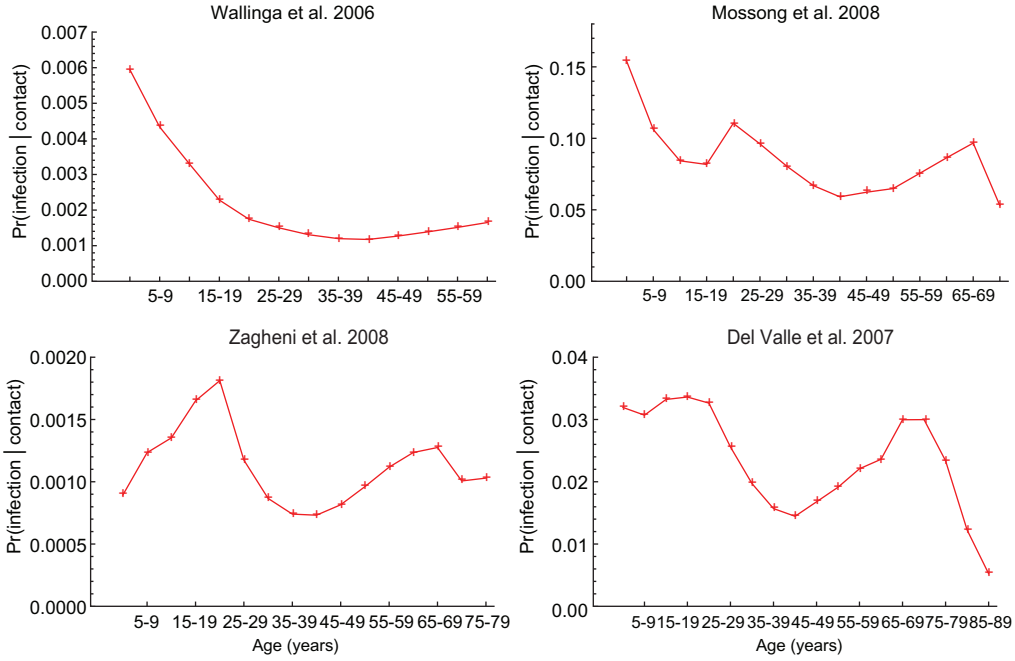

Fig. 5. Probabilities of infection with influenza per conversation, period together, or effective contact, ordered left to right and top to bottom as in Fig. 3. Patterns are similar despite quantitative differences reflecting average weekly [28] or daily intimate and casual conversations [22], daily periods (minutes) with others [31], and numbers of contacts × probabilities of transmission per contact derived from mean contact durations [9].

involves off-diagonal matrix elements. In cross-classified population models, the main diagonal is responsible for direct effects; other matrix elements are responsible for indirect ones. Thus, modelers using extant formulae, in which mixing between groups is proportional to their respective contacts, may have underestimated indirect effects.

Suppose that one wished to assess the impact of vaccinating young parents on pertussis among infants. Immunity-modified disease can be quite mild, but that among infants lacking maternal antibodies typically is severe and occasionally even fatal. Mothers with prolonged cough illnesses may not be very infectious, but mother–infant contacts are particularly intimate. Which mixing formula should one use? What about the impact of vaccinating older children against influenza on mortality among elderly adults, who occasionally die of pneumonia, which may complicate influenza? These groups are no more connected than any others in the existing formulation, but infants and young adults are directly connected in ours. And children and elderly adults are connected via their parents and children, respectively.

Delta formulations are convenient mathematically, but do not allow the age range of one's contemporaries to vary as one ages (e.g., the range narrows perceptively among adolescents), much less differences between the age ranges of one's contemporaries and one's parents or children. By virtue of secular patterns in childbearing, moreover, the age ranges of parents and children may change with age. But the Gaussian formulation allows such variation, reproducing the essential features of these observations (cf. Fig. 3a–d), and can inform – given constant or age-specific susceptibilities to infection on contact – infection rates for transmission modeling. Truncating the Gaussian, or using the lognormal or gamma distributions would ensure that ages were positive, but not that $a \leqslant L$. Infants' contemporaries have such a narrow age range and persons aged $a > L$ are so few that such complications would be an unwarranted distraction. Applications of continuous distributions to biological phenomena require common sense.

