The SILAS-model: Sexual Infections as Large-scale Agent-based Simulation

Stefan Scholz[1], Manuel Batram[2], and Wolfgang Greiner[1]

[1]Dept. of Public Health, Bielefeld University
[2]Dept. of Business Administration & Economics, Bielefeld University

Introduction

Agent-based models (ABM) simulate decision-making and interacting entities as autonomous software agents. Hence, ABM is a method of interest to answer a variety of research questions in the field of health economics, ranging from classical modeling for health technology assessment to assessment of policy impact on social security systems. Independent of the research question at hand, a realistic depiction of all relevant population characteristics in the agent population is the basis for retrieving valid answers. In this context, the economic evaluation of interventions (behavioral or technological) of infectious disease must rely on a sufficiently realistic depiction of the disease dynamic. The spread of infectious diseases not only depends on the course of the disease, but also on the population structure and on behavioural aspects relating on the heterogeneity of patients or patient groups. Therefore, effects caused by certain individuals may not be covered by the models mostly used in epidemiology or public health so far, as only expected values of a population are incorporated in those models (i.e. Markov, dynamic transmission). A famous example to demonstrate the influence of the non-average behavior of a single individual is the spread of HIV in the US via a flight attendant in the 1980's. (see Auerbach et al., 1984) Modeling the spread of HIV using only expected values regarding mobility and number of sex-partners would have failed to include this kind of behavior and therefore the spread of the disease would have been underestimated. Agent-based modeling (ABM) can incorporate the heterogeneity of individuals and the variety of their behaviors to provide more valid and reliable results in epidemiology, public health and health economic evaluation of infectious diseases.

This paper presents first results from the *Sexual Infections as Large-Scale Agent-based Simulation* (SILAS)-model, on the spread of syphilis in Germany. As additional feature to existing literature on ABM in STDs, which reports results for the subgroup of men who have sex with men (MSM)(Gray et al., 2011), the agent population in SILAS is representative for the whole German population.

