Age-dependent Contact Patterns and Time Use Data

Emilio Zagheni , *University of California, Berkeley*
Francesco C. Billari, *Università Bocconi, Milano*
Piero Manfredi, *Università di Pisa*
Joel Mossong, *Laboratoire National de Santé, Luxembourg*

September 21, 2006

ABSTRACT: Social contact patterns are the critical explanatory factor of the spread of directly transmitted infectious diseases. Both indirect (via observed epidemiological data) and direct (via diaries that record 'at risk' contacts) approaches to the measurement of contacts by age have been proposed in the literature. In this paper we systematically discuss the possibilities offered by Time Use Surveys (TUS) for the estimation of contact patterns. TUS provide diary-structured data about the activities undertaken by sampled individuals and/or the locations where the activities occurred. Focusing on the notion of 'mixing by activity/location and time slot', we develop a methodology to estimate time of exposure contact matrices and mixing matrices. We provide estimates of these matrices with regard to the U.S. and we show that the allocation of time to activities is the major source of assortativeness of contact patterns by age.

Social contact patterns are critical factors in the explanation of the transmission of infectious agents. For several diseases, such as measles, mumps, rubella and influenza, the transmission of the pathogen from person to person occurs via small airborne infectious droplets: effective transmission requires close contacts between individuals. Mixing patterns thus play an important role in the transmission dynamics of respiratory infections and their analysis provides decision-makers with useful information concerning the specific groups in a population that should be targeted by vaccination measures. In spite of the relevance of a good understanding of mixing patterns to forecast the spread of infectious agents and the efficacy of prevention programmes, the dynamics of contact patterns that lead to the diffusion of infections are still unclear. On the one hand, there is a need to obtain direct estimates of contact behaviors; on the other hand, it is difficult to define a suitable measure of at-risk contact (Edmunds et al. (1997)).

In this paper we try to quantify contact patterns in terms of time of exposure to other people, within and between age groups. We estimate a time of exposure contact matrix for the U.S., from time use data, and we try to get some insights on age-dependent mixing patterns.

The problem of relating the patterns of disease spread to the underlying socio-epidemiological mechanisms of the spread of the disease agent has been addressed, in the standard mathematical modeling, by developing a matrix representation of contact patterns. Three types of matrices have been widely used: contact matrices, mixing matrices and transmission parameter matrices. Contact matrices give the number of contacts that people in group i have with people in group j, per unit of time. Mixing matrices give the fraction of the contacts that people in group i have with people in group j, per unit of time. Transmission parameter matrices include the effects of the probability of transmission per single contact between a susceptible person in group i and an infectious individual in group j.

In the first phase of epidemiological modeling, the host population was assumed to homogeneously mix, i.e. individuals were assumed to randomly mix, with identical contacts and transmission parameters. Nonetheless, it was recognized that populations are made up of subgroups and that heterogeneity must affect the process of disease transmission.

The early attempts to introduce heterogeneity into the modeling of contact patterns led to the proportionate mixing approach (Nold (1980)): within this framework, contacts between individuals are still considered random, but in a heterogeneous population. Age, geographical separation and population density are some of the main sources of heterogeneity and they are taken into account when modeling disease transmission.

The proportionate mixing approach represents a first step towards the improvement of the homogeneous mixing scheme, but it does not take into account assortativeness (the fact that people tend to have more contacts with individuals similar to them, for instance in terms of age). In order to overcome this limitation and to model heterogeneity in a more realistic way, several approaches have been developed. On the one hand, Anderson and May (1985) suggested an approach based on the 'WAIFW'(Who Acquires Infection From Whom) matrix. On the other hand, several authors (e.g. Nold (1980), Jacquez et al. (1988) and Hethcote (1989)) have contributed to the development of the 'preferred' and 'structured' mixing models. Finally, a further line of development concentrated on directly filling in the elements of a contact matrix (e.g. Edmunds et al. (1997)).

Anderson and May (1985) focused on indirectly assessing the patterns of contacts within and between age groups. They defined a n x n WAIFW matrix, the elements of which represent effective contacts (those contacts that are likely to result in the transfer of infection) between individuals in age group i and age group j. Estimates of this matrix were derived from estimates of the force of infection for each of the n age groups. This approach has been widely used, but it shows several limitations: for example, the number of distinct mixing rates is constrained to be n, instead of n x n. As a consequence, the researcher is required to choose between certain constrained matrix structures that are unlikely to accurately represent the true contact patterns.

The preferred mixing approach was developed to improve the proportional mixing scheme. It is an attempt to introduce assortativeness: the contact matrix is built as a convex combination of a proportionate mixing matrix and a diagonal (fully assortative) mixing matrix. The basic idea behind this scheme is that people are allowed to reserve an arbitrary fraction of their group's contacts for within-group contacts while the remaining contacts are subject to proportional mixing. This type of matrix was used, for instance, by Jacquez et al. (1988) in their studies about HIV/AIDS transmission.

Edmunds et al. (1997) tried out a direct approach to estimate contact rates: they defined an 'at-risk' contact as a two way conversation and they collected data through diaries in which respondents were asked to record some information about the people they had conversations with throughout the day.

In this paper we investigate contact patterns by means of time use data. We assume that the time of exposure to people, by age, is a good measure for contact patterns that could lead to the spread of airborne infections. In that sense our aim is to build a contact matrix based on time use data and to get insights on mixing patterns, such as the level of assortativeness by age.

MATERIALS AND METHODS

Data

Our analysis mainly relies on data from the "American Time Use Survey (ATUS) - 2003". The ATUS is a continuously conducted survey on time use in the United States, whose goal is to measure how people divide their time among life's activities. ATUS covers all residents living in households in the United State that are at least 15 years old[1]. The ATUS sample is composed of approximately 40500 households annually.

