The Relative Importance of Frequency of Contacts and Duration of Exposure for the Spread of Directly Transmitted Infections

Elisabetta De Cao Emilio Zagheni Piero Manfredi
Alessia Melegaro

Abstract

The recent availability of survey data on social contacts (e.g., Polymod) has strongly improved our understanding of the social determinants of the spread of close-contact infections. However, little is known about the relationship between two critical factors that explain the transmission of infections: frequency of contacts and duration of exposure. Using classic results from probability theory, this paper combines these two factors by obtaining a new relevant epidemiological quantity: the number of "suitable" contacts (i.e., contacts that involve a sufficiently long time of exposure to allow transmission). Model parameters, estimated against serological data, regulate the length of the minimal duration of exposure for a suitable contact. A wide range of age-specific matrices of suitable contacts can be derived to infer the level of transmissibility for different infections. The model has been tested using data on time use, number of contacts, and seroprevalence for varicella in Italy. The results show that the minimal duration of exposure for transmission of varicella is very small, confirming that varicella is highly transmissible. The proposed approach shows the relative importance of number of contacts versus time of exposure. This is relevant to design public health interventions, whose outcome critically depends on social contact patterns.

Keywords: Time use data; Contact data; Varicella Zoster Virus; Bayesian Melding.

Abbreviations: TUD, Time Use Data; TU, Time Use; VZV, Varicella Zoster Virus; NGM, Next Generation Matrix; MSD, Minimal Suitable Duration; SIR, Susceptible Infective Removed.

Social contact patterns are the key factors underlying the spread of close-contact infectious diseases, like measles or varicella [1, 2]. Early approaches to measure social contacts were indirect [3]. Matrices of transmission rates (e.g., the Who-Acquires-Infection-From-Whom matrices [1] and the proportionate-preferred matrices [4]) were based on theoretical hypotheses about social mixing patterns. Even though the indirect approach has been very influential [3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14], it has an important limitation: model results are highly sensitive to the assumed mixing structures.

Recently, three main approaches to measure contacts directly from social data have been proposed [15, 16, 17]. These approaches produce estimated matrices whose entries represent the average number of contacts or time of exposure that individuals in age group i have with individuals in age group j, per unit of time. A first approach relies on contact surveys in which the respondent self reports the number of contacts she/he had during a randomly sampled day (and additional information: age of contacted persons, type of contact, etc.) [15, 18, 19, 20, 21]. A second approach relies on Time Use Data (TUD): time of exposure matrices are estimated from time use (TU) diaries, assuming that simple mixing rules (e.g., proportionate mixing) hold at the level of single locations and for short time slots [16]. In a third approach, contact matrices are estimated from the simulation outputs of individual-based models, appropriately calibrated to socio-demographic and TUD [17, 22], to generate the underlying contact network structure of the population.

TUD and contact surveys have been used independently to model two different dimensions of the transmission process. How can we integrate the two sources of information into a unified model? What is the role of number of contacts versus

duration of exposure? Is it specific to infections, or groups of infections? This article addresses these and related questions. We propose a novel methodology, based on the so-called "occupancy problem", to combine the two data sources. In our approach, contacts "suitable for transmission" are the ones which are expected to last long enough to include an infectious event. Our method is very powerful and flexible since it can model a large family of contact matrices, possibly reflecting different degrees of infection transmissibility. We show an application to an epidemiological model for varicella.

MATERIALS AND METHODS

Data

As part of the European Sero-Epidemiology Network (ESEN2) [23] and Polymod project [21], Italian serological data were collected and tested for antibodies to varicella zoster virus (VZV) infection (sample size: 2,517; age-range: 0-79 years; date 1997-2003). Given that no mass vaccination program for VZV was in place in Italy at the time, the serological data describe the natural history of the infection.

TUD were collected by the Italian National Statistical Agency (ISTAT) in 2002-2003 on a sample of about 24 thousand households. Each respondent reported in a diary all the activities that she/he did during a randomly assigned day, as well as the locations where the activities took place.

Contact data for Italy were collected in 2006 as part of the Polymod project [21]. Respondents (n=849) were asked to report the number of different people they had contact with during a randomly assigned day, demographic information,

characteristics of the contacts (i.e., age, gender, location, frequency) and whether the contact was physical or non-physical.

We used data for Italy, the only country for which we have a full combination of the three relevant data sources: a Time Use study covering young children ($\geq$ 3 years), a social contacts survey, and serological data for close-contact infections.

Suitable contact matrices

Contact matrices, C, have elements c_{ij} representing the average number of contacts (per unit of time, e.g. per day) that individuals in age group i have with individuals in age group j [15, 17, 21, 22]. Duration of exposure matrices, E, have elements e_{ij} which represent the average time (e.g., in minutes) that individuals in age group i are 'exposed' to individuals in age group j [16]. We propose a novel measure of contacts that we summarize in matrices of "suitable contacts". In our approach, a contact is "suitable" when, under the assumption that transmissibility cumulates over the duration of a contact, the underlying duration of exposure is sufficiently long to allow for transmission of the infection. We evaluate suitable contacts from TUD and contact surveys using a probabilistic result from the "occupancy problem", that is the problem of computing the probability that a number of given boxes receive at least one ball, when a fixed number of balls are randomly assigned to these boxes (see the Appendix for details). We assume that duration of exposure between age groups is randomly allocated to the respective number of contacts in "discrete packages" (analogously to balls and boxes), and that each infection is characterized by a minimal duration "package" that is necessary for transmission: a "minimal suitable duration" (MSD). Essentially, given indepen-

dent information on average number of contacts and average duration of contacts, we want to compute the expected value of the number of contacts that last longer than a minimal threshold for transmission. For example, consider an infection for which one minute is the MSD, and assume that age groups i and j are exposed to each other, e_{ij}, for 20 minutes/day, and have an average number of contacts c_{ij} of 10/day. The average duration of contacts, e_{ij}/c_{ij}, is two minutes. However, if the 20 minutes are randomly allocated to contacts, there is also a quite large (binomial) probability (more than 12%) that a contact lasts less than a minute, therefore being "not suitable" for infection transmission.

