The social structure underlying the spread of childhood close-contact infectious diseases. Combining time use and contact surveys to explain age-specific seroprevalence profiles

Emilio Zagheni
University of California, Berkeley, USA
Francesco C. Billari
Università Bocconi, Milano, Italia
Piero Manfredi
Università di Pisa, Pisa, Italia

September 22, 2008

Abstract

Demographic and social processes play an important role in preventing or sustaining the spread of infectious diseases in human societies. In this paper, we discuss the possibilities offered by two independent social data sources - time use data and contact surveys - to explain the observed age-specific seroprevalence profile for two close-contact childhood infections (varicella and parvovirus B19). We first discuss the epidemiological consequences of the assumption that age-specific potentially infectious contacts are proportional to self-reported age-specific number of social contacts or age-specific estimates of time of exposure. Then we propose a model that combines information from age-specific social contact and time of exposure matrices. The model is used to evaluate the impact of behavioral changes and specific interventions (e.g., school closure) on the spread of close-contact infections. Empirical analysis is based on data for Italy, where a comprehensive and up-to-date collection of serologic and social data is available.

EXTENDED ABSTRACT

Introduction

Social mixing patterns are a relevant explanatory factor for the spread of close-contact infectious diseases [1, 2]. The exposure and frequency of contacts between people belonging to different age groups is strongly dependent on demographic and social variables. Population age structure and household size are important demographic determinants of observed contact patterns. Social structure is another crucial element: for instance, the likelihood of interaction between people of different age groups depends on norms that influence the location where they spend time (e.g. work, school, home, restaurant, public transportaton, etc.) and the time slots of the day during which specific activities take place.

Three main approaches have been suggested in the literature to estimate mixing patterns using social data. A first approach relies on contact surveys in which the respondent self reports the number of contacts he has during a randomly sampled day, together with some additional information (i.e., age of contacted person, whether the contact is physical or not, etc.) [3, 4, 5, 6, 7]. A second approach relies on micro-simulation: contact and time of exposure matrices are obtained as output of a simulation informed by secondary data such as transportation data [8]. A third approach relies on time use surveys: time of exposure matrices are estimated from time use diaries, assuming that proportional mixing holds at the level of single location and for short time slots [9].

In this paper, we first discuss the role of demographic and social structure in shaping mixing patterns. We then propose a model that combines information from time use and contact surveys. Finally, we test the ability of the model to fit serologic data and we use the model to evaluate the impact of specific public health interventions.

Data

We use data for Italy, for which we have one of the most comprehensive and up-to-date collections of social data (time use and contact surveys) and serologic data for varicella zoster virus and parvovirus B19.

Data on time use were collected by the Italian National Statistical Agency (ISTAT) in 2002-2003 on a sample of about 24 thousand households. Time use data were collected in the form of diaries in which the respondent records the activities that he did during the day and the location where the activity took place.

The contact survey for Italy was collected as part of POLYMOD, a project funded by the European Union. A sample of 849 respondents were asked to self report the number of contacts they had during a randomly sampled day, together with some additional information (see [7] for details).

Serological samples were collected and tested as part of POLYMOD for antibodies to varicella zoster virus and parvovirus B19. These are infections for which vaccination programmes are not yet in place in Italy. Therefore the data describe the natural history of the disease. The sample size is 2517 and the age of participants ranges from 0 to 79 years.

The Model

In epidemiology, a fundamental quantity is the age-specific force of infection (λ_i), that is the rate at which susceptible individuals in the age group i become infected. In standard infectious diseases modeling, the force of infection is proportional to the transmission rates between and within age groups, β_{ij}:

$$\lambda_i = \sum_j \beta_{ij} \times Y_j \tag{1}$$

where Y_j is the number of infectives at steady state in age group j. Traditionally, the transmission rates, which form the "who-acquires-infection-from-whom" matrix, are estimated 'indirectly' from epidemiological data, under suitable simplifying assumptions [1]. More recently 'direct' approaches have been suggested: the transmission rates matrix is assumed to be proportional to either a contact matrix or a time of exposure matrix estimated from sample surveys [5, 9].

In the case of Italy, we have two independent data sources that give us information on time of exposure (i.e., time use survey) and number of contacts (i.e., contact survey). In this paper we propose a model that combines these two data sources.