Data are collected in the form of diaries in which the respondent is supposed to record his daily activities chronologically. For most activities the respondent is also asked to specify where the activity took place (e.g. respondent's home, respondent's workplace, restaurant, school, mall, car, bus, etc.) and whether he was in company of somebody else while doing the activity (e.g. alone, spouse, own household child, roommates, friends, coworkers, etc).

On the one hand, the ATUS provides us with data on the amount of time that people spend in presence of their family members, by age. On the other hand, the ATUS gives us detailed information on how people schedule their activities throughout the day, according to their age. As a matter of fact, each respondent records the minute during which he began the activity and the minute during which it ended it. In addition, the respondent provides information on where the activity took place, and whether acquaintances or unknown people where present during the activity.

In order to complement the ATUS with data on children's time use, we make use of the 1989-90 Activity Pattern Survey of California Children and the 1987-88 Activity Pattern Survey of Californians. Both these time-diary surveys were collected by the Survey Research Center at the University of

[1]The survey does not include active military personnel and people residing in institutions such as nursing homes and prisons.

California, Berkeley, for the Air Resources Board. The surveys were designed to measure the activities and locations of Californians, on a typical day, and to assess their exposure to sources of air pollution. The two surveys cover, respectively, children aged 11 or younger and people who are 12 or older.

Time of exposure contact matrices

In this section we set up a methodology to estimate the daily amount of time that people spend together, on average, according to their age.

Several time use surveys ask the respondent to record some information about the presence of other people, for each of the activities the respondent did during the day. This is useful information, since it allows us to compute, for each activity, the average time spent by the respondents alone or in presence of somebody else. In particular, several time use surveys provides detailed information about the people who were present during the activity, especially if members of the respondent's family. The age of the people who were present during the activity is one piece of information that is particularly relevant to our purposes. Time use data sets usually allows us to directly extract this information for the members of the respondent's family, so that we can easily have a representation of the contact patterns between members of the same families, by age. We represent these contact patterns by means of a matrix whose entry ij is the average time that people in the age group i spend with their family members in the age group j throughout an average day.

In addition to family members, several other people may be present during the activity considered, but the available data do not allow us to identify them. Consider, for instance, activities like being on public transportation, at school, or in a pub: in these cases, people have contacts - in the sense of sharing the same room or having conversations - with a number of people. In order to estimate the amount of time that people spend with non-family members, according to their age and during such activities, we need to develop an indirect approach.

Our approach is based on the assumption that, for those activities such as being on a bus or at school, people ultimately divide their time among other participants to the activity in a proportional way. A matrix that represents the time of exposure to people by age can thus be obtained by aggregating matrices computed under the hypothesis of 'proportional allocation of time'. The time of exposure contact matrix, that we obtain by aggregating those matrices, diverges from the one obtained under the assumption of homogenous mixing by a factor that is related to the process of aggregation.

We divide the whole day into 1440 time slots, each of which consists of one minute. We have that the number of people belonging to the age group i that are in the location h during the time slot z is equal to the number of minutes spent by the population in the age group i, in the location h and during that particular time slot z of the day considered. We refer to this quantity as T_i^{hz} and we interpret it as a measure of person-minutes. For example, let n be the number of age groups in the population, then we have that $\sum_{k=1}^{n} T_k^{hz}$ is the total amount of time spent by the population in the location h during the time slot z. Conversely, $\sum_{z=1}^{1440} T_i^{hz}$ represents the total amount of time spent by people in the age group i in the location h throughout the entire day.

We assume that, for each location and each time slot, the hypothesis of proportional mixing holds. This means that the time of exposure of people in the age group i to people in the age group j, in the location h and for the time slot z, is equal to:

$$\frac{T_i^{hz} \times T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}} \tag{1}$$

The total time of exposure of people in the age group i to people in the age group j, for all locations and throughout an entire day, is obtained by aggregating as follows:

$$\sum_{h=1}^{q} \sum_{z=1}^{1440} \left(\frac{T_i^{hz} \times T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}}\right) \tag{2}$$

where q is the number of activities or locations considered. Later on we will refer to 'activities' and 'locations' as exchangeable terms, since the disaggregation by activities or locations is conceptually equivalent and it is related only to the availability of data.

This method is based on the assumption of random mixing within each time slot and activity, but the process of aggregation over time slots and activities allows for a deviation from random mixing, due to the fact that people heterogeneously structure their day according to their age. Students, for instance, schedule their daily activities according to their scholastic engagements. As a consequence, they share the same locations for several hours a day: they use the public transportation system and they share classrooms mainly during the same time slots. Our representation captures the assortativeness by age that is related to how people schedule their activities throughout the day. In particular, by using this method we account for the fact the people of similar ages tend to do the same activities and tend to schedule those activities during the same time slots.

By summing the value obtained in the expression 2 and the respective one for the time of exposure between family members, we get an estimate of the overall daily time of exposure between people of the age groups i and j. When we extend this procedure to all age groups considered, we get a matrix whose entries represent estimates of the average daily time of exposure between age classes.

For several time use surveys, data on the exact time of beginning and ending of activities are not available. In some cases, only aggregated values by activity or location are provided. This means that, for those time use surveys, we do not know the schedule of activities for each respondent: we only know the total time spent by the respondent in each location throughout a day. We can thus compute the total time spent by the respondents in the age group i in the location h throughout a day, that is equivalent to $\sum_{z=1}^{1440} T_i^{hz}$.

In this case we can still compute the overall time of exposure between people of different age groups, during activities such as being on public transportation or at school, but we are not able to capture the fact that people of the same age groups tend to schedule their activities during the same time slots. We only capture the fact that people of the same age groups tend to do the same activities.