Now consider contacts and exposure over longer time periods, under the assumption that the average duration per contact e_{ij}/c_{ij} remains constant over time. Let u_{ij} be a random variable representing the number of suitable contacts between age groups i and j. The expected number of suitable contacts $\bar{u}_{ij} = E(u_{ij})$ is given by the product of the average number of contacts c_{ij} and the proportion of these contacts that are suitable for transmission $(1 - \exp(e_{ij}/c_{ij}))$. The larger the average duration of contacts e_{ij}/c_{ij}, the larger the proportion of contacts suitable for transmission. In our previous example, the proportion of suitable contacts would be $(1 - \exp(-2))$, i.e. about 86%. The average number of suitable contacts $\bar{u}_{ij}$ is smaller than the average number of contacts c_{ij} as some contacts may not last long enough and therefore are "not suitable for transmission".

In principle, if the MSD were known for different infections, we could compute suitable contact matrices U for any close-contact infectious disease. However, this information is generally not available and needs to be estimated or assumed. Our approach shows how the shape of a suitable contact matrix, or assortativeness in contacts, varies as the minimal (infection-specific) duration varies. Consider, for

example, a disease with a lower level of transmissibility than before, say 2 minutes instead of 1 as MSD. Then we would have a smaller proportion of suitable contacts, given by $(1 - \exp(-10/10))$, i.e. about 63%. The impact of changes in MSD on the overall shapes of the suitable contact matrix depends on the overall shape of both E and C matrices. However, it can be shown that, for standard situations, less transmissible infections have a less assortative suitable matrix U, than their original contact matrix C.

In our empirical analysis we estimate the fraction q_2 of total exposure time between age groups which is suitable for transmission, by expressing the expected number of suitable contacts as $\bar{u}_{ij} = c_{ij}(1 - \exp(-q_2 e_{ij}/c_{ij}))$. The resulting suitable matrix is validated against observed seroprevalence data by computing the respective predicted values using an age-structured SIR (Susceptible-Infective-Removed) model at endemic equilibrium and under the social contact hypothesis [13, 15, 17, 24]. The latter assumes that age-specific transmission is proportional to the number of suitable contacts through a single disease-specific (here assumed age-independent) parameter $q_1 : k_{ij} = q_1 \times \bar{u}_{ij}$.

The elements of the next generation matrix (NGM) [15, 25] are obtained as the product of k_{ij} and duration of infectiousness d (which is about 7 days for varicella): $NGM_{ij} = q_1 \times \bar{u}_{ij} \times d$. We call this transmission model the *basic* model. The parameters q_1 and q_2 can be interpreted as 'level' and 'shape' parameters, respectively. High values of q_2 give little importance to the exposure matrix and more importance to the contact matrix. The level parameter q_1 rescales the overall structure of suitable contacts to account for infection transmissibility that is not age-specific (both susceptibility and infectivity).

Our approach could be extended in several directions. For example, the im-

portance of different types of contacts for disease transmission can be evaluated by stratifying the contact matrix C by the level of intimacy of contacts, $C = C_1 + C_2$ where C_1 is the physical contacts matrix and C_2 the non-physical one [24, 26]. Since this information (i.e., proximity of contacts) is not available in the TUD, a simplified transmission model, called *intimacy* model, was considered here for illustrative purposes whereby exposure durations were assigned to contacts independently of the level of intimacy of contacts (e_{ij}/c_{ij}). We obtain $k_{ij} = (q_{1,p} \times c_{ij,p} + q_{1,np} \times c_{ij,np}) \times \bar{u}_{ij}$, where $q_{1,p}$ and $q_{1,np}$ correspond to the disease-specific parameters for, respectively, physical and non-physical contacts.

Statistical inference on transmission parameters

A second important contribution of this article is the application of state-of-the-art Bayesian techniques to evaluate parameter uncertainty and to incorporate prior epidemiological knowledge. The standard approach of estimating transmission parameters by fitting an age-structured SIR model at equilibrium based on the chosen contact matrix via maximum-likelihood, to cross-sectional unlinked data (contact and serological data) has been widely used in recent years [for extensive discussion see: 16, 17, 24, 26, 27, 28, 29]. In this article, we apply a novel approach, known as Bayesian melding [30, 31], developed to make statistical inference for models, as the standard age-structured SIR model, that transforms a vector of input parameters (e.g., θ) into a set of outputs (e.g., ρ: seroprofiles and R_0) in a deterministic way: $\rho = M(\theta)$. In our specific case, the vector θ of unknown parameters is, for the basic model $\theta = (q_1, q_2)$, and for the intimacy model $\theta = (q_{1,p}, q_{1,np}, q_2)$.