The observed mixing matrices are qualitatively similar (Fig. 3a–d left), but differ quantitatively. In particular, the sub- and super-diagonals from both time-use studies exceed those from both studies of face-to-face conversations. While people certainly can share spaces without conversing, careful analysis of similarities and differences between these and other encounters by which respiratory diseases might be transmitted would be illuminating. And, if protocols permitted comparison of contact patterns, as in the European countries studied by Mossong et al. [22], differences among societies could increase our understanding of the manner in which social structure mediates infectious disease transmission.

By using published observations in this manner, we have sidestepped the exceedingly important issue of what exactly constitutes a contact [11]. In our defense, the authors of these empirical studies were thoughtful people who aimed to illuminate a general phenomenon whose details may nonetheless differ from one respiratory disease to another. Other infectious diseases are transmitted via fomites or sexual relations.

Authors of two of these studies "weighted" contacts by duration or intimacy. The observations of Del Valle et al. [9] are $P_{ij} = 1 - \exp(-\sigma T_{ij})$, where σ is the mean number of transmissions per hour of contact between fully infectious and susceptible people (the authors assumed 0.2) and T_{ij} are mean contact durations (in hours) per day (Fig. 3d). Similarly, Mossong et al. [22] reported all daily conversations and ones involving physical contact (Fig. 3b), so casual conversations can be obtained by difference. While contacts among persons of working age are more common among casual than intimate conversations (this is more apparent in contour plots), contacts among contemporaries predominate in both.

Observations from the earlier Dutch study [28] are average weekly conversations and from the other time-use study, they are average daily periods (in minutes) sharing spaces [31]. Given proportions with protective antibodies, probabilities of infection are qualitatively similar. Except for the study by Zagheni et al. [31], susceptibility to influenza generally is highest among children and declines with age. In several studies, susceptibility increases transiently when adults could be involved in childcare (i.e., young adults and their parents, the young children's grandparents). Susceptibility to varicella is highest among children too, but increases again among older adults in the time-use studies.

We have also estimated β_j alone and both β_i and β_j simultaneously for several respiratory diseases, where members of group i are susceptible and those of group j infectious. These alternative formulations are $\lambda_i = a_i \sum_j c_{ij} \beta_j y_j$ and $\lambda_i = a_i \beta_i \sum_j c_{ij} \beta_j y_j$, where $y_j = I_j/N_j$. Somewhat surprisingly, our estimates of β_i are more coherent or intelligible than those of β_j. Using these varicella observations, for example, β_j increases irregularly with age (results not shown). That is, susceptibility seems to vary more with age than infectiousness does. Even if other respiratory diseases proved similar in this respect, the corresponding probabilities for infectious diseases transmitted via fomites or sexual contact need not follow suit.

Differences between influenza and varicella presumably reflect population immunity. We assumed that everyone was susceptible during the 1957 influenza pandemic, but insofar as new influenza strains evolve from earlier ones, this conventional wisdom cannot be correct. During the 2009 pandemic, for example, older adults had residual immunity for reasons molecular biologists have recently elucidated [30]. Varicella afflicts children primarily, but adults who escaped infection during childhood remain susceptible, and their risk of exposure may increase during their reproductive years. In common with many infectious diseases, varicella is more serious among older than younger people. Our mixing formulae will enable population modelers to faithfully reproduce such patterns.

Our reproduction number for varicella resembles others' (e.g., [29] obtain 3–12 depending on mixing), but our estimate for influenza is high, presumably because the risk of infection in households exceeds that in communities [20]. We did not consider asymptomatic infections or residual or cross-immunity. Immunity affects estimates directly insofar as populations at risk actually are smaller than believed [21]. This follows from the relationship, $\frac{I_i}{N_i} = \frac{\lambda_i}{\gamma} \times \frac{S_i}{N_i}$. Given observed "attack rates," exaggeration of S_i/N_i diminishes λ_i/γ, implying a smaller β_i and thus $\Re_0$. The second effect depends on the extent to which asymptomatic people are infectious [14].