For the q activities considered, the total time of exposure between people in the age group i and people in the age group j, for all locations and throughout an entire day, may be represented as follows:

$$\sum_{h=1}^{q}\left(\frac{\sum_{z=1}^{1440} T_i^{hz} \times \sum_{z=1}^{1440} T_j^{hz}}{\sum_{k=1}^{n}\sum_{z=1}^{1440} T_k^{hz}}\right) \tag{3}$$

Mixing matrices within a time use framework

In the previous section we set up a methodology to estimate time of exposure contact matrices. In this section we discuss mixing patterns within a time use framework. First, we show a formal representation of mixing matrices based on time use data. Then we discuss some insights on the structure of these matrices and the determinants of their deviation from the homogeneous case.

Mixing matrices by age give the fraction of contacts that people in the age group i have with people in the age group j, per unit of time. A time use version of classical mixing matrices is a matrix whose ij entry is the ratio of the ij entry of the respective time of exposure contact matrix and the sum of its entries in the row i.

Mixing matrices for contacts with family members can be easily obtained from time use data: as we introduced it in the previous section, time use surveys generally provide data about the time of exposure to family members. It is thus straightforward to compute time of exposure contact matrices and mixing matrices for contacts between family members.

In what follows we focus on mixing matrices for those social activities, like being at school, at work or on public transportation, for which we do not have direct measures of age-dependent time of exposure between individuals.

The simplest assumption that we can make about mixing patterns is the hypothesis of random mixing. This means that we assume that the probability of having contacts with people of other age groups is dependent only on the size of the age groups. Therefore, the fraction of contacts of an individual in the age group i with people in the age group j, p_{ij}, is dependent only on the age structure of the population. We refer to this case as homogeneous mixing and we express the ij entry of the corresponding mixing matrix as:

$$p_{ij} = \frac{P_j}{P} = f_j \tag{4}$$

where P_j is the size of the age group j and P is the size of the entire population.

Homogeneous mixing has been used extensively in epidemiological models, although it is far from a realistic representation of mixing patterns. For instance, the size of the age group may not prove to be a good indicator of participation to social activities. A first improvement of the homogeneous mixing scheme would thus consider the amount of time that people of different age groups spend doing social activities. If we assume that the fraction of contacts that an individual in the age group i has with people in the age group j is proportional to the time spent on social activities by the people in the age group j, then we have:

$$p_{ij} = \frac{T_j}{\sum_{k=1}^{n} T_k} \tag{5}$$

where T_j is the daily amount of time that people in the age group j spend doing social activities.

We refer to this mixing scheme as proportional mixing: our approach is based on the idea that ultimately people divide their time proportionally to other people. What it is relevant is the level at which the assumption of proportionality is reasonable.

First, we can refine the proportional mixing scheme by assuming proportionality only within each activity/location. A further improvement would assume proportionality at the level of both single locations and single time slots.

Consider the activity/location h: The fraction of contacts that an individual in the age group i has with people in the age group j, while being in the location h is

$$p_{ij}^h = \frac{T_j^h}{\sum_{k=1}^n T_k^h} = \frac{\overline{T}_j^h \times \overline{N}_j^h}{\sum_{k=1}^n \overline{T}_k^h \times \overline{N}_k^h} \tag{6}$$

where $\overline{T}_j^h$ is the average daily time that people in the age group j spend in the location h; $\overline{N}_j^h$ is the average daily number of people in the age group j that participate to the activity/location h. We refer to this mixing scheme as proportional mixing by activity/location.

In order to grasp the determinants of the deviation of the proportional mixing by activity/location from the homogeneous mixing scheme, we can compare the two models. Consider, for instance, the ratio of their respective entries in the mixing matrices:

$$\begin{aligned}\frac{p_{ij}^h}{f_j} &= \frac{\overline{T}_j^h \times \overline{N}_j^h}{\sum_{k=1}^n \overline{T}_k^h \times \overline{N}_k^h} \times \frac{P}{P_j} = \\ &= \frac{\overline{N}_j^h / \overline{N}^h}{P_j / P} \times \frac{\overline{T}_j^h \times \overline{N}^h}{\sum_{k=1}^n \overline{T}_k^h \times \overline{N}_k^h} = \\ &= \frac{\pi_j^h}{f_j} \times \frac{\overline{T}_j^h}{\frac{\sum_{k=1}^n \overline{T}_k^h \times \overline{N}_k^h}{\overline{N}^h}} = \frac{\pi_j^h}{f_j} \times \frac{\overline{T}_j^h}{\overline{T}^h}\end{aligned} \tag{7}$$

where $\pi_j^h = \overline{N}_j^h / \overline{N}^h$.

This decomposition shows that the ratio p_{ij}^h / f_j may be factorized into two terms: the first one gives the relative participation of individuals in the age group j to the activity/location h, compared to the share of the population that they represent. The second factor evaluates, for those who participate to the activity/location h, the average time of participation of those in the age group j, compared to the average time of participation of all participants.

When we aggregate the whole set of activities/locations, we get that the fraction of contacts of an individual in the age group i with people in the age group j is:

$$p_{ij} = \frac{\sum_{h=1}^{q}(\frac{T_i^h \times T_j^h}{\sum_{k=1}^{n} T_k^h})}{\sum_{h=1}^{q} T_i^h} =$$

$$= \sum_{h=1}^{q}(\frac{T_i^h}{\sum_{h=1}^{q} T_i^h} \times \frac{T_j^h}{\sum_{k=1}^{n} T_k^h}) \tag{8}$$

where q is the number of activities/locations considered.