We translate our epidemiological knowledge into probabilistic statements by expressing a prior distribution for the Basic Reproductive Number (R_0) of the infection considered. R_0 is an output of our SIR model and the most important summary measure of infection transmissibility [1]. A large set of estimates and expert opinions for R_0 are available for many infections. Therefore, R_0 is a natural candidate for us to express a priori knowledge. In mathematical terms, R_0 is the dominant eigenvalue of the NGM [25, 32]. It follows that inputs and outputs in our model are linked through the deterministic function $R_0 = M(\theta)$, that maps the vector of unknown parameters into the dominant eigenvalue of the underlying NGM. A prior distribution on the outputs implicitly defines a prior distribution on the inputs [30], conditional on time of exposure and contact matrices. The implicitly defined prior is the so-called induced prior distribution: $p^*(\theta)$.

From the serological data (W) we compute the binomial likelihood, $p(W|M(\theta))$, i.e. the probability of the data conditional on the model input θ, given by:

$$L(\theta) = \prod_{i=1}^{N} p(W|M(\theta)) \propto \prod_{i=1}^{N} \pi(a_i)^{y_i}(1-\pi(a_i))^{(n_i-y_i)}$$

where $i = 1, .., N$ ($N = 79$) denotes the ith year of age; n_i is the corresponding serological sample size; y_i is the observed number of seropositive, and $\pi(a_i)$ is the prevalence predicted by the model for age group i. Infants younger than 1 year old are not considered because assumed to be protected by maternal antibodies.

The posterior distribution for the parameters of interest is obtained by combining priors and likelihoods using the Bayes rule:

$$p(\theta|W) \propto p^*(\theta)p(W|M(\theta))$$

We use the Sampling-Importance-Resampling algorithm to compute the posterior distributions of the input parameters [33, 34]. For each model specification (basic and intimacy), we use uniform distributions for direct priors of input and output parameters (qs, R_0). In particular, varicella R_0 is assumed to vary between 1 and 8 [24, 26]. Posterior Bayes factors, which are the standard goodness of fit measures for these Bayesian approaches [35] are used to compare different model fits.

RESULTS

Suitable contact matrices

By considering different values of the MSD, which we assume to be a proxy for infection transmissibility, our approach allows for an entire family of suitable contact matrices.

Figure 1 illustrates the main differences between three suitable matrices obtained by combining the TU baseline matrix (measured in minutes) with the Polymod matrix, under the assumption of an MSD of, respectively, 1, 10 and 20 minutes. In particular Figure 1(a) reports mean numbers of contacts along the main diagonals of the three suitable matrices considered, showing a massive decline in children's contacts (i.e. age groups 0-14), which are up to six times less as the MSD increases from 1 to 20 min, whereas older age groups are much less penalized. Figure 1(b),(c),(d) reports contacts of children aged respectively 0-4, 5-9, and 10-14, i.e. the groups mostly contributing to transmission, with other age groups. Again the stronger decline (still up to a factor six) regards contacts with other children age groups, while contacts with older individuals are much less

affected. Overall these effects sharply reduces the matrix assortativeness.

The simplest summary measure of this decline in the role of children in transmission is represented by the dominant eigenvalue of the contact matrices, i.e. the Basic Reproduction Number that would be observed under the social contact hypothesis sub a 100% probability of transmission per single contact [15]. It happens that as the MSD passes from 1 to 10 to 20 min, the dominant eigenvalue declines from 21.9 (very close to the figure of the Polymod matrix, given by 22.1) to 16.1 and to 12.5, suggesting a substantial decline in the overall transmissibility.

Fitting suitable matrices to data

Our basic model, based on the combination of contact and TU matrices, fits observed varicella serological profiles rather well (see Table 2). The fit is very similar to the one obtained using the Polymod contact matrix (Posterior Bayes Factor of 239.688 vs. 238.927), which is known to perform extremely well for varicella [17, 24, 26]. This result suggests that the concept of suitable contact is valuable, especially because it provides additional insights on the determinants of infection transmission.

The estimated posterior modes and credible intervals for q_1, q_2, and R_0 are reported in Table 2, and a comparison with the estimate of the q parameter for the Polymod baseline model is shown. The proportion q_2 of exposure that is suitable for transmission is around 26% (9.6%-94.1%) assuming that the MSD for varicella is one minute.

Prior and posterior distributions for parameters q_1, q_2, and R_0 are shown in Figure 2. Non informative uniform priors (solid lines) have been assumed. It is

worth noting that the posterior of q_2 shows a certain degree of skewness, whereas the ones for q_1 and R_0 are much more symmetric around their mean value of, respectively, 0.032 and 4.522.

In Figure 3, the appropriateness of the combined Polymod baseline and TU baseline fitting model to reproduce the observed serological data is shown: the estimated age-specific force of infection is high at around the ages of 5-6 and 40 years, and relatively low for other age groups [as in previous works: 24, 36].

The results of the intimacy model are reported in Table 2, for comparison with the model based on Polymod data only (Polymod baseline) [26, 27]. The posterior mode for $q_{1,p}$, is larger than the corresponding mode of $q_{1,np}$ confirming that physical contacts are more important in explaining infection transmission than non-physical contacts [24, 26]. However, the unavailability of proximity information for the TUD limits the validity of the 'intimacy' experiment. Indeed, the posterior Bayes factor from using the suitable matrix is higher in the intimacy model than in the basic model (242.204 vs 239.688).