Methods

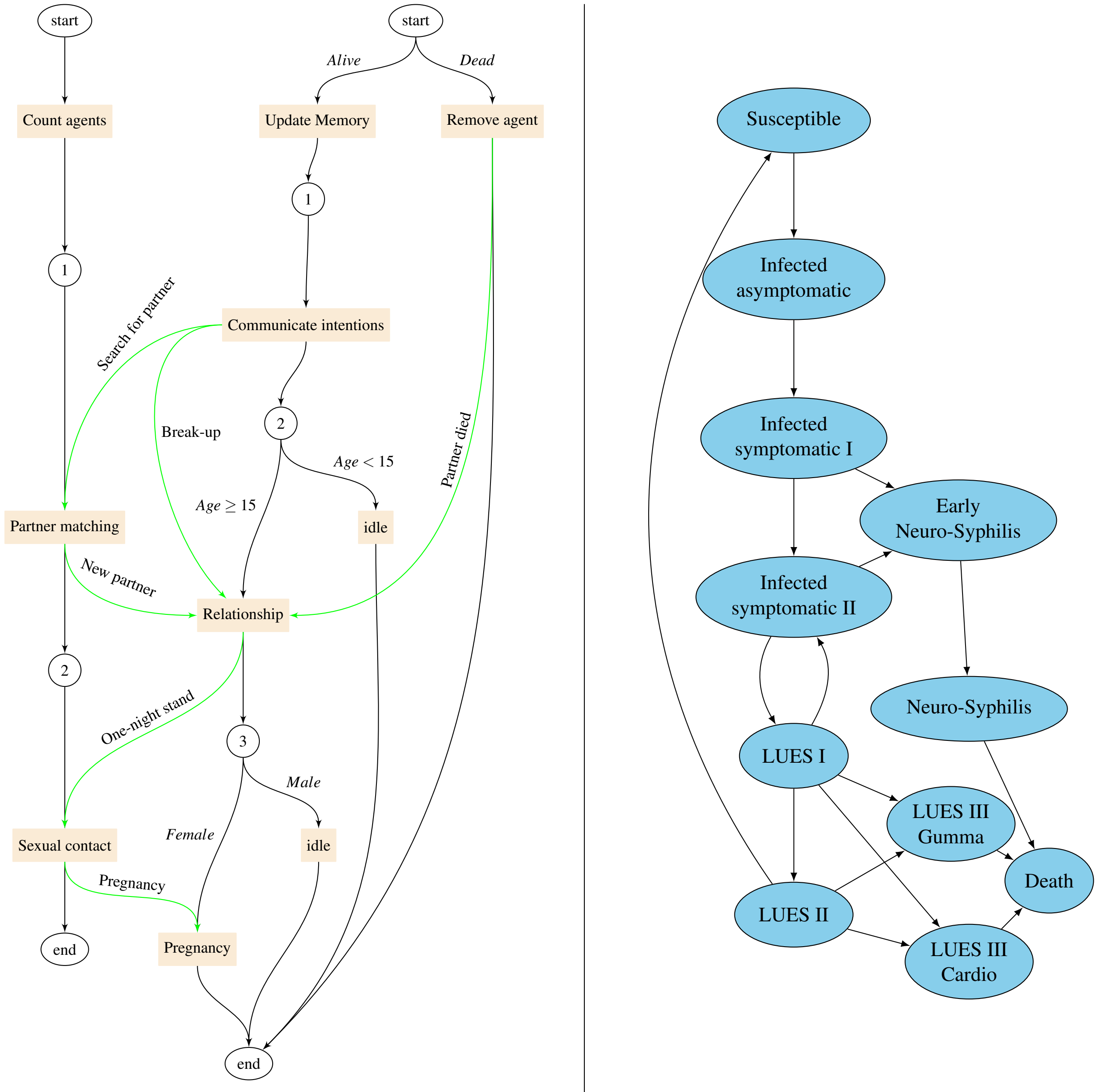
Demographic related behavior

The SILAS-model is implemented in the FLAME-framework (Software Engineering Group, 2014). The software is using the X-machine framework to operationalize agent behavior and interaction via a messageboard system (Coakley/ Smallwood/Holcombe, 2006). Each agent has a set of 21 memory variables (e.g. age, sex, sexual orientation and relationship status) with its individual value sets. Agents in the starting population are intialised with their memory variables being representative for the general German population. In addition, age-, sex- and sexual orientation-specific relationship rates and according age differences between the partners are applied. Within each model iteration (one day) agents run through the model according to the model flow depicted in figure 1a.

In the first period of each iteration, agents update their memory variables independent of interactions with other agents. For example, the day of pregnancy is increased, the age is incremented if the iteration matches the agent's birthday or methods of contraception are chosen. Also, at this point in time, agents determine wether they die on the next day, in which case they get removed from the agent population.

The second period of time serves for intention-building and the communication of intentions related to relationships and sexual contacts. Thereby, the intention for a new relationship is communicated to a helping "stat-agent" and break-up intentions are directly communicated to the partner. The "stat-agent" collects all agents with the intention for finding a new partner, matches these agents according to age and sexual orientation and communicates the result back to the agents in the next iteration. This approach was chosen for computational reasons to avoid redundand readings of the messageboard by all agents. All relationship-related memory variables are updated in the relationship-function.
Sexual intentions are transformed to sexual contacts in relationships without the use of messages assuming symmetric intentions for sexual contact in relationships (for details please see Batram/Scholz, 2014). Sexual intentions of singles are communicated to the "stat-agent" in the same way as the intention for a new relationship. Unfaithful behavior is possible, whereas a non-single agent behaves like a single agent.

In the final period single or unfaithful agents are notified, whether a sexual contact was realized and if it lead to pregnancy considering individual fertility periods and conception probabilities of female agents (de La Rochebrochard/ Thonneau, 2002; Gnoth, 2003; Keulers et al., 2007; Wilcox/Weinberg/Baird, 1995; Barrett/Marshall, 1969; Colombo/ Masarotto, 2000). Again, all related memory variables will be updated. The actual birth (i.e. the creation of a new agent) is realized in the first period under the consideration of age-specific rates of induced and spontaneous abortions. (Nybo Andersen et al., 2000; Federal Statistical Office, 2014)



(a) **General, daily model-flow of the SILAS model:** Circles denote the different states of the agents, squares denote functions, green lines denote messages sent and black lines denote the flow of agents through the model. The flow on the left depicts the statistical agent, the right side the person agents.

(b) **Course of the disease of syphilis within each agent:** Each node represents a Markov-state, arrows denote the direction in which agents in a certain health state can move.