It turns out that p_{ij} is a weighted average of the p_{ij}^h, for each of the q activities/locations considered. The weights are the relative participation of individuals in the age group i to the single activity/location, compared to their participation to all activities/locations.

Our methodology can be further refined by assuming that the hypothesis of proportionality holds only for people who are both involved in the same activity/location and during the same time slots. If we divide the 24-hours day into 1440 time slots, then we will have to aggregate $(1440 \times q)$ mixing matrices in order to get the overall proportional mixing matrix by activity/location and time slot.

By applying the same procedure that we used for equation 8, in order to aggregate over activities and time slots, we get:

$$p_{ij} = \frac{\sum_{h=1}^{q}\sum_{z=1}^{1440}(\frac{T_i^{hz} \times T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}})}{\sum_{h=1}^{q}\sum_{z=1}^{1440} T_i^{hz}} =$$

$$= \sum_{h=1}^{q}\sum_{z=1}^{1440}(\frac{T_i^{hz}}{\sum_{h=1}^{q}\sum_{z=1}^{1440} T_i^{hz}} \times \frac{T_j^{hz}}{\sum_{k=1}^{n} T_k^{hz}}) \tag{9}$$

This means that the overall mixing matrix is a weighted average of the proportional mixing matrices built at the level of single activity/location and single time slot. The weights represent the relative importance of the activity/location and the time slot considered, compared to all activities/locations and time slots.

When we compare the mixing matrix that we obtained by disaggregating over both activities/locations and time slots, with the homogeneous mixing one, we obtain:

$$\frac{p_{ij}}{f_j} = \sum_{h=1}^{q} \sum_{z=1}^{1440} \left(\frac{\pi_j^{hz}}{f_j} \times \frac{\overline{T}_j^{hz}}{\overline{T}^{hz}} \times \frac{T_i^{hz}}{\sum_{h=1}^{q} \sum_{z=1}^{1440} T_i^{hz}} \right) \tag{10}$$

Three elements basically concur to make this aggregate mixing matrix deviate from the homogeneous mixing one. The first one is the relative participation to the activity/location during single time slots, compared to the age structure. The second element is the average time that individuals in the age group j spend participating to the activity/location, compared to all participants. Finally, the amount of time that people in the age group i spend on single activity/locations and time slots, relative to the amount of time that they spend on all social activities.

RESULTS

In the previous sections we proposed a methodology to analyze time use data in order to get insights on contact patterns by age. In this section we present some results with regard to the U.S.

First, ATUS data provides us with information on the time of exposure to family members, by age. Figure 1 shows a graphical representation of the time of exposure contact matrix between family members, by age: the contact pattern is strongly related to the kinship structure of the U.S. People tend to spend a relevant portion of their time with siblings and parents, especially at young ages. In addition, the youngest age groups show a peak in time of exposure to old people, namely their grandparents. Peaks in correspondence of people with the same age, or with one generation of difference, are noticeable for all age groups.

We built a time of exposure contact matrix by summing up the time of exposure to family members and the time of exposure to people, by age, in the following locations: workplace, restaurant/bar, school, grocery store, mall, bus, subway/train, place of worship. We observed a pronounced heterogeneity in contact patterns by age and we explained it in terms of both heterogeneous participation to activities and heterogeneous scheduling of the same activities throughout the day. Figures 2, 3 and 4 show a graphical representation of time of exposure contact matrices for activities that took place in a restaurant or bar, at workplace or at school, respectively. The ij entry of these matrices is the total daily time that U.S. people in the age group i and U.S. people in the age group j spend on average together in the location considered. For each location, the time of exposure contact matrix is built by summing up 1440 matrices, each of which is representative for a one-minute time slot. It is apparent that contact patterns are structured by activity/location: people of different age tend to spend their time in different places. For example, adults spend a relevant share of their time at work, while adolescents spend, on average, more time at school. The same way, young adults spend more time at restaurants or bars than adolescents or elderly. Social norms make people assume different roles in a society according to their age: duties, preferences, financial situation and habits are factors that are strongly dependent on age and that concur in making people of different age do different activities. Within a time use framework, heterogeneous mixing by age is thus partly explained by the fact that people allocate their time to different activities according to their age.

Different ways of scheduling activities throughout the day introduce a second element of heterogeneity by age. Consider, for instance, the location 'school': among those who spend some time at school during the day, it is likely that adolescents are in classrooms in the morning and in the early afternoon, while those adults who go to school are more likely to spend time in classrooms during the late afternoon or in the evening.

Both different allocations of time to activities and their different scheduling throughout the day are elements of assortativeness by age: people tend to spend more time with those of the same age. Assortativeness translates into time of exposure contact matrices whose elements on the main diagonal are bigger than the other ones and bigger than the elements on the main diagonal for the homogeneous mixing case. The same way, time of exposure contact matrices computed by disaggregating over activities and time slots have bigger values on the main diagonal than the respective matrices computed by disaggregating only over activities.

We evaluated the amount of additional information that time of exposure contact matrices provide, with respect, for instance, to the homogeneous case, by measuring the distance between the respective mixing matrices. We considered mixing matrices, instead of time of exposure contact matrices, in order to avoid scale effects (by construction, the rows of mixing matrices sum to 1). We measured the distance between two matrices by using the Frobenius norm of their difference: let B and C be two $n \times n$ mixing matrices and $A = (B - C)$, then we measure the distance between B and C as

$$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{trace(A^T A)} = \sqrt{\sum_{i=1}^{n} \lambda_i^2} \tag{11}$$

where the λ_is are the singular values of A.