DISCUSSION

In recent years, the availability of social contacts data [15, 21, 37] as well as TUD [16] has improved our understanding of mixing patterns, infectious disease modeling, and the evaluation of alternative intervention strategies [9, 38, 39, 40].

In previous works, the focus was on the concept of "contacts". However, the number of social contacts is only one among the critical variables that characterize individual interactions and cause infection transmission. Ideally, we would like to know other quantities relevant for transmission, such as the amount of excreted

infectious virus (e.g., through coughing, sneezing or exhalation) and its propagation dynamics [41, 42, 43]. In practice, the difficulties in measuring these variables induce us to consider possibly correlated quantities, such as the duration of exposure between individuals. For instance, if a sneeze during a contact substantially increases the probability of transmission, then a longer contact is more likely to lead to transmission of the infection, as it is more likely that at least one sneeze occurs during the contact. In other words, it is highly likely that the occurrence of some "suitable events" (e.g., kiss, handshake, sneeze, etc.) is positively related to the duration of the contact. For highly transmissible infections (e.g., measles) a short duration of exposure between an infected individual and a susceptible one might be sufficient for transmission. If the infection is not highly transmissible, a longer duration of exposure may be needed. Does this mean that many contacts are in this case "wasted" for transmission? Is there an MSD of exposure below which transmission cannot occur?

The present paper attempts to answer the previous questions by integrating into a single unified contact model the number of encounters (from contact survey) and the duration of exposure (from TUD). The main idea is that a contact is "suitable" for transmission only if it lasts for a sufficiently long (infection-specific) time.

Our approach allows us to generate a large class of contact matrices by appropriately varying the dimensional unit of the exposure matrix, taken as a proxy of the MSD across different infections. For small values of the MSD, we obtain the standard contact matrix itself (i.e., Polymod matrix), suggesting that for highly transmissible infections only the number of contacts matters. For larger values of the MSD (infections with lower transmissibility), the number of contacts becomes

less critical, as some of these contacts might not receive enough exposure duration and therefore are not suitable. Thus, the ensuing contact matrix is less assortative because age groups with large numbers of contacts are more heavily penalized for the lack of exposure time.

Our method was tested against Italian seroprevalence data for varicella. The results are, in terms of goodness of fit, consistent with previous works [24, 26]. For varicella, the required duration of exposure is fairly small, confirming that varicella is a highly transmissible infection.

Our work is the first attempt to develop a consistent framework to disentangle the relative importance of number of contacts and duration of exposure using the most relevant available data sets. In addition, we use Bayesian techniques to formally incorporate epidemiological knowledge about quantities of interest. The unified framework generates entire posterior distributions for model parameters. The distributions may reveal asymmetries in uncertainty that are important to consider when planning public health interventions.

This work opens several possibilities for future research. First, it will be important to test the model against less transmissible infections such as bacterial infections [44, 45, 46], or viruses like parvovirus B19 and Hepatitis A [47, 48, 49], or pandemic influenza in outside-of-school settings [50]. Second, it will be important to improve the approach to generate suitable matrices by deepening our understanding of contacts and TUD, and their inter-relationships. For example, the duration of exposure matrix used here relies on the assumption of no assortativeness at the level of single activities and time slots recorded by the TUD [as in 16]. Clearly, duration of exposure might be more assortative and this could be investigated by integrating TUD and Polymod contact data stratified by the

duration of contacts. Third, it will be relevant for control interventions to study if the proportion of contact suitable for transmission is constant in different settings (ie. household, school, work-place).

In conclusion, our approach generates noteworthy insights regarding the fraction of time of exposure between age groups that is suitable for transmission. This new methodology is relevant to evaluate the impact of public health interventions such as vaccination, screening or distance-based measures, on disease burden. For the first time, both number and duration of contacts are considered, and their relative importance in explaining infection data is evaluated. Though more work is needed to fully disentangle their role, our analysis provides a tool for generating a class of contact matrices that can be used to model infections with different levels of transmissibility.

Acknowledgments

Authors affiliations

University of Groningen, Department of Pharmacy, Groningen, The Netherlands. (Elisabetta De Cao); Max Planck Institute for Demographic Research, Rostock, Germany (Emilio Zagheni); Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Pisa, Italy (Piero Manfredi); and Dondena Centre for Research on Social Dynamics, Bocconi University (Alessia Melegaro).

Authors contributions

All authors contributed fully to the final manuscript. EZ conceived the question addressed by this article and developed the concept of suitable contact matrix. EDC developed the programme in R, incorporated the Bayesian framework and produced results under the supervision of AM. PM analyzed the properties of suitable contact matrices and contributed to the interpretation of the results. EDC and AM wrote the first draft of the manuscript. AM and PM reshaped some parts of the manuscript which was then revised and approved by all authors.

Grants and funding

EDC, PM, AM were partly funded by the European Centre for Disease Prevention and Control (ECDC) under grant 2009/002 (project title: "Vaccine preventable diseases modelling in the European Union and EEFTA countries") to the Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, coordinator Piero Manfredi.

Acknowledgments

The authors thank Tommi Asikainen, Pierluigi Lopalco, Paloma Carrillo, Bruno Ciancio, Chantal Quinteen, Piero Benazzo and Luca Fulgeri from ECDC for scientific advice and financial support; Rebecca Graziani and Nele Goeyvaerts for valuable comments, and the Epidemics[3] Conference participants for helpful discussions on the topic of this work.

