

Transparency and Documentation in Simulations of Infectious Disease Outbreaks: Towards Evidence-Based Public Health Decisions and Communications

Joakim Ekberg¹, Toomas Timpka¹, Magnus Morin³, Johan Jenvald³,
James M. Nyce⁴, Elin A. Gursky⁵, and Henrik Eriksson²

¹Dept. of Medicine and Health Sciences, Linköping University, SE-581 83 Linköping, Sweden
joakim.ekberg@liu.se, tti@ida.liu.se

²Dept. of Computer and Information Science, Linköping University, Linköping, Sweden
her@ida.liu.se

³VSL Research Labs, Linköping, Sweden
magnus.morin@vsl.se, johan.jenvald@vsl.se

⁴Dept. of Anthropology, Ball State University, Muncie, IN, USA
jnyce@rocketmail.com

⁵National Strategies Support Directorate, ANSER/Analytic Services Inc, Arlington, VA, USA
elin.gursky@anser.org

Abstract. Computer simulations have emerged as important tools in the preparation for outbreaks of infectious disease. To support the collaborative planning and responding to the outbreaks, reports from simulations need to be transparent (accessible) with regard to the underlying parametric settings. This paper presents a design for generation of simulation reports where the background settings used in the simulation models are automatically visualized. We extended the ontology-management system Protégé to tag different settings into categories, and included these in report generation in parallel to the simulation outcomes. The report generator takes advantage of an XSLT specification and collects the documentation of the particular simulation settings into abridged XMLs including also summarized results. We conclude that even though inclusion of critical background settings in reports may not increase the accuracy of infectious disease simulations, it can prevent misunderstandings and less than optimal public health decisions.

Keywords: outbreak simulation, ontologies, report generator.

1 Introduction

Recognizing the threat of a destructive pandemic influenza outbreak, the World Health Organization has urged countries to develop preparedness plans [1]. In these preparations, computer simulations have emerged as an important tool for analyzing competing interventions, producing forecasts to be used in exercises, and for support of policy making [2].

Simulations of outbreaks of infectious disease can be powerful tools in local and regional pandemic-response planning [3], but to be most useful for this purpose, there

must be methods for disseminating the simulation results. Without a sufficiently transparent documentation procedure, the advantages simulations offer public health planners become less than desired. In particular, simulation reports must be truthful and clear with regard to the simulation parameters and their values, even when the reports are detached from the immediate simulation context.

The research question this study addresses is how simulation results can be reported responsibly to multiple target audiences, such as when the reports are used as decision support in policy-making contexts. The specific aim is to investigate how the settings used in simulation models can be incorporated in the generation of simulation reports. Previously, we reported an ontology-based approach to outbreak simulation with separate models [4,5] specifically designed to support exploration of alternative interventions under varying hypothetical conditions in local communities [6,7]. This paper describes the method used for results dissemination, in particular a procedure for routine compilation and summary of simulation results.

2 Methods

We extended the existing simulator architecture [5] with a prototypical simulation report generator. The architecture uses the ontology management system Protégé [8] for representing and manipulating simulator settings. The simulator settings are transferred to a simulation engine, designed for computational efficiency, which runs the simulation according to the settings. As a step in an overall assumptions management effort [9], the implemented report generator can access both settings and results to produce a report from the simulation displaying specified subset(s) of the underlying assumptions.

In the development of the report generator, we took into account that many different sets of factors can be chosen to be modelled. There are many factors that theoretically may affect the progress of an actual infectious disease outbreak. The problem is not only to enumerate relevant factors, but also to select significant (i.e., sufficiently important) factors, and to make comparisons possible between different results. In our ontology-based approach to simulation, we separate models for the community, the disease, and interventions. The community model can be made to match factual communities to varying degrees. The disease model includes the epidemiological data available for the specific disease modelled and the intervention model contains different strategies modelled and available and accessible to policy makers.

The Models of Infectious Disease Agent Study (MIDAS) project has documented model profiles, which are standardized descriptions of simulation models devised to aid comparisons between models by detailing modelling, assumptions, data sources and implementation issues [<https://www.epimodels.org/midas/modelProfilesFull.do>]. We separated components from these profiles and used these to develop a typology of explicit simulator settings [9].

The ontology management system of the simulator was augmented by devising a Protégé extension for tagging different categories of settings. The simulator produces documentation from each step in the simulation process in a standardized format using XML (eXtensible Markup Language) documents, thus making a range of trace information available for post-processing and inspection.

For implementing the prototypical report generator, we used the scripting language XSLT (eXtensible Stylesheet Language Transformations), which has been specifically designed for transforming XML documents. The task of compiling simulation results and settings to produce simulation reports in various forms was carried out by a collection of XSLT scripts.

