

## Modelling heterogeneous social contact patterns in a Simulator for the Transmission of Infectious Diseases (Stride)

*Elise Kuylen - PhD student at the University of Antwerp.*

In today's world, infectious diseases can spread quickly - and therefore we should be able to react swiftly and correctly when they manifest. We want to be able to predict both the way in which an epidemic will likely spread, as well as the effect that certain interventions, such as vaccination campaigns, will have.

However, conducting large-scale clinical trials is often infeasible, due to budget, ethical reasons, or time constraints. This is why modelling and simulation play such an important role in modern infectious disease epidemiology. Building mathematical models allows us to compare 'what-if' scenarios that would be impossible to evaluate in a controlled way in real life.

Over the last few years, individual-based models, where each individual in a population is represented by a unique entity, have become increasingly popular. These kind of models are able to take into account the multiple levels heterogeneity that exist within a population in a more direct and intuitive manner than compartmental, population-level models can.

One way in which this heterogeneity manifests is social behaviour. The average number of contacts a person has per day, and with whom those contacts are, varies from person to person. Modelling this is not trivial: social mixing patterns are very complex, and are influenced by many factors. However, they can have an important effect on the spread of an infectious disease, and thus should be included in a model for the transmission of such a disease (1, 2).

Several individual-based models for the transmission of infectious diseases, that take into account heterogeneous social mixing, have already been proposed. However, many of these models focus on very specific diseases or situations (3, 4) and only a few, such as FluTE (5) and FRED (6) are publicly available as open-source projects.

Stride (a Simulator for the TRansmission of Infectious DisEases) is an open-source simulator for

the transmission of infectious diseases. It has been designed to be applicable to multiple situations, populations and diseases: through input files characteristics of the environment, the population and the disease can be specified. Special care has been taken to ensure optimal runtimes, even when working with populations of up to several millions of individuals.

In Stride, the influence of age, context and type of day on social mixing patterns is explicitly modelled (7, 8): a fifteen-year-old individual will have different social behaviour than someone who is in their forties, and that same fifteen-year-old will make different contacts at school than at home.

After briefly introducing our model, we illustrate it by simulating the spread of Influenza in a synthetic population for Miami-Dade (Florida, USA). Besides using age- and context-dependent social contact frequencies, we also compare three social contact hypotheses to examine the influence of the type of day (weekday, weekend day or holiday) on mixing patterns and disease spread.

### *The Stride simulator*

Since Stride is based on an individual-based model, the central entity of the simulator is the Person. Each person is characterized by their age, gender, and health status, which can be one of either susceptible, exposed, infectious, symptomatic, recovered or vaccinated/immunized.

Each person is also a member of four different Clusters. A cluster represents a group of individuals that may contact each other during a simulation day. Every individual belongs to a Household cluster. Individuals younger than 18 are also assigned a School cluster, while those older than 18 are assigned to a Work cluster - with a certain probability, reflecting the unemployment rate in the simulated population. Finally, every person is also a member of two more general Community clusters. These represent other contacts an individual may have during the day, such as

visiting public spaces or meeting with friends.

For each of these different cluster types, a Contact Profile is read from an input file at the beginning of the simulation. This contains the rates at which members of these clusters contact each other, depending on their ages. From these rates, a contact probability can be derived, which is then compared to a randomly drawn number to determine if contact actually occurs between two given members of a cluster.

The simulator moves forward in discrete time steps of one day. Each simulation day, we first update the health status of all individuals in the population. This means that for exposed persons, we check whether they should become infectious and/or symptomatic, and for infectious and/or symptomatic persons, we check whether they should recover. The durations of the incubation period and of the infectious and symptomatic periods are unique for each individual, and are drawn from a distribution at the beginning of the simulation.

Secondly, we decide for each cluster, which individuals are present on the current simulation day. This depends on the type of the cluster (household, school, workplace, or community cluster) and on whether the simulated day is a weekday, weekend day or a holiday. As we will see in the example below, different kinds of regimes can be implemented here.

Next, for each cluster, we simulate the actual contacts and transmissions that take place. To ensure optimal runtimes, we implemented two different algorithms to simulate contacts in a cluster. Which algorithm is used depends on which contacts we are interested in: do we want to study all possible contacts that take place within a cluster, or are we only interested in those contacts that can lead to a transmission of the disease? In the latter case, we will use an optimised algorithm, which only simulates the contacts between infectious and susceptible members of a cluster (9).