Tables 1, 2 and 3 provide the overall mixing matrices, with regard to the U.S., for the cases of homogeneous mixing, proportional mixing by activities/locations and proportional mixing by activities/locations and time slots, respectively. Figure 5 shows a graphical representation of these mixing matrices that have been computed by applying to the ATUS data the methods that we presented in the previous section. It turns out that the matrix distance between homogeneous mixing and mixing by activities is 0.8401, whereas the distance between homogeneous mixing and mixing by activities and time slots is 0.8535. The distance between mixing by activities and mixing by activities and time slots is 0.0248.

Heterogeneous participation to activities/locations explains most of the deviation from the homogeneous mixing case. By including heterogeneous scheduling of activities into the mixing by activities/locations scheme, the deviation from the homogeneous case slightly increases: this is mainly due to the addition of a further element of assortativeness.

So far we have applied our methodology to the ATUS data set: first we got estimates of time of exposure contact matrices and mixing matrices; then we measured the deviation between different mixing matrices, that we discussed from a theoretical viewpoint in the previous sections. We thus presented a methodology and a rigorous application to a data set. However, the ATUS do not cover people aged 15 or younger: we thus tried to extract some information from other surveys in order to get a picture of contact pattern for children as well. We analyzed the 1989-90 Activity Pattern Survey of California Children and 1987-88 Activity Pattern Survey of Californians: we grouped the activities/locations in order to make them coherent with the ATUS classification and we computed the overall age-related time of exposure contact matrix by aggregating the time of exposure contact matrices for single activities/locations. Figure 6 shows a graphical representation of our estimates for the average daily time of exposure between people, by age groups, that we obtained by combining the different data sources.

DISCUSSION

Our study is the first, as far as the authors are aware, that provide estimates of mixing matrices based on time use surveys. The measurement of human contact patterns has been pursued, in the literature, by applying both indirect and direct methods. The indirect approach relies on epidemiological data, such as estimates of the force of infection for each of the age groups considered. The direct approach is based on the collection of diaries where the respondents record the number and age of people they had a conversation with during the day (See Wallinga et al. (1999) for a review). In both cases the relevant epidemiological variable is the number of contacts by age. Our methodology, on the other hand, relies on the assumption that the time of exposure to other people, by age, is a relevant epidemiological variable for modeling the spread of airborne infectious diseases.

The methodology we proposed allowed us to estimate time of exposure contact matrices and mixing matrices. These matrices give us insights on human contact patterns and can be plugged into mathematical models for the diffusion of airborne-spread infectious agents in order to evaluate, for instance, different vaccination target programmes.

We showed that, within a time use framework, a mixing matrix can be obtained by aggregating proportional mixing matrices over activities/locations and time slots. Three elements basically concur to make the ij entry of mixing matrix by activity/location and time slots differ from respective entry in the homogeneous mixing matrix: the first one is the relative participation to the activity/location during single time slots, compared to the age structure. The second element is the relative amount of time that people in the age group j spend participating to the activity/location, compared to all participants. Finally, the amount of time that people in the age group i spend on single activity/locations and time slots, relative to the amount of time that they spend on all social activities.

We observed a strong heterogeneity in contact patterns by age and we explained it in terms of both heterogeneous participation to activities and heterogeneous scheduling of the same activities throughout the day. Within our scheme, assortativeness is explained in terms of age-dependent allocations of time to activities and their scheduling throughout the day.

In addition to a methodology and its rigorous application to the ATUS data set, we evaluated the deviance between mixing matrices under the assumptions of both proportional mixing by activity/location and proportional mixing by activity/location and time slot. We showed that the allocation of time to activities is by far the major source of assortativeness of contacts by age. This has important implications for the application of our methodology to time use surveys that do not provide diary-structured information: our methodology, as a matter of fact, can be applied to those data sets with a relative small loss of information.

References

[1] Anderson, R.M. and May, R.M. (1985), Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes, *J. Hyg., Camb.*, 94, 365-436.

[2] Edmunds, W.J., O'Callaghan, C.J. and Nokes, D.J. (1997), 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. London B Biol. Sci.* 264, 949-957.

[3] Hethcote, H.W. (1989), Three basic epidemiological models. In *Applied Mathematical Ecology*, L. Gross, T.G. Hallam and S.A. Levin (eds.), Springer-Verlag, 119-144.

[4] Horn, R.A. and Johnson C.R., *Matrix Analysis*, Cambridge University Press, 1990.

[5] Jacquez, J.A., Simon, C.P., Koopman, J., Sattenspiel, L. and Perry, T. (1988), Modelling and the analysis of HIV transmission: the effect of contact patterns, *Mathematical biosciences* 92, 119-99.

[6] Nold, A. (1980), Heterogeneity in disease transmission modeling, *Mathematical Biosciences*, 52, 227-240.

[7] Wallinga, J., Edmunds, J.W. and Kretzschmar, M. (1999), Perspective: human contact patterns and the spread of airborne infectious diseases, *Trends in Microbiology*, 9, 372-377.