3 Results

The augmented simulation environment keeps track of associations between settings and simulation results. For instance, it is recognized (and shows policy makers) that the reporting of transmission rates is dependent on settings in the simulation model regarding sources for population data, transmission probabilities between different age groups, behavioural settings, such as proportion of individuals staying home from work, and intervention settings.

3.1 Settings and User Groups

For functional reasons, the parametric settings governing simulations of infectious disease transmission were divided into broad categories that reflect the response planning situation. These settings reflect a condition that can be made subject for intervention (is modifiable) and if the setting is based on verified empirical observations or heuristically estimated.

Model settings address the level at which communities, diseases, and interventions need to be represented. The settings include the selected granularity of population data and mixing group structure. For instance, it is often assumed that schools constitute a homogenous mixing group with regard to infectious disease transmission [10]. The community model settings are often based on reports of the daily close contact probabilities for pairs of individuals from different age groups [6]. But the inclusion of too much detail means that simulations not only become more complex, but also can impede validation against empirical data. The model settings can be varied in the simulation environment through substitutable components of community models or disease models. Model assumptions and paradigm assumptions are usually fundamentally associated with the simulation context to be interpreted and analyzed in most policy-making contexts. Comparisons of analyses using different community model settings are also possible but mostly for research and development purposes. Complete reports of settings, with the addition of random seed, software versions, and other technical information may serve as a technical documentation.

Scenario settings include fixed environmental assumptions that normally cannot be controlled by the health response managers, for example the surge capacity of the health care system. Using these scenario assumptions, it is possible to evaluate the effect of an intervention by varying the degree of response and efficiency of the response organizations. Scenario assumptions also include non-controllable epidemiological settings based on probable values on biological parameters derived from the current understanding of the infectious disease in question, such as transmissions rates, the average incubation period, and efficacy of pharmaceuticals. Modifiable

scenario settings, such as the choice and timing of interventions are the most important model components to be adjusted and compared in reports, because most health policy analyses are designed to be able to explore alternative interventions.

Behavioural assumptions address action patterns in populations. These include changes in behaviour due to increased risk awareness during an ongoing pandemic, including voluntary social distancing.

3.2 Implementation

In the extended version of the simulation environment, the user defines a simulation job by specifying the simulation model, instance data, and experiment parameters, and submits it to the computational environment, which is a discrete-event simulator designed for efficiency.

The simulation job generates a number of stochastically-generated populations (usually $n=100$) with randomized initial cases. These are run in the computational environment, simulating each specification 1,000 times. All this documentation is collected into a repository of XML documents.

The report generator uses XSLT specification to collect the documentation for the particular simulation settings, including the settings used for the simulation job, the generated mixing group specification for the simulator, and the simulator results of all collections of simulations into a raw XML report. This intermediate raw XML report is thereafter processed to produce different types of abridged XMLs that summarize results and settings. These report XMLs can easily be manipulated and transformed into a range of formats with built-in features of XSLT as illustrated in Figure 2 (e.g., HTML and PDF). In the current approach, the transformation of settings to textual reports does not take advantage of any natural-language processing, but are reported as stated in the simulator user interface.

In the implementation of the automated report generator in XSLT, the stochastically generated population model documentation was about 20 megabytes in size. With 100 generated populations, this size was too large to be practically manageable. To overcome this, we assumed that the simulation results, including information on how age and mixing group members were infected, were sufficient for documentation purposes. Reports of the variation between different generated populations were deemed necessary only when tracing when and how specific individuals transmitted infection.

3.3 'Swine flu' Use Example

Involvement of outbreak response specialists in provision of model settings and interpretation of simulation reports can be seen in an analysis of high schools closures in preparation of response to the 'swine flu' outbreak in 2009 (Figure 1). The simulation specialist provides (a) community model settings, the health care manager provides (b) non-modifiable scenario settings, the chief epidemiologist provides (c) non-modifiable epidemiological settings, and the policy-maker specifies (d) the intervention. Figure 2 shows the report presenting the analysis results.

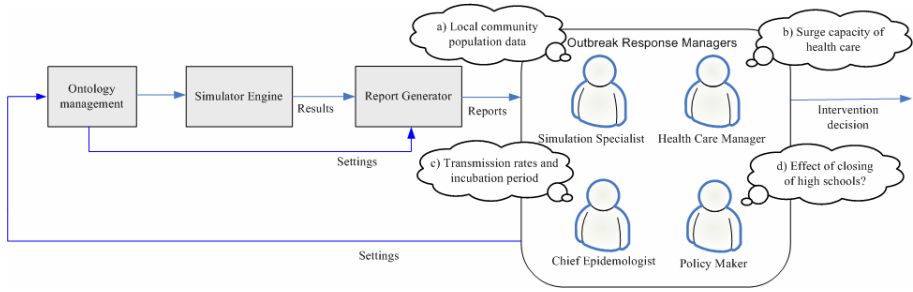


Fig. 1. Overview of the ‘transparent’ predictive modelling process

Simulation report generated from SimulationJob_1/Scenario_1#1 using one SimCore indatda XML and all SimCore outdata XML, generated in approx 4 sec using Saxon6.5.5