Presentations

EDC presented a poster at the Alpine Population Conference (Alp-Pop), La Thuile, Italy, January 15-18, 2012. EDC presented a poster at the Third International Conference on Infectious Disease Dynamics (EPIDEMICS3), Boston, US, November 29-December 2, 2011. EDC presented a poster at the PAA 2011 Annual Meeting, Washington, US, March 31-April 2, 2011. EZ presented a poster at the PAA 2009 Annual Meeting, Detroit, US, April 30-May 2, 2009, and he won the Poster Session.

Conflict of interest

None declared.

References

[1] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, United Kingdom: Oxford University Press, 1991.

[2] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. New York, NY: John Wiley and Sons, Inc, 2000.

[3] R. M. Anderson and R. M. May. Vaccination against rubella and measles: Quantitative investigations of different policies. *J Hyg (Lond)*, 90:259–325, 1983.

[4] H. W. Hethcote. *Models for infectious human diseases: Their structure and relation to data*, chapter Modelling heterogeneous mixing in infectious diseases dynamics, pages 215–238. Cambridge: Isham V and Medley G., 1995.

[5] M. E. Halloran, S. L. Cochi, T. A. Lieu, M. Wharton, and L. Fehrs. Theoretical epidemiologic and morbidity effects of routine varicella immunization of preschool children in the united states. *American Journal of Epidemiology*, 140:81–104, 1994.

[6] H. R. Babad, D. J. Nokes, N. J. Gay, E. Miller, P. Morgan-Capner, and R. M. Anderson. Predicting the impact of measles vaccination in England and Wales: Model validation and analysis of policy options. *Epidemiol Infect*, 114:319–344, 1995.

[7] W. J. Edmunds, G. F. Medley, and D. J. Nokes. Evaluating the cost-

effectiveness of vaccination programmes: A dynamic perspective. *Stat Med*, 18:3263–3282, 1999.

[8] W. J. Edmunds, O. G. van de Heijden, M. Eerola, and N. J. Gay. Modelling rubella in europe. *Epidemiol Infect*, 125:617–634, 2000.

[9] M. Brisson, W. J. Edmunds, N. J. Gay, B. Law, and G. De Serres. Modelling the impact of immunization on the epidemiology of varicella zoster virus. *Epidemiol Infect*, 125:651–669, 2000.

[10] M. Brisson, W. J. Edmunds, and N. J. Gay. Varicella vaccination: impact of vaccine efficacy on the epidemiology of VZV. *Journal of Medical Virology*, 70(1):S31–7, 2003.

[11] N. J. Gay. The theory of measles elimination: Implications for the design of elimination strategies. *J Infect Dis*, Suppl 1:S27–S35, 2004.

[12] C. L. Trotter, N. J. Gay, and W. J. Edmunds. Dynamic models of meningococcal carriage, disease, and the impact of serogroup c conjugate vaccination. *American Journal of Epidemiology*, 162:89–100, 2005.

[13] A. Melegaro, Y. H. Choi, R. George, W. J. Edmunds, E. Miller, and N. J. Gay. Dynamic models of pneumococcal carriage and the impact of the heptavalent pneumococcal conjugate vaccine on invasive pneumococcal disease. *BMC Infect Dis*, 10(90), 2010.

[14] A. J. van Hoek, A. Underwood, M. Jit, et al. The impact of pandemic influenza H1N1 on health-related quality of life: a prospective population-based study. *Plos One*, 6(3):e17030, 2011.

[15] J. Wallinga, P. Teunis, and M. Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164:936–944, 2006.

[16] E. Zagheni, F. C. Billari, P. Manfredi, et al. Using time use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*, 168(9):1082–1090, 2008.

[17] F. Iozzi, F. Trusiano, F. C. Billari, et al. Little italy: An agent-based approach to the estimation of contact patterns-fitting predicted matrices to serological data. *PLoS Computational Biology*, 6(12), 2010.

[18] J. Wallinga, W. J. Edmunds, and M. Kretzschmar. Human contact patterns and the spread of airborne infectious diseases. *Trends Microbiol.*, 7:372–377, 1999.

[19] W. J. Edmunds, C. J. O'Callaghan, and D. J. Nokes. Who mixes with whom? a method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc R Soc Lond B*, 264:949–957, 1997.

[20] P. Beutels, Z. Shkedy, and *et al.* Aerts, M. Social mixing patterns for transmission models of close contact infections: Exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiol Infect*, 134(6):1158–66, 2006.

[21] J. Mossong, N. Hens, and M. others Jit. Social contacts and mixing patterns relevant to the spread of infectious diseases. *Plos Med*, 5(3):74, 2008.

[22] S. Y. Del Valle, J. M. Hyman, H. W. Hethcote, et al. Mixing patterns between age groups in social networks. *Soc Networks*, 29:539–554, 2007.

[23] A. Nardone, F. de Ory, M. Carton, et al. The comparative sero-epidemiology of varicella zoster virus in 11 countries in the european region. *Vaccine*, 25:7866–7872, 2007.

[24] N. Goeyvaerts, N. Hens, B. Ogunjimi, M. Aerts, Z. Shkedy, P. Van Damme, and P. Beutels. Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society: Series C*, 59(2):255–277, 2010.