Figures and Tables

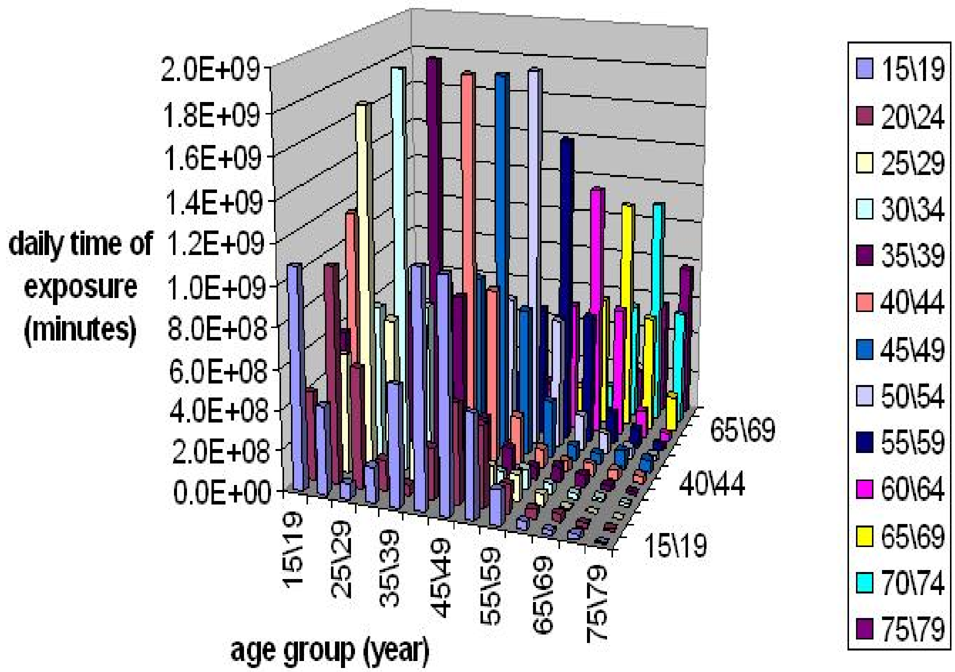

Figure 1: Overall time of exposure between family members and relatives, by age groups, in the U.S. Data source: ATUS 2003.

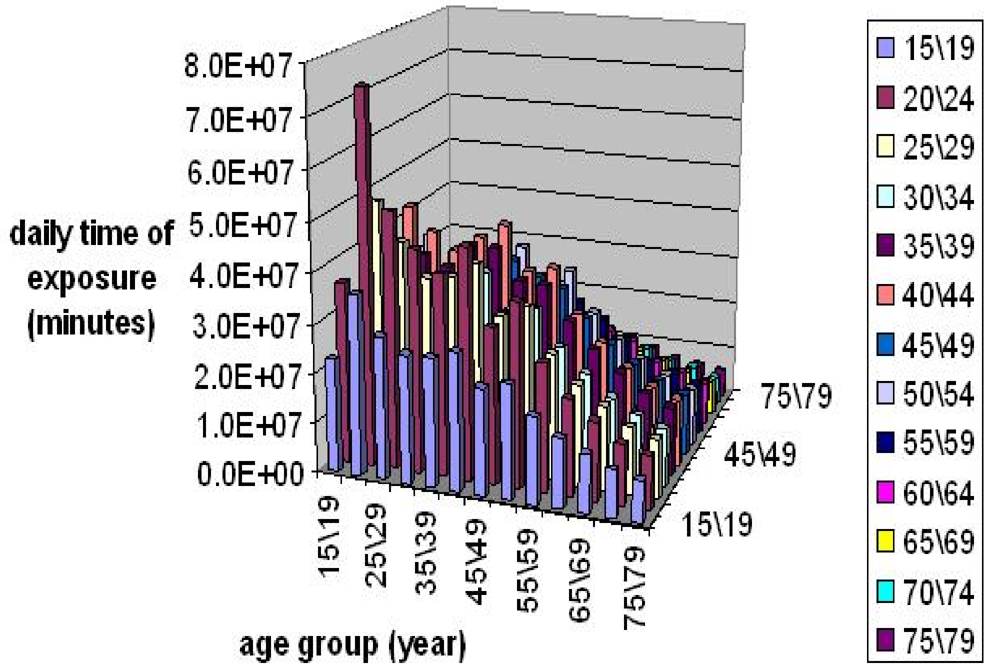

Figure 2: Overall daily time of exposure between people living in the U.S., by age groups, while they are in restaurants or bars. Data source: ATUS 2003.

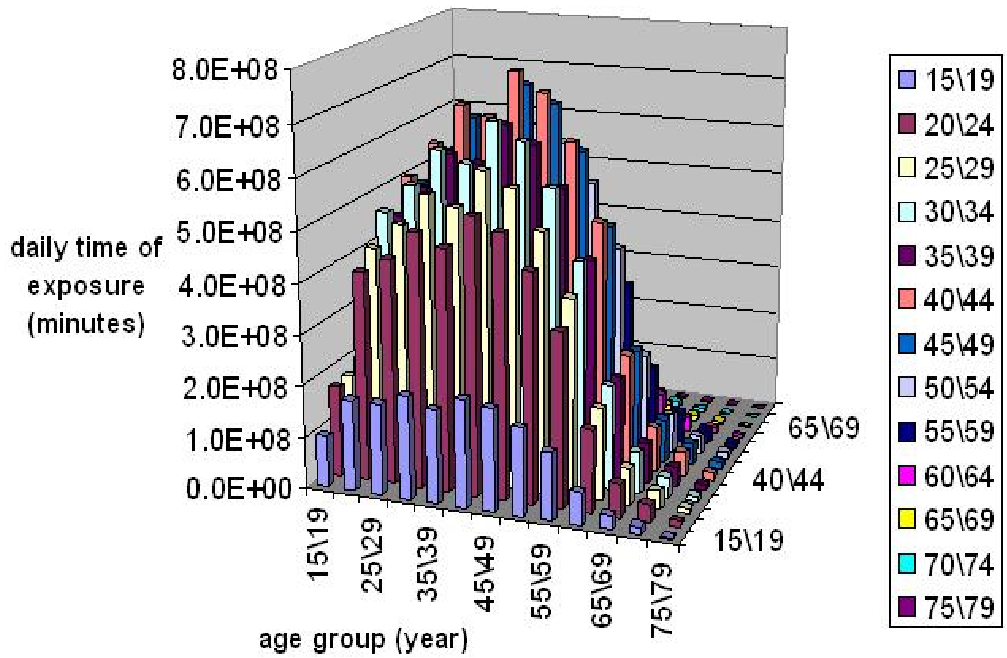

Figure 3: Overall daily time of exposure between people living in the U.S., by age groups, while they are at their workplace. Data source: ATUS 2003.