Results:

Type of simulation: R0

R0: 1.9965800000000014

Distribution of mean R0

1.42 1.48 1.57 1.60 1.65 1.66 1.68 1.69 1.72 1.72 1.72 1.73 1.75 1.76 1.78 1.78 1.79 1.79 1.80 1.81
 1.81 1.82 1.83 1.84 1.84 1.84 1.84 1.85 1.85 1.85 1.85 1.85 1.87 1.87 1.89 1.89 1.90 1.90 1.91 1.91
 1.92 1.93 1.94 1.94 1.94 1.95 1.95 1.96 1.96 1.96 1.97 1.97 2.00 2.00 2.01 2.02 2.02 2.03 2.03 2.03
 2.03 2.03 2.04 2.04 2.05 2.05 2.06 2.06 2.07 2.08 2.09 2.10 2.12 2.12 2.12 2.13 2.14 2.14 2.16 2.17
 2.17 2.18 2.19 2.19 2.20 2.22 2.25 2.26 2.28 2.33 2.33 2.33 2.33 2.38 2.38 2.44 2.48 2.50 2.59 2.75

Max: 2.757
 Min: 1.427
 Range: 1.33
 Standard diversion between mean R0: 0.2347481427084153
 Variance: 0.05510669050505051
 Number of generated populations (iterations): 100
 Number of simulations for each population (repetitions): 1000
 Number of initial cases: 1
 Simulation mode (R0 or duration): R0

Model Settings

Community Model

Population data source: [Linköping;no workgroup](#)
 Types of Mixing Groups used: [Household](#), [Neighborhood](#), [Community](#), [PlavGroup](#), [Daycare](#), [ElementarySchool](#), [MiddleSchool](#), [Highschool](#)

Epidemiological settings

base-model [H1N1-influenza](#)
 Age Groups: [- 5][6 - 18][19 - 64][65 -]

Types of Mixing Groups used: [Household](#), [Neighborhood](#), [Community](#), [PlavGroup](#), [Daycare](#), [ElementarySchool](#), [MiddleSchool](#), [Highschool](#)

Transmission Rates

Household				
Age Groups	- 5	6 - 18	19 - 64	65 -
- 5	0.08	0.08	0.03	0.03
6 - 18	0.08	0.08	0.03	0.03
19 - 64	0.03	0.03	0.04	0.04
65 -	0.03	0.03	0.04	0.04

Neighborhood				
Age Groups	- 5	6 - 18	19 - 64	65 -
- 5				
6 - 18				
19 - 64				
65 -				

Fig. 2. Example of ‘swine flu’ analysis report with explicit simulator settings

4 Discussion

During an ongoing crisis, such as the ‘swine flu’ outbreak in 2009, health-care policy-makers at a local level need to deal with complex problems. Typically, they need to get a clear picture of what part of the situation (and a simulation) can be affected by their decisions and what is largely out of their control. Simulations can present the dynamics of pandemic outbreaks for policy makers and help them choose between and prioritize decision alternatives.

To facilitate this, heuristic estimates need to be easily recognizable as factual data becomes available and scenarios can be updated. For instance, among avian influenza cases in Indonesia, the mean incubation period appears to have been approximately 5 days, which is nearly twice as long as for past pandemic strains and current inter-pandemic strains of influenza [11]. Redefinition of the non-modifiable epidemiological settings according to new data can dramatically change the simulation outcome. Without proper documentation of what knowledge source has been used, it becomes impossible to determine on what basis predictive modelling is performed. For instance, during the presently (July 2009) ongoing ‘swine flu’, reports on reproduction ratio (average number of secondary cases per primary case) and generation intervals (the time between primary and secondary case infection) [12,13] can easily be compared to previously run simulations when new data emerge.

Similarly, the behavioural assumptions made in pandemic simulations tend to overlook the variety of possible behavioural responses to an epidemic and thereby be flawed [14,15]. The result is that these simulations rest on a more or less simplistic representation of human and social behavioural response. Without proper communication of the behavioural assumptions made, simulations may be interpreted as accurate regarding social dynamics when they are not.