We have performed similar calculations for pertussis, based on age-specific proportions with antibodies passively-acquired from mothers or indicative of recent infection that decay and wane, respectively, albeit at different rates [15].

5. Summary

Motivated by recent empirical observations of proxies for contacts by which respiratory diseases might be transmitted, we augmented the mixing model of Jacquez et al. [19] to include preferential contacts between parents and children and among co-workers as well as contemporaries. Hethcote [16] suggested replacing the deltas commonly used in theoretical studies with the Gaussian or another distribution because contemporaries need not be exactly the same age. As similar reasoning also applies to parents and children, we reformulated our model to include several Gaussian kernels. We first demonstrate that our mixing model fits the observed contact matrices, whose elements are C_{ij}, numbers of conversations (or periods) that members of group i have

(spend) with those of group j. Off-diagonal matrix elements are solely responsible for indirect effects. Having thus failed to disprove our mixing model, a hypothesis about phenomena structuring these extraordinarily important contacts, we turn to a practical objective. We use published or otherwise available attack rates and forces of infection to calculate the probabilities of infection on contact, the infection rates required for transmission modeling, and the reproduction numbers for two vaccine-preventable diseases. In these calculations, we use the empirical observations. Should they not suffice for transmission modeling (i.e., the observed age classes not match or include those modeled), our mixing model could be used instead.

Acknowledgments

We are grateful to Roger Germundsson for guiding J.G. during the 2009 Advanced Mathematica™ Summer School, Karl Hadeler for helpful discussions, two anonymous reviewers for constructive suggestions, and Nathaniel Hupert and Michael Washington for support. Recognizing the importance of mixing in population modeling, John Edmunds pioneered empirical studies of inter-personal contacts, several of which permitted this theoretical study and applications. Z.F.'s research is partially supported by IPA908630 between the CDC and Purdue University and NSF grant DMS-1022758.