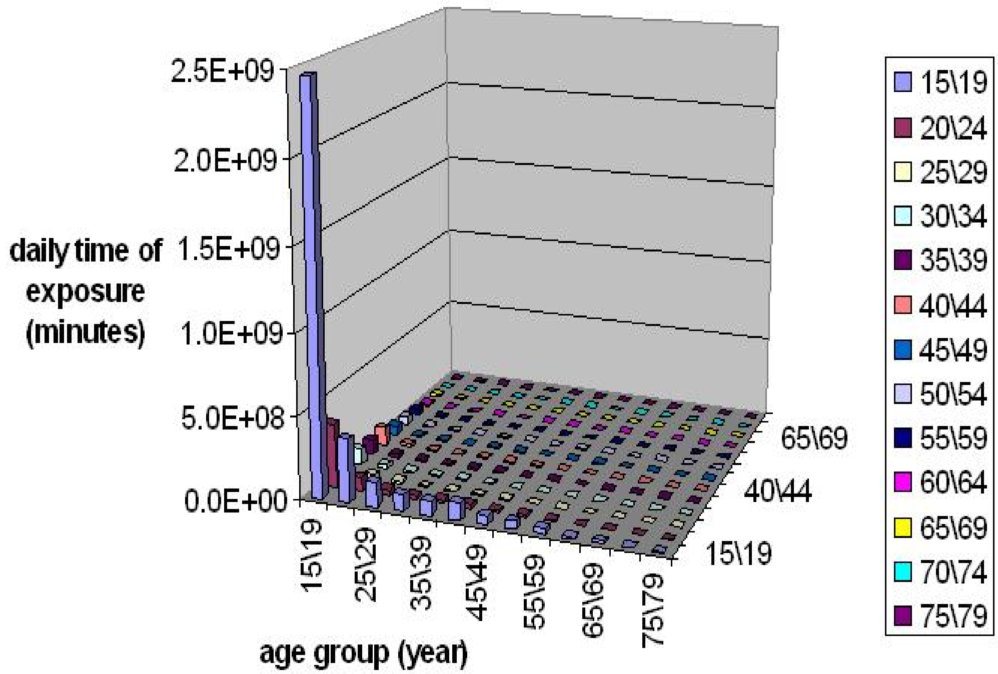

Figure 4: Overall daily time of exposure between people living in the U.S., by age groups, while they are at school. Data source: ATUS 2003.

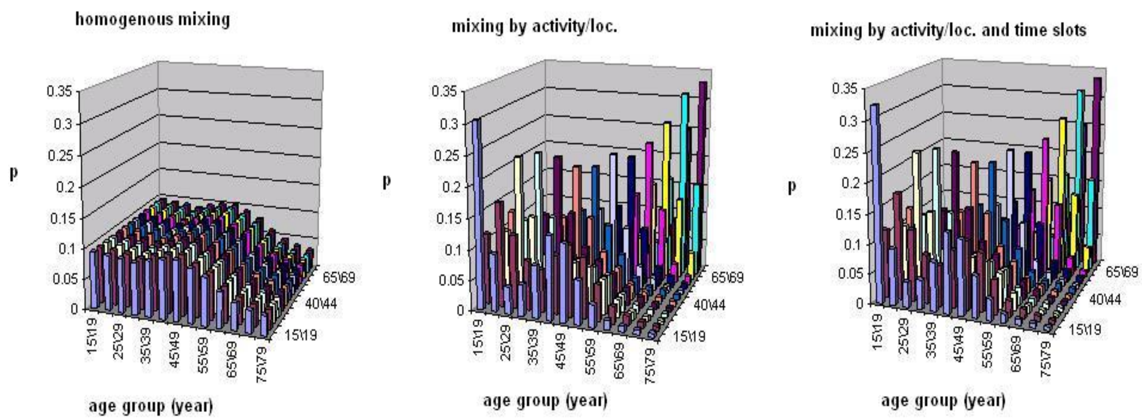

Figure 5: Graphical representation of mixing matrices for people living in the U.S., for the cases of homogeneous mixing, proportional mixing by activities/locations and proportional mixing by activities/locations and time slots. Data source: ATUS 2003.