[25] O. Diekmann, J. A. Heesterbeek, and 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(4):365–382, 1990.

[26] A. Melegaro, M. Jit, E. Zagheni, et al. What types of contacts are important for the spread of infections? using contact survey data to explore european mixing patterns. *Epidemics*, 3(3-4):143–151, 2011.

[27] B. Ogunjimi, N. Hens, N. Goeyvaerts, M. Aerts, and P. Beutels. Using empirical social contact data to model person to person infectious disease transmission: An illustration for varicella. *Mathematical Biosciences*, 278(2):80–87, 2009.

[28] N. Hens, G. Minalu Ayele, and N. others Goeyvaerts. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight european countries. *BMC Infectious Diseases*, 9(187):1–12, 2009.

[29] M. Kretzschmar, P. F. M. Teunis, and R. G. Pebody. Incidence and reproduction numbers of pertussis: Estimates from serological and social contact data in five european countries. *PLoS Med*, 7(6):e1000291, 2010.

[30] A. E. Raftery, G. H. Givens, and J. E. Zeh. Inference from a deterministic population dynamics model for bowhead whales (with discussion). *J. Amer. Statist. Assoc.*, 90:402–416, 1995.

[31] D. Poole and A. E. Raftery. Inference for deterministic simulation models: The bayesian melding approach. *J. Amer. Statist. Assoc.*, 95:1244–1255, 2000.

[32] J. A. Heesterbeek. R_0. PhD thesis, University of Leiden, 1992.

[33] D.B. Rubin. Comment, on "The calculation of posterior distributions by data augmentation", by M. Tanner and W. H. Wang. *J. Amer. Statist. Assoc.*, 82:543–546, 1987.

[34] D.B. Rubin. Using the sir algorithm to simulate posterior distributions. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics 3*, pages 395–402. Oxford, UK: Clarendon Press, 1988.

[35] M. Aitkin. Posterior bayes factor. *J. R. Statist. Soc. B*, 53(1):111–142, 1991.

[36] J. Mossong, L. Putz, and F. Schneider. Seroprevalence and force of infection of varicella-zoster virus in Luxembourg. *Epidemiol Infect.*, 132(6):1121–1127, 2004.

[37] P. Horby, P. Quang Thai, N. Hens, et al. Social contact patterns in vietnam and implications for the control of infectious diseases. *PLoS One*, 6(2), 2011.

[38] Y. H. Choi, M. Jit, G. Nigel, et al. 7-valent pneumococcal conjugate vaccination in England and Wales: Is it still beneficial despite high levels of serotype replacement? *Plos One*, 6(10):e26190, 2011.

[39] A. J. van Hoek, A. Melegaro, E. Zagheni, W. J. Edmunds, and N. Gay. Modelling the impact of a combined varicella and zoster vaccination programme on the epidemiology of varicella zoster virus in England. *Vaccine*, 29(13):2411–20, 2011.

[40] A. J. van Hoek, A. Melegaro, N. Gay, et al. The cost-effectiveness of varicella and combined varicella and herpes zoster vaccination programmes in the united kingdom. *Vaccine*, 30(6):1225–34, 2011.

[41] P. F. M. Teunis, N. Brienen, and M. E. E. Kretzschmar. High infectivity and pathogenicity of influenza a virus via aerosol and droplet transmission. *Epidemics*, 2(4):215–22, 2010.

[42] T. P. Weber and N. I. Stilianakis. Inactivation of influenza A viruses in the environment and modes of transmission: a critical review. *J Infect*, 57(5):361–73, 2008.

[43] T. P. Weber and N. I. Stilianakis. Review of aerosol transmission of influenza A virus. *Emerging Infect Dis*, 12(11):1657–66, 2006.

[44] T. Leino, K. Auranen, PH Mäkelä, et al. *Haemophilus influenzae* type b and cross-reactive antigens in natural *Hib* infection dynamics; modelling in two populations. *Epidemiol Infect*, 129(1):78–83, 2002.

[45] T. Leino, T. Takala, K Auranen, et al. Indirect protection obtained by

Haemophilus influenzae type b vaccination: analysis in a structured population model. *Epidemiol Infect*, 132(5):959–66, 2004.

[46] A. Melegaro, Y. Choi, R. Pebody, and N. Gay. Pneumococcal carriage in united kingdom families: Estimating serotype-specific transmission parameters from longitudinal data. *American Journal of Epidemiology*, 166(2):228–235, 2007.

[47] N. Hens, M. Aerts, Z. Shkedy, et al. Estimating the impact of vaccination using age-time-dependent incidence rates of hepatitis b. *Epidemiol Infect*, 136(3):341–51, 2008.

[48] N. Goeyvaerts, N. Hens, M. Aerts, and P. Beutels. Model structure analysis to estimate basic immunological processes and maternal risk for parvovirus b19. *Biostat*, 12(2):283–302, 2011.

[49] M. Andraud, O. Lejeune, J. Z. Musoro, et al. Living on three time scales: the dynamics of plasma cell and antibody populations illustrated for hepatitis a virus. *PLoS Comput Biol*, 8(3), 2012.

[50] A. C. Ghani, M. Baguelin, J. T. Griffin, et al. The early transmission dynamics of H1N1pdm influenza in the united kingdom. *PLoS Curr*, 2009.