We can think that a fraction s of the average time of exposure between groups i and j, (E_{ij}), is suitable for transmission of the disease in terms of proximity of contact, physical condition of the location of contact, etc. People in the age groups i and j have a certain daily number of contacts on average, (C_{ij}), which differ in terms of duration. If we assume that a person randomly distributes her/his suitable minutes for transmission to people she/he has contact with, then some people may receive more than one suitable minute, whereas some others may not receive any of them. If the disease is highly transmissible, what matters for transmission between two people is that they have a contact with at least one suitable minute of exposure. We call this kind of contact 'suitable contact'. We do not know what kind of contacts are 'suitable', but we assume that the likelihood that people experience at least one such contact is positively related to the duration of their contacts.

If we think of our problem in these terms, we can use some results from classic probability problems, such as the 'occupancy problem', to obtain the expected number of suitable contacts between age groups i and j, (A_{ij})[1]:

$$E[A_{ij}] = C_{ij}(1 - e^{-sE_{ij}/C_{ij}}) \tag{2}$$

If we assume that the age-specific transmission rates are proportional to the age-specific number of suitable contacts, then, by multiplying the quantity in expression 2 by a parameter l that represents a disease-specific infectivity parameter, we obtain:

$$\beta_{ij} = l \times C_{ij}(1 - e^{-sE_{ij}/C_{ij}}) \tag{3}$$

The parameters l and s can be interpreted as 'level' and 'shape' parameters, respectively. In this setting, high values of s give little importance to the exposure matrix and more importance to the contact matrix. The level parameter l then 'adjusts' the structure of suitable contacts to account for the degree of infectivity of the disease.

The force of infection can be obtained by plugging equation 3 into equation 1. The parameters s and l can be then estimated from serologic data, using a maximum likelihood technique.

Preliminary Results and Discussion

Figure 1 shows a contour plot of the estimated average daily time that people in the age group i spend with people in the age group j in Italy. The estimates are obtained, respectively, by using the approach developed in [9] on Italian time use data and from the contact survey for Italy [7]. In both cases, the highest values are on the main diagonal, implying assortativeness. In the case of the contact matrix, the highest values are more concentrated along the main diagonal, compared to the estimated time of exposure.

Figure 2 shows the fit of the model to serologic data, based on maximum likelihood estimates for the parameters l and s: we observe a good fit to the seroprevalence data.

Both data on time use and number of contacts are available for single locations (e.g., school, home, workplace): we can thus use this information to obtain a more detailed representation of contact patterns and we can write the elements of the β matrix as a combination of suitable contacts in the n different settings considered:

$$\beta_{ij} = l \times [C^1_{ij}(1 - e^{-s^1 E^1_{ij}/C^1_{ij}}) + \cdots + C^n_{ij}(1 - e^{-s^n E^n_{ij}/C^n_{ij}})] \tag{4}$$

The representation in equation 4 allows us to evaluate the impact of specific interventions such as school closure (e.g., elimination of the setting 'school' and adjustment of the other settings to levels observed during vacation time) or behavioral changes (e.g., modification of the s parameter for a specific setting). The effect of these interventions is discussed in the paper.

Notes

[1]The result can be derived as follows. If we think of the number of suitable minutes of contact between groups i and j, (sE_{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

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

We have

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

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

$$E[Z] = \sum_{i=1}^{C_{ij}} E[Z_i] \approx \sum_{i=1}^{C_{ij}} e^{-sE_{ij}/C_{ij}} = C_{ij} e^{-sE_{ij}/C_{ij}}$$