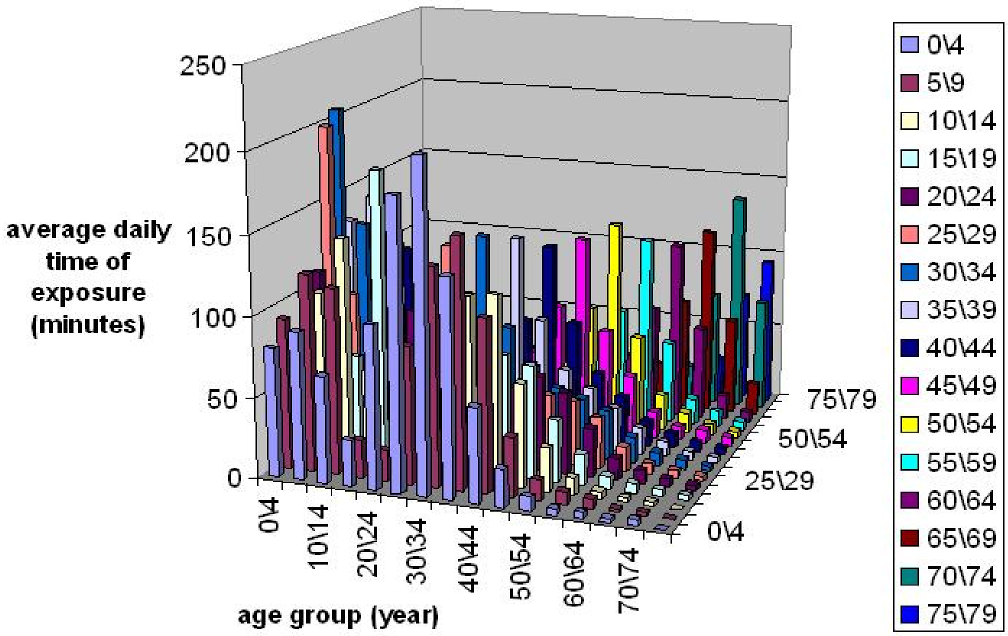

Figure 6: Average daily time of exposure between people living in the U.S., by age groups. Data source: ATUS 2003, 1989-90 Activity Pattern Survey of California Children, 1987-88 Activity Pattern Survey of Californians.

	15\19	20\24	25\29	30\34	35\39	40\44	45\49	50\54	55\59	60\64	65\69	70\74	75\79
15\19	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
20\24	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
25\29	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
30\34	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
35\39	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
40\44	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
45\49	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
50\54	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
55\59	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
60\64	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
65\69	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
70\74	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032
75\79	0.095	0.092	0.090	0.086	0.093	0.099	0.100	0.090	0.080	0.060	0.046	0.038	0.032

Table 1: Mixing matrix by age for people living in the U.S., under the assumption of homogeneous mixing. Data source: ATUS 2003.

	15\19	20\24	25\29	30\34	35\39	40\44	45\49	50\54	55\59	60\64	65\69	70\74	75\79
15\19	0.305	0.096	0.044	0.052	0.087	0.138	0.129	0.073	0.037	0.016	0.010	0.008	0.006
20\24	0.117	0.171	0.121	0.082	0.070	0.098	0.117	0.101	0.060	0.028	0.017	0.011	0.007
25\29	0.053	0.117	0.240	0.146	0.087	0.085	0.080	0.076	0.060	0.030	0.014	0.009	0.005
30\34	0.056	0.071	0.131	0.242	0.140	0.098	0.077	0.067	0.057	0.029	0.015	0.010	0.007
35\39	0.088	0.059	0.075	0.134	0.232	0.142	0.088	0.069	0.050	0.030	0.018	0.010	0.007
40\44	0.124	0.072	0.065	0.083	0.126	0.212	0.129	0.075	0.051	0.027	0.017	0.011	0.008
45\49	0.119	0.089	0.062	0.067	0.080	0.133	0.208	0.111	0.065	0.028	0.016	0.014	0.009
50\54	0.078	0.088	0.069	0.068	0.072	0.089	0.129	0.225	0.102	0.039	0.021	0.012	0.008
55\59	0.051	0.067	0.069	0.072	0.067	0.078	0.095	0.129	0.217	0.096	0.031	0.019	0.010
60\64	0.033	0.048	0.052	0.057	0.061	0.062	0.062	0.075	0.146	0.237	0.124	0.032	0.012
65\69	0.028	0.038	0.031	0.037	0.046	0.049	0.045	0.053	0.061	0.160	0.268	0.138	0.046
70\74	0.025	0.028	0.024	0.031	0.033	0.041	0.050	0.037	0.045	0.050	0.165	0.313	0.159
75\79	0.028	0.028	0.023	0.032	0.032	0.042	0.050	0.036	0.036	0.030	0.086	0.248	0.331

Table 2: Mixing matrix by age for people living in the U.S., under the assumption of proportional mixing by activities/locations. Data source: ATUS 2003.

	15\19	20\24	25\29	30\34	35\39	40\44	45\49	50\54	55\59	60\64	65\69	70\74	75\79
15\19	0.324	0.095	0.043	0.050	0.083	0.135	0.127	0.070	0.035	0.015	0.010	0.008	0.005
20\24	0.116	0.179	0.122	0.083	0.069	0.098	0.116	0.100	0.059	0.027	0.016	0.010	0.006
25\29	0.051	0.118	0.242	0.146	0.087	0.085	0.079	0.076	0.060	0.029	0.013	0.008	0.005
30\34	0.053	0.072	0.131	0.244	0.141	0.098	0.077	0.067	0.057	0.029	0.015	0.010	0.007
35\39	0.085	0.058	0.074	0.134	0.234	0.143	0.088	0.069	0.050	0.030	0.018	0.010	0.007
40\44	0.121	0.072	0.065	0.083	0.126	0.214	0.130	0.075	0.052	0.027	0.016	0.011	0.008
45\49	0.117	0.088	0.062	0.067	0.080	0.133	0.209	0.112	0.065	0.028	0.016	0.014	0.009
50\54	0.075	0.087	0.069	0.068	0.073	0.089	0.129	0.226	0.102	0.039	0.021	0.012	0.008
55\59	0.048	0.065	0.069	0.072	0.067	0.078	0.096	0.130	0.218	0.096	0.031	0.019	0.010
60\64	0.031	0.046	0.051	0.057	0.061	0.062	0.063	0.076	0.146	0.238	0.124	0.032	0.013
65\69	0.027	0.036	0.030	0.037	0.046	0.049	0.045	0.053	0.061	0.161	0.270	0.138	0.047
70\74	0.025	0.026	0.023	0.030	0.032	0.040	0.050	0.037	0.045	0.050	0.166	0.315	0.160
75\79	0.026	0.024	0.021	0.031	0.032	0.041	0.050	0.036	0.037	0.031	0.087	0.249	0.334

Table 3: Mixing matrix by age for people living in the U.S., under the assumption of proportional mixing by activities/locations and time slots. Data source: ATUS 2003.