Appendix: Occupancy problem

Consider the number of suitable minutes of contact between groups i and j, $(q_2 e_{ij})$, as 'balls', and the number of contacts between groups i and j, (c_{ij}), as 'boxes', then the expected number of suitable contacts between the two age groups can be thought of as the expected value of occupied boxes from randomly assigned balls.

To compute this expected value, define the indicator function:

$$Z_i = \begin{cases} 1 & \text{if the contacted person } i \text{ receives } \textit{zero} \text{ suitable minutes;} \\ 0 & \text{otherwise.} \end{cases}$$

We obtain (by Poisson approximation)

$$E[Z_i] = Pr[Z_i = 1] = (1 - \frac{1}{c_{ij}})^{q_2 e_{ij}} \approx e^{-q_2 e_{ij}/c_{ij}}$$

Consider now $Z = \sum_{i=1}^{c_{ij}} Z_i$. The variable Z is the total number of contacted people who do not receive any suitable minute of transmission. Its expected value is:

$$E[Z] = \sum_{i=1}^{c_{ij}} E[Z_i] \approx \sum_{i=1}^{c_{ij}} e^{-q_2 e_{ij}/c_{ij}} = c_{ij} e^{-q_2 e_{ij}/c_{ij}}.$$

Therefore the expected number of suitable contacts between age groups i and j is $c_{ij}(1 - e^{-q_2 e_{ij}/c_{ij}})$.

Figures and Tables

Figure 1: Patterns of suitable contacts in different age groups for three distinct values of the msd (in minutes): msd=1, msd=10, msd=20. Panel(a): contacts along the main diagonal of the suitable matrix; panel (b): contacts of children 0-4 years old; panel (c): contacts of children 5-9 years old; panel (d): contacts of children 10-14 years old. The vertical axis reports the ratio between numbers of suitable contacts and the smallest entry observed in the corresponding matrix.

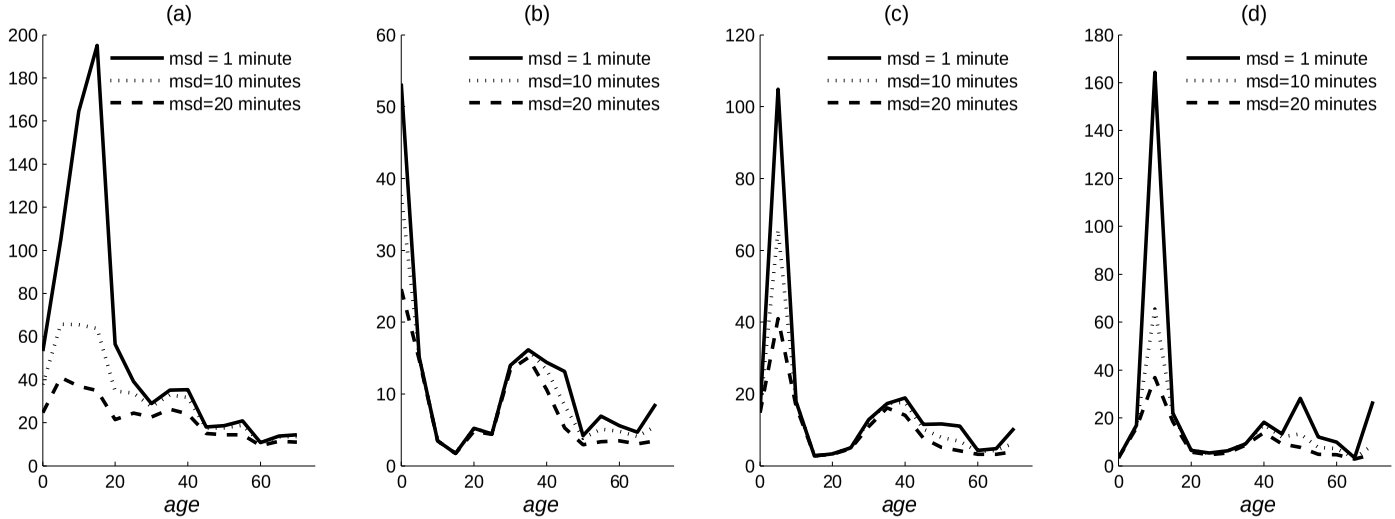

Figure 2: Prior (solid line) and posterior (histograms) distributions for the input parameters q_1 and q_2 and the output R_0.