Disease modeling

Agents of the starting population have an age-, sex- and sexual orientation-specific probabilities of syphilis infection. The course of the disease itself is modeled as a Markov-model within each agent as depicted in figure 1b. Once the simulation started, the spread of the disease is only governed by the behavior rules and interactions of the agents. Depending on the health state of the agent, different probabilities of transmission by unprotected sexual contact between agents and age- and sex-specific diagnosis and treatment rates are applied from reported data. (Robert Koch Institute, 2014) Furthermore, congenital transmission at birth is modeled explicitly.

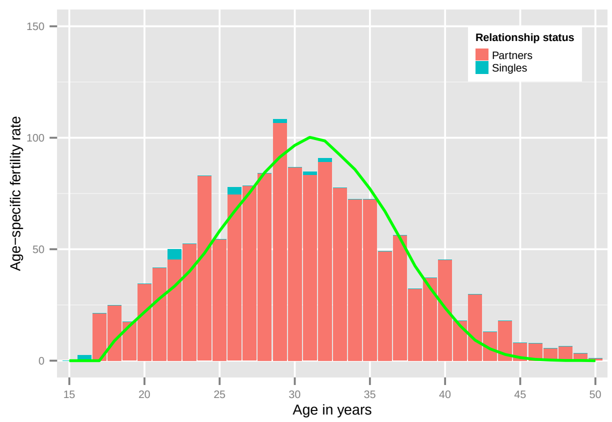
Behavior estimation

The agent behavior (interaction and memory updating) is modeled stochastically and is estimated using data of the German federal statistics office, the Robert Koch-Institute (Robert Koch Institute, 2014), the PAIRFAM panel dataset (Nauck et al., 2014; Huinink et al., 2011), and published literature. Thereby, all distribution-parameters of agent-functions or behavior rules (e.g. mean and variance of a normal distribution) have been estimated in dependence on the agent-memory variables using generalized additive and linear models. For example, all four parameters of the zero-one-inflated-beta distribution, which is used to model all probabilities in the model (e.g. probability of a sexual contact), have been estimated conditional on age, sexual orientation, pregnancy, length of relationship, etc. Agents use the resulting coefficients to calculate their individual, conditional probability distributions and sample their daily behavior from those distributions. A multinomial logit-model was estimated for contraception-use of the agents. Contraception options include no contraception, condoms, birth control pill or both of the latter. All parameters of the according distributions are estimated using the GAMLSS-package (Rigby/Stasinopoulos, 2005) or mlogit-package (Henningsen/Toomet, 2011) for R (R Core Team, 2014).

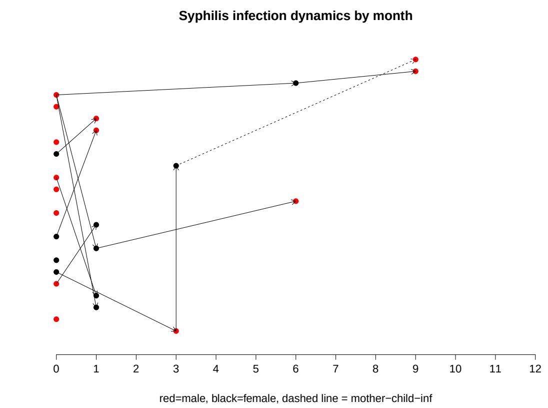
The possibility of computational parallelization in the FLAME-framework as well as in the software tool programmed for the creation of the starting population allows for acceptable run-times for large populations of 1 million agents (~ 8 minutes on a standard-workstation).

Preliminary Results

As can be seen in figure 2a the model sufficiently approximates the observed age-specific fertility rates (ASFR) in Germany. Deviations may be due to the stoachstic nature of the model in combination of the agent population of 100,000 agents, which is used for the step-by-step validation in the current development phase of the model. Hereby, the bottom-up implementation of the demographic behavior rules leads to valid results on the macro-level without any feedback-loop between macro- and micro-level. Infertility, age-specific conception probabilities, induced and spontaneous abortions show to have a relevant effect on the birth rates.



(a) **Demographic results:** The figure depicts the age-specific fertility rate within and outside of relationships produced by the model in the first year of the simulation in contrast to the observed German age-specific fertility rate (green line).



(b) **Disease spread between agents:** Each node represents a Markov-state, arrows denote the direction in which agents in a certain health state can move.

Using the ASFR as indicator for the realistic sexual behavior of the agents, it can be concluded that the frequency of sexual contacts and contraception methods correspond to the German population, thus allowing a realistic simulation of STI spread. Figure 2b shows the infection paths between agents in the SILAS-model for preliminary assumptions about the course of syphilis. In combination with the memory variables of the agents, the role of homo-, hetero and bisexual agents in outbreaks as well as mother-to-child infection can be analyzed.