In the first case, when all possible contacts between members of a cluster are simulated, it is possible to mark a certain fraction of the population as 'survey participants'. As such, only the contacts made by these 'survey participants' are logged, to limit the

output of the simulator. For example, the resulting contact patterns presented in this article are based on a sample of 3000 survey participants from the total simulated population.

As mentioned before, whether contact occurs between two members of a cluster depends on the type of cluster and the age of the individuals: this probability can be looked up through a contact profile specific to the cluster type. When contact does occur between an infectious and a susceptible individual, the transmission probability determines whether a transmission of the disease also occurs.

More information on the structure and implementation of Stride can be found in a GIT repository on Bitbucket [http://www.bitbucket.org/stride\\_ua/stride](http://www.bitbucket.org/stride_ua/stride).

#### *Influenza in Miami-Dade, USA*

To illustrate the Stride model and elaborate on the importance of social contact modelling, we performed a case study on influenza transmission in Miami-Dade (Florida, USA). Except for using age- and context-related contact probabilities, we also compared three social contact hypotheses. First, we ran the simulator with only weekday contact patterns ("only weekdays" scenario). Then, we ran the simulator with contact rates that were the weighted average of week- and weekend days ("7-day average" scenario). Finally, we used separate contact profiles for weekdays and weekend days ("week/weekend" scenario).

We used a synthetic population of 2.45 million people in 867 251 households, 441 schools and 180 772 workplaces, extracted from the 2010 U.S. Synthetic Population Database (Version 1) of RTI International (10). To construct the aforementioned Community clusters, we aggregated households (which were sorted on ID: in our model proximity in household IDs reflects spatial proximity as well) until a threshold of 2000 people was reached (5). The contact rates we used were partly derived from a survey in Flanders (2), since these data were not available at a population level for the USA.

After someone has been exposed, they will, on average, become infectious after two days, while

they recover, on average, after 6 days. After a person recovers in this model, they will become immune against future influenza infections. The basic reproduction number  $R_0$  - the number of secondary cases an infected individual causes on average - was estimated to be 1.56 for the Miami-Dade population. We repeated a simulation running for 150 simulated days 20 times, each simulation starting with 245 infected persons in the population. Next we examined both the social contact patterns emerging from these simulations, and the number of influenza cases that manifested over time.

In figure 1, age-related mixing patterns observed in the “7 day average” scenario are presented. Yellow represents a high contact frequency, while red represents a lower frequency. For the other two scenarios, similar patterns, which closely mimic those observed in the POLYMOD social contact study (1), were observed (results not shown here). We can see that people are mainly contacting people of the same age, while there are also contacts between children and their (grand)parents - although these are less frequent. Individuals between 18 and 65 can be seen to mix with a broader range of ages, which can be attributed to their contacts in the workplace

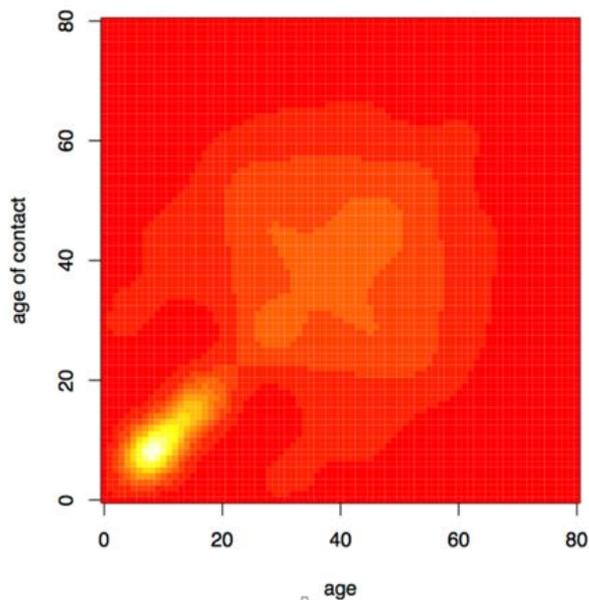


Figure 1. Age-age mixing patterns in Miami-Dade population for the three scenario. (A) “only weekdays”, (B) “7 day average”, (C) “week/weekend”. Yellow indicates high contact frequencies, while red indicates lower contact frequencies.