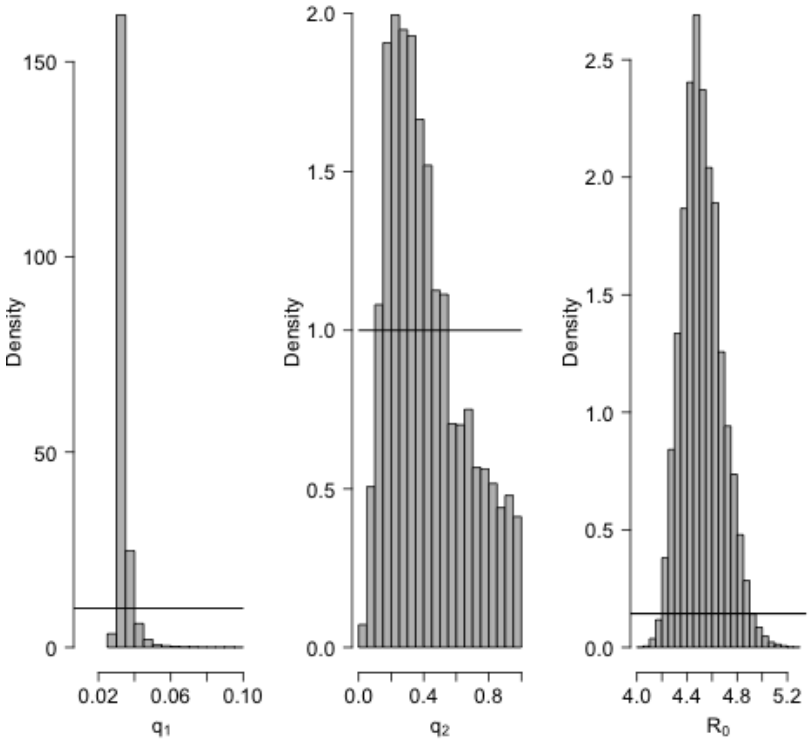

Figure 3: Fit of the model to serological VZV data for Italy. Points are observed serological data with size proportional to the corresponding sample size; solid line is the estimated prevalence when q_1 and q_2 are equal to their posterior modes; dashed lines are the estimated prevalence when q_1 and q_2 are equal to 2.5% and 97.5% quantiles of their posterior distribution. The bold solid line at the bottom of the graph is the estimated force of infection with its minimum and maximum values reported.

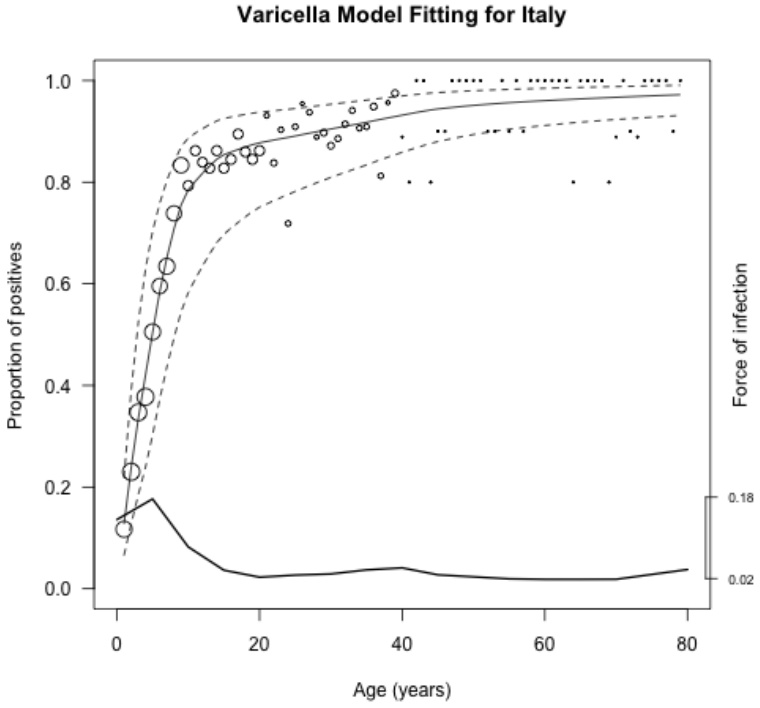

Table 1: Definition and Description of the Alternative "C" and "E" Matrices Used as Inputs for the Suitable Contact Matrices.

Matrix	**Type (contact/exposure)**	**Source**
Polymod_baseline	contact (all)	Survey data (Mossong et al 2008)
Polymod_physical	contact (physical)	Survey data (Mossong et al 2008)
Polymod_nonphysical	contact (non-physical)	Survey data (Mossong et al 2008)
TU_baseline	duration of exposure	Italian TUD 2002-2003 Zagheni et al. (2008) method

Table 2: Fits of the Basic and Intimacy Models Based on the Suitable Matrix. Fits From Standard Contact and Time Use Matrices are Also Reported for Comparison Purposes. Italy, 1997-2006†.

Basic Model	**Posterior Bayesian Factor**	**Posterior Mode** q_1		**95% Credible Interval**		**Posterior mean** q_2	**95% Credible Interval**	**Posterior Mode** R_0	**95% Credible Interval**
Polymod_baseline-TU_baseline	239.688	0.033		0.030,0.043		0.260	0.096,0.941	4.381	4.243,4.307
		Posterior Mode q		**95% Credible Interval**				**Posterior Mode** R_0	**95% Credible Interval**
Polymod_baseline	238.927	0.031		0.030,0.033				4.734	4.520,4.987
TU_baseline	249.340	0.0012		0.0011,0.0013				4.255	4.039,4.449
Intimacy Model	**Posterior Bayesian Factor**	**Posterior Mode** $q_{1,n}$	**95% Credible Interval**	**Posterior Mode** $q_{1,np}$	**95% Credible Interval**	**Posterior Mode** q_2	**95% Credible Interval**	**Posterior Mode** R_0	**95% Credible Interval**
Polymod_baseline-TU_baseline	242.204	0.043	0.034,0.063	0.0027	0.0004,0.0400	0.554	0.067,0.976	3.560	3.362,4.656
Polymod_baseline	239.871	0.039	0.033,0.041	0.010	0.010,0.030			3.956	3.864,4.710

†Serological samples were collected between 1997 and 2003, Time Use Data in 2002-03, Polymod data in 2006.
Estimates of the q parameters are not comparable due to the different scales of the underlying matrices.