Concluding Remarks

The SILAS model seems to provide a valid foundation for analyses in the context of epidemiological, public health and health technology assessment in the field of infectious diseases. The agent-based approach provides a tool to identify special patterns of the STI spread thus allowing the evaluation of tailored interventions or the effect of health care utilization (e.g. time to diagnosis). Planned extensions to the model include a regional distribution of agents and the inclusion of sex workers as important parameters to disease spread.

Acknowledgements

This paper uses data from the German Family Panel pairfam, coordinated by Josef Brüderl, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz Neyer, and Sabine Walper. pairfam is funded as long-term project by the German Research Foundation (DFG)

References

Auerbach, David M. et al.: Cluster of Cases of the Acquired Immune Deficiency Syndrome, The American Journal of Medicine, 76 1984 (March 1964), pp. 487–492.

Barrett, John C/Marshall, John: The Risk of Conception on Different Days of the Menstrual Cycle, Population Studies, 23 1969 (3), pp. 455–461.

Batram, Manuel/Scholz, Stefan: Intention action, 2014 ⟨URL: `http://silas-model.com/the-project/special-topics/intention-action/`⟩.

Coakley, Simon/Smallwood, Rod/Holcombe, Mike: Using X-Machines as a Formal Basis for Describing Agents in Agent-Based Modelling, Proceedings of the 2006 Agent-Directed Simulation Conference, 2006, pp. 33–40, ISBN 1565553039.

Colombo, Bernardo/Masarotto, Guido: Daily Fecundability: First Results from a New Data Base, Demographic Research, 3 September 2000 (5) ⟨URL: `http://www.demographic-research.org/volumes/vol3/5/`⟩, ISSN 1435–9871.

de La Rochebrochard, E./Thonneau, P.: Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study, Human Reproduction, 17 June 2002 (6), pp. 1649–1656 ⟨URL: `http://humrep.oxfordjournals.org/lookup/doi/10.1093/humrep/17.6.1649`⟩, ISSN 0268–1161.

Federal Statistical Office: Number of induced abortions 2013, Wiesbaden, Germany, 2014 – Technical report.

Gnoth, C.: Time to pregnancy: results of the German prospective study and impact on the management of infertility, Human Reproduction, 18 September 2003 (9), pp. 1959–1966 ⟨URL: `http://humrep.oxfordjournals.org/lookup/doi/10.1093/humrep/deg366`⟩, ISSN 1460–2350.

Gray, Richard T et al.: Will changes in gay men's sexual behavior reduce syphilis rates? Sexually transmitted diseases, 38 December 2011 (12), pp. 1151–8 ⟨URL: `http://www.ncbi.nlm.nih.gov/pubmed/22082727`⟩, ISSN 1537–4521.

Henningsen, Arne/Toomet, Ott: maxLik: A package for maximum likelihood estimation in R, Computational Statistics, 26 2011 (3), pp. 443–458.

Huinink, Johannes et al.: Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design, Zeitschrift für Familienforschung - Journal of Family Research, 23 2011, pp. 77–101.

Keulers, MJ et al.: The length of the fertile window is associated with the chance of spontaneously conceiving an ongoing pregnancy in subfertile couples, 22 2007 (6), pp. 1652–1656.

Nauck, Bernhard et al.: The German Family Panel (pairfam), Cologne, Germany, 2014.

Nybo Andersen, Anne-marie et al.: Maternal age and fetal loss: population based register linkage study, BMJ General practice, 320 2000 (24 June), pp. 1708–1712.

R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria, 2014 ⟨URL: `http://www.r-project.org`⟩.

Rigby, R A/Stasinopoulos, D M: Generalized additive models for location, scale and shape,(with discussion), Applied Statistics, 54 2005 (3), pp. 507–554.

Robert Koch Institute: SurvStat@RKI 2.0, 2014 ⟨URL: `https://survstat.rki.de`⟩ – visited on 05/12/2014.

Software Engineering Group: FLAME, Sheffield, UK, 2014 ⟨URL: `http://www.flame.ac.uk/`⟩.

Wilcox, Allen J/Weinberg, Clarice R/Baird, Donna D: Timing of Sexual Intercourse in Relation to Ovulation, The New England Journal of Medicine, 333 1995 (23), pp. 1517–1521.