In figure 2, the number of new influenza cases over time is plotted for the three different scenarios that we tested. We see that, although contact patterns between the three scenarios are very similar, the transmission dynamics are heavily influenced by the timing of those contacts. Not only is the attack rate (fraction of the population that acquired the infection by the end of the outbreak) higher in the “only weekdays” scenario - which may be partially attributed to the higher number of contacts typically occurring on weekdays -, the peak of the epidemic also occurs earlier than in the other two scenarios. Furthermore, the number of daily cases in the “week/weekend” scenario showed an irregular shape, indicating decreased transmission during weekend days.

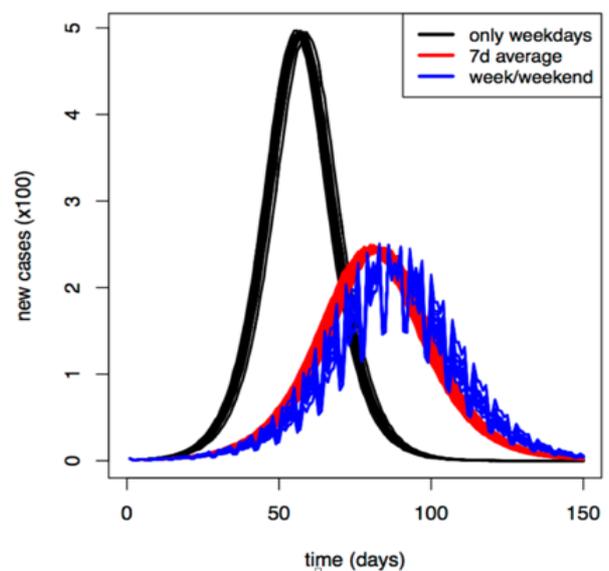


Figure 2. Comparison between three social contact hypotheses of Influenza cases over time for a simulated epidemic in the Miami-Dade population, with  $R_0 = 1.56$ .

#### Future work

Stride explicitly models the influence of age, context and type of day on social contact patterns. However, there are many other factors that influence mixing patterns, which might be included in our model in the future. Examples of these are general policies such as school closures and quarantines, but also fluctuations in people’s social behaviour during an epidemic. We expect future research to focus on including these and other types of adaptive behaviours.

*This article was based on the following paper: Kuylen E, Stijven S, Broeckhove J, Willem L. Social Contact Patterns in an Individual-based Simulator for the Transmission of Infectious Diseases (Stride). Procedia Computer Science. 2017; 108: 2438-2442.*

**Elise Kuylen** - PhD student at the University of Antwerp. She divides her time between the IDLab and the VAXINFECTIO research groups there. Her areas of interest are Modelling & Simulation and Artificial Intelligence. [elise.kuylen@uantwerpen.be](mailto:elise.kuylen@uantwerpen.be)

#### References:

1. 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.
2. Willem L, Van Kerckhove K, Chao DL, Beutels P. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS ONE.* 2012; 7(11): e48695.
3. Liu F, Enanoria WTA, Zipprich J, et al. The role of vaccination coverage, individual behaviors, and the public health response in the control of measles epidemics: an agent-based simulation for California. *BMC Public Health.* 2015; 15(1): 447.
4. Nurhonen M, Cheng AC, Auranen K. Pneumococcal transmission and disease in silico: a microsimulation model of the indirect effects of vaccination. *PLoS ONE.* 2013; 8(2): e56079.
5. Chao DL, Halloran ME, Obenchain VJ, Longini IMJ. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol.* 2010; 6(1): e1000656.
6. Grefenstette JJ, Brown ST, Rosenfeld R, et al. FRED (a framework for reconstructing epidemic dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health.* 2013; 13(1).
7. Ajelli M, Poletti P, Melegaro A, Merler S. The role of different social contexts in shaping influenza transmission during the 2009 pandemic. *Sci Rep.* 2014; 4.
8. Hens N, Goeyvaerts N, Aerts M, Shkedy Z, Van Damme P, Beutels P. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect Dis.* 2009; 9(1): 5.
9. Willem L, Stijven S, Tijskens E, Beutels P, Hens N, Broeckhove J. Optimizing agent-based transmission models for infectious diseases. *BMC Bioinform.* 2015; 16(1): 183.
10. RTI International. 2010 RTI U.S. synthetic population ver. 1.0. 2014. [<http://www.epimodels.org>] Accessed 27 November 2017.