References

[1] A.D. Barbour, Macdonald's model and the transmission of bilharzia, Trans. R. Soc. Trop. Med. Hyg. 72 (1978) 6.
[2] S.P. Blythe, C. Castillo-Chavez, Like-with-like preference and sexual mixing models, Math. Biosci. 96 (1989) 221.
[3] S. Busenberg, C. Castillo-Chavez, A general solution of the problem of mixing of sub-populations and its application to risk- and age-structured epidemic models for the spread of AIDS, IMA J. Math. Appl. Med. Biol. 8 (1991) 1.
[4] F. Carrat, E. Vergu, N.M. Ferguson, et al., Time lines of infection and disease in human influenza: a review of human volunteer studies, Am. J. Epidemiol. 167 (2008) 775.
[5] T.D.Y. Chin, J.F. Foley, I.L. Doto, C.R. Gravelle, J. Weston, Morbidity and mortality characteristics of Asian strain influenza, Public Health Rep. 75 (1960) 149.
[6] C. Castillo-Chavez (Ed.), 1989. Mathematical and Statistical Approaches to AIDS Epidemiology. Lecture Notes in Biomathematics, vol. 83, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, 405p.
[7] C. Castillo-Chavez, S. Busenberg, K. Gerow, Pair formation in structured populations, in: J. Goldstein, F. Kappel, W. Schappacher (Eds.), Differential Equations with Applications in Biology, Physics and Engineering, Marcel Dekker, NY, 1990, pp. 47–65.
[8] S. Cauchemez, F. Carrat, C. Viboud, A.J. Valleron, P.Y. Boëlle, A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data, Stat. Med. 23 (2004) 3469.
[9] S.Y. Del Valle, J.M. Hyman, H.W. Hethcote, S.G. Eubank, Mixing patterns between age groups in social networks, Soc. Networks 29 (2007) 539.
[10] O. Diekmann, J.A. Heesterbeek, J.A. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (1990) 365.
[11] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation, John Wiley and Son, Ltd., Chichester, UK, 2000. 303p.
[12] D.J.D. Earn, J. Dushoff, S.A. Levin, Ecology and evolution of the flu, Trends Ecol. Evol. 17 (2002) 334.
[13] C.P. Farrington, Modeling risks of infection for measles, mumps and rubella, Stat. Med. 9 (1990) 953.
[14] Z. Feng, in preparation. On the estimation of $\Re_0$ when some infectious people are asymptomatic.
[15] Z. Feng, J. Glasser, M. Andersson, H. Hallander, P. Tüll, P. Olin, unpublished manuscript. Modeling risks of infection when immunity wanes: application to Bordetella pertussis in Sweden.
[16] H.W. Hethcote, Modeling heterogeneous mixing in infectious disease dynamics, in: V. Isham, G. Medley (Eds.), Models for Infectious Human Diseases: Their Structure and Relation to Data, Cambridge University Press, Cambridge, UK, 1996, pp. 215–238.
[17] D.L. Heymann, Control of Communicable Diseases Manual, 18th ed., American Public Health Association, 2004. 699p.
[18] S.-F. Hsu Schmitz, C. Castillo-Chavez, Parameter estimation in non-closed social networks related to the dynamics of sexually-transmitted diseases, in: E.H. Kaplan, M.L. Brandeau (Eds.), Modeling the AIDS Epidemic: Planning, Policy, and Prediction, Raven Press, NY, 1994, pp. 533–560.
[19] J.A. Jacquez, C.P. Simon, J. Koopman, L. Sattenspiel, T. Perry, Modeling and analyzing HIV transmission: the effect of contact patterns, Math. Biosci. 92 (1988) 119.
[20] I.M. Longini Jr, J.S. Koopman, A.S. Monto, J.P. Fox, Estimating household and community transmission parameters for influenza, Am. J. Epidemiol. 115 (1982) 736.
[21] J.D. Matthew, C.T. McCaw, J. McVernon, E.S. McBryde, J.M. McCaw, A biological model for influenza transmission: pandemic planning implications of asymptomatic infection and immunity, PLoS One 2 (11) (2007) e1220, doi:10.1371/journal.pone.0001220.
[22] J. Mossong, N. Hens, M. Jit, et al., Social contacts and mixing patterns relevant to the spread of infectious diseases, PLoS Med. 5 (2008) 381.
[23] US Department of Health and Human Services (DHHS), 1996. National Center for Health Statistics. Third National Health and Nutrition Examination Survey, 1988–1994, NHANES III Laboratory Data File (CD-ROM). Public Use Data File Documentation Number 76200. Centers for Disease Control and Prevention, Hyattsville, MD.
[24] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167.
[25] A. Nold, Heterogeneity in disease transmission modeling, Math. Biosci. 124 (1980) 59.
[26] E. Renshaw, Modeling Biological Populations in Space and Time, Cambridge University Press, Cambridge, UK, 1991. 403p.
[27] M.G. Roberts, M.I. Tobias, Predicting and preventing measles epidemics in New Zealand: application of a mathematical model, Epidemiol. Infect. 124 (2000) 279.
[28] J. Wallinga, P. Teunis, M. Kretzschmar, Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents, Am. J. Epidemiol. 164 (2006) 936.
[29] H.J. Whitaker, C.P. Farrington, Infections with varying contact rates: application to varicella, Biometrics 60 (2004) 615.
[30] R. Xu, D.C. Ekiert, J.C. Krause, R. Hai, J.E. Crowe, I.A. Wilson, Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. Published online March 25 (2010). 10.1126/science.1186430.
[31] E. Zagheni, F.C. Billari, P. Manfredi, A. Melegaro, J. Mossong, W.J. Edmunds, Using time-use data to parameterize models for the spread of close-contact infectious diseases, Am. J. Epidemiol. 168 (2008) 1082.