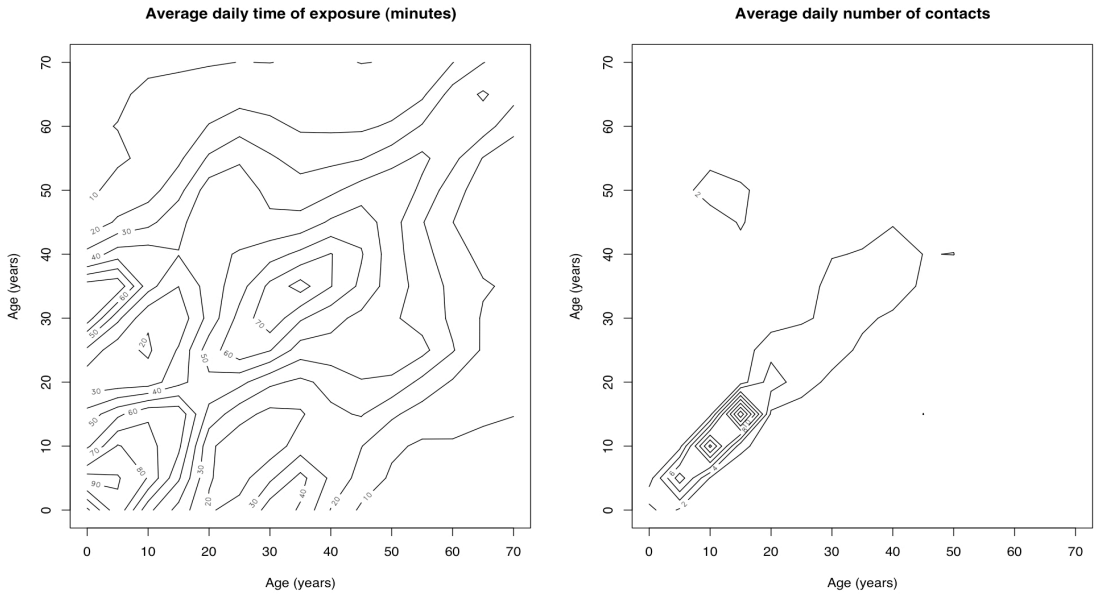

Figure 1: Mixing patterns for Italy estimated from time use and contact surveys.

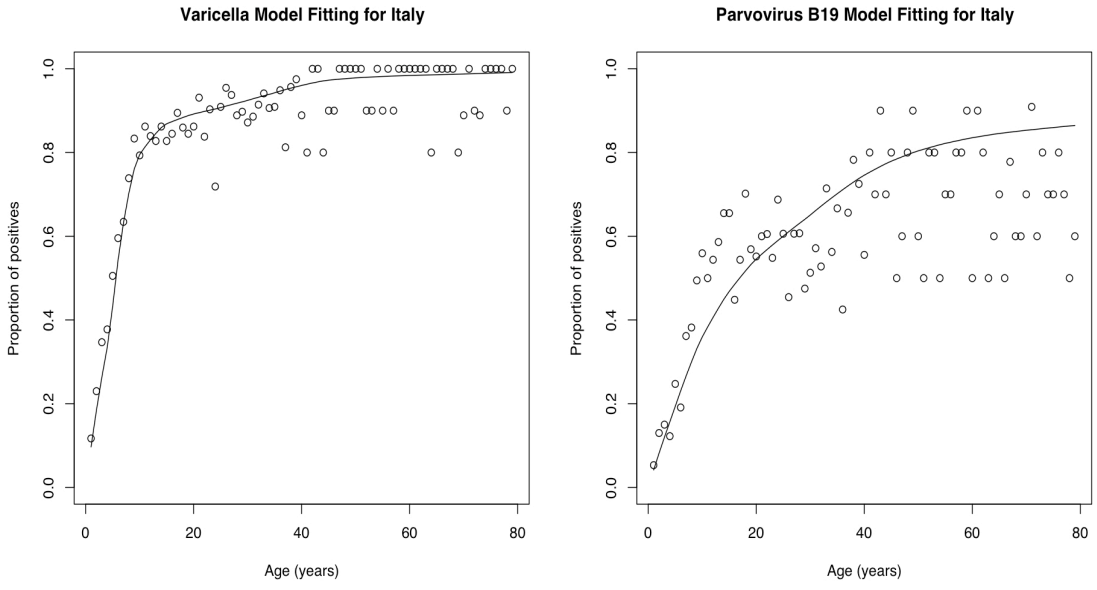

Figure 2: Fit of the proposed model to serologic data for Italy. Points are serologic observations; dashed lines represent the fit of the model.

Bibliography

[1] Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford, United Kingdom: Oxford University Press, 1991.

[2] Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. New York, NY: John Wiley and Sons, Inc, 2000.

[3] Wallinga J, Edmunds WJ, Kretzschmar M. Human contact patterns and the spread of airborne infectious diseases. Trends Microbiol. 1999;7:372-7.

[4] Edmunds WJ, O'Callaghan CJ, Nokes DJ. 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 1997;264:949-57.

[5] Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol. 2006;164:936-944.

[6] Beutels P, Shkedy Z, Aerts M, et al. 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 2006;134(6):1158-66.

[7] Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. Plos Med 2008;5(3) e74.

[8] Del Valle SY, Hyman JM, Hethcote HW, et al. Mixing patterns between age groups in social networks. Soc Networks 2007; 29:539-554.

[9] Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J and Edmunds, WJ. Using Time Use Data to Parameterize Models for the Spread of Close-contact Infectious Diseases. Am J Epidemiol. 2008; doi: 10.1093/aje/kwn220