The range of possible explicit settings in reports of an infectious-disease simulation precludes any serious attempt to find a complete or even optimal set of explicit simulator settings. Some more or less well founded settings are explicitly modelled, while others by necessity have to be assumed non-variable or non-relevant. Correspondingly, it is possible to categorize the settings used in predictive modelling of infectious diseases in several ways, and the present categorization is not attempted to be theoretically complete. The visualization of possible settings has to be dynamic and accessible, therefore it has to be organized in a both systematic and manageable way. One of the benefits of using an ontology approach is that it is possible to reorganize and inspect the way the settings are categorized and defined.

5 Conclusions

Simulation can be a powerful resource that can help decision makers prepare for outbreaks of infectious disease. We have described how the background settings used in simulation models can be incorporated in the generation of simulation reports. This transparency does not necessarily increase the accuracy or validity of simulations per se. However, without transparent and accessible documentation procedures, it is difficult for policy makers to interpret the results. To have this functionality built into simulation tools will improve collaborative planning and the kinds of choices health care decision maker make in outbreaks of infectious disease.

Acknowledgements. This work was supported by the Swedish Research Council under contracts 2006-4433 and 2008-5252.

References

1. WHO. Pandemic influenza preparedness and response WHO guidance document Geneva: World Health Organization (2009)
2. Straetmans, M., Buchholz, U., Reiter, S., Haas, W., Krause, G.: Prioritization strategies for pandemic influenza vaccine in 27 countries of the European Union and the Global Health Security Action Group: a review. *BMC Public Health* 7(1), 236 (2007)
3. Jenvald, J., Morin, M., Timpka, T., Eriksson, H.: Simulation as decision support in pandemic influenza preparedness and response. In: ISCRAM, Delft, The Netherlands, May 13-16 (2007)
4. Timpka, T., Morin, M., Jenvald, J., Eriksson, H., Gursky, E.: Towards a simulation environment for modeling of local influenza outbreaks. In: AMIA Annu. Symp. Proc., pp. 729–733 (2005)
5. Eriksson, H., Morin, M., Jenvald, J., Gursky, E., Holm, E., Timpka, T.: Ontology based modeling of pandemic simulation scenarios. *Stud. Health Technol. Inform.* 129, 755–759 (2007)
6. Holm, E., Timpka, T.: A discrete time-space geography for epidemiology: from mixing groups to pockets of local order in pandemic simulations. *Stud. Health Technol. Inform.* 129, 464–468 (2007)
7. Timpka, T., Morin, M., Jenvald, J., Gursky, E., Eriksson, H.: Dealing with ecological fallacy in preparations for influenza pandemics: use of a flexible environment for adaptation of simulations to household structures in local contexts. *Stud. Health Technol. Inform.* 129, 218–222 (2007)
8. Gennari, J.H., Musen, M.A., Ferguson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.: The evolution of Protégé: An environment for knowledge-based systems development. *Int. J. Hum. Comp. Stud.* 58(1), 89–123 (2003)
9. Eriksson, H., Morin, M., Ekberg, J., Jenvald, J., Timpka, T.: Assumptions management in simulation of infectious disease outbreaks. In: AMIA Annual Symp. Proc. (in press, 2009)
10. Germann, T.C., Kadau, K., Longini Jr., I.M., Macken, C.A.: Mitigation strategies for pandemic influenza in the United States. *Proc. Natl. Acad. Sci. U S A* 103(15), 5935–5940 (2006)
11. Yang, Y., Elizabeth Halloran, M., Sugimoto, J.D., Longini Jr., I.M.: Detecting Human-to-Human Transmission of Avian Influenza A (H5N1). *Emerg. Infect. Dis.* 13(9), 1348–1353 (2007)
12. Boelle, P., Bernillon, P., Desenclos, J.: A preliminary estimation of the reproduction ratio for new influenza A (H1N1) from the outbreak in Mexico. *Euro. Surveill* 14(19) (March-April 2009)
13. Fraser, C., Donnelly, C.A., Cauchemez, S., Hanage, W.P., Van Kerkhove, M.D., Hollingsworth, T.D., Griffin, J., Baggaley, R.F., Jenkins, H.E., Lyons, E.J., Jombart, T., Hinsley, W.R., Grassly, N.C., Balloux, F., Ghani, A.C., Ferguson, N.M., Rambaut, A., Pybus, O.G., Lopez-Gatell, H., Apluche-Aranda, C.M., Chapela, I.B., Zavala, E.P., Guevara, D.M., Checchi, F., Garcia, E., Hugonnet, S., Roth, C.: Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science* (2009)
14. Ferguson, N.: Capturing human behavior. *Nature* 446, 733 (2007)
15. Timpka, T., Eriksson, H., Gursky, E., Nyce, J., Morin, M., Jenvald, J., Strömgren, M., Holm, E., Ekberg, J.: Population-based simulations of influenza pandemics: validity and significance for public health policy. *Bull. World Health Organ.* 87, 305–311 (2009)