Micro-simulation of a smallpox outbreak using official register data

L Brouwers (lisa.brouwers@smi.se)1,2, M Boman3,4, M Camitz3,4,5, K Mäkilä1, A Tegnell6
1. Swedish Institute for Infectious Disease Control, Department of Epidemiology, Solna, Sweden
2. Royal Institute of Technology, KTH/ICT/SCS, Kista, Sweden
3. Swedish Institute of Computer Science (SICS), Kista, Sweden
4. Department for Biostatistics and Epidemiology, Karolinska Institute, Solna, Sweden
5. Department of Sociology, Stockholm University, Sweden

Introduction

Should an infection of a contagious disease occur, the potential threat must be met by swift countermeasures. In Sweden, relatively accurate and complete population data as well as environment data are available from governmental institutions. We have used these official register data as part of the input to our computer-based micro-simulation model of the spread of infectious disease. We have studied different pathogens and scenarios, but this report concentrates on our results for smallpox, which is an example of a predominantly airborne, fairly contagious vaccine-preventable disease for which reliable data on some basic parameters exist [1]. To explore the value of a micro-level representation, meaning that we explicitly represent each of the micro-units – here individuals – instead of aggregating them into groups, we performed a number of simulation experiments on the efficacy of various policy interventions for smallpox outbreaks. To the best of our knowledge, our model is the first based on real register data at the level of individuals [2].

To explore the efficacy of four vaccine-based policy strategies (ring vaccination, targeted vaccination, mass vaccination, and pre-vaccination of healthcare personnel combined with ring vaccination) for controlling smallpox outbreaks in Sweden, disease transmission on a spatially explicit social network was simulated. The mixing network was formed from high-coverage official register data of the entire Swedish population, building on the Swedish Total Population Register, the Swedish Employment Register, and the Geographic Database of Sweden. The largest reduction measured in the number of infections was achieved when combining ring vaccination with a pre-vaccination of healthcare personnel. In terms of per dose effectiveness, ring vaccination was by far the most effective strategy. The results can to some extent be adapted to other diseases and environments, including other countries, and the methods used can be analysed in their own right.

A large number of models have been produced to describe the spread of infectious disease, in order to better understand the mechanisms behind incidence and speed, as well as to evaluate countermeasures. In 1905, William Hamer put forth the so-called mass action principle by concluding that an epidemic process is in part governed by the degree of contact between infectious and susceptible individuals. The principle states that the speed of an outbreak’s development is proportional to the product of the number of individuals in these two groups. It is true under the most simple assumption possible concerning the structure of human contacts that everybody is equally likely to meet anybody else, so-called homogeneous mixing [3]. Even today, most models assume homogeneous mixing. The widely used SIR model [4], for instance, contains three groups of individuals: susceptible, infectious, and recovered/removed (SIR). The numbers of individuals in the three groups are functions of time, and the process is often described using partial differential equations [5,6]. Macro-models of this simple kind, in which the behaviours of individuals are not modelled, can be shown to be sufficient for some diseases, such as measles [5,7]. For diseases that are less infectious a close contact between the infectious and the susceptible is required for transmission. Macro-level models assume homogenous mixing, which means that the chance for any two people in the model to meet is equally great – not the case in reality, where geography and contact patterns make it much more likely to meet a family member or a neighbour than a distant stranger. It has recently been established that contact patterns may influence epidemics significantly [8,9]. Real contact networks are highly structured into families and other social groupings, and the rate of contacts varies considerably in the different settings. To identify and to model the key elements in social structures and behaviour are major challenges in disease modelling: levels of detail need to be neither too low nor too high [10]. The computer readily lends itself to random simulations, due to the conceptual ease with which different
assumptions can be implemented, forming different scenarios.

Compared with other models [11,12-14], our model stands out because all individuals – as well as their homes, workplaces, and certain behaviours – are explicitly represented. What is more, the underlying data are real in so far that each individual’s home, family, and workplace is modelled on official register data. In addition, all dwellings and workplaces are spatially explicit – i.e. represented by their real geographical coordinates. This explicit representation allows for exploration of tailored interventions; towards individuals within particular sectors of the work force, geographical regions, or specific age groups.

In a step towards an individual-based model, a pattern of contacts may be devised. This enables us to model the application of control measures such as ring vaccination. Through certain assumptions, it is possible to mimic the effects of contact tracing [15], without explicitly modelling a contact network. The model of Eubank et al. [12] is the most detailed in its population structure and to some lengths mimics a real population by using extremely detailed transportation data. It is through that dataset possible to connect people to places and so generate a contact network. Halloran et al. [16] also use an individual-based approach with a population structured in groups at various levels, such as homes, schools, and clinics. Within each group, contacts take place through homogeneous mixing.

Method
Modelling the population
‘Microsim’ is a structured micro-model for simulating outbreaks of infectious disease [17,18]. It represents the entire Swedish population, with geographically explicit connections to family members, dwellings, and workplaces. Microsim is built to run on a standard personal computer. Updating the status of nine million people in the model is time consuming; thus we put much effort on increasing the speed of execution. A simulation run over 150 days takes about one hour to run, which we found acceptable. The Microsim model

![Figure 1](image1.png)
Age distribution of the Swedish population, 2009


![Figure 2](image2.png)
Household size distribution, Sweden, 1990


![Figure 3](image3.png)
Size distribution of workplaces (number of employees at workplace sites), Sweden, 2009


![Figure 4](image4.png)
Spatial distribution, Swedish population, 2002

Brighter colours indicate higher density.
Modelling the disease
We divided the incubation and symptomatic phases into two parts. We assumed that vaccination is effective only in the first three days of incubation in order to demonstrate the efficacy of different vaccination strategies. Likewise, we assumed that patients are highly infectious for the first four days of symptoms, after which their infectivity decreases. An individual can hence pass through up to six phases with different characteristics in the form of time, health status, infectivity, and more (Table 1). The relevant time-distributions are either uniform or point distributions, with the exception of the second incubating period [11,20-22]. It is assumed that 30% of the unvaccinated individuals will die, with death occurring seven to 14 days into the symptomatic phases.

Micro-modelling individual behaviour
The behaviour of the simulated individuals is defined by a simple set of rules for daily routines or special circumstances. In the morning, each individual checks his/her state of health, which determines activity for the next eight hours. For the working population, or children attending school or day care, this means moving to another location. The probability of making longer journeys within the country is set to 0.03 (3%), based on the average value for daily domestic journeys in excess of 100 km (215,000: probability 0.025, or 2.5%, rounded up to include shorter journeys between regions). Each day, we assume that 5% of the total population are prevented from attending work or school by illnesses other than smallpox. We assume that 1% of the population will seek medical care daily (in 1999 there were 25 million visits to doctors in Sweden [365 days; 9 million inhabitants]) [23]. This amounts to 4,500 emergency room (ER) and hospital visits daily. The daily routine of individuals infected with smallpox and still in incubation is unaffected on the first day of the prodromal phase. On the second day of the prodromal phase, 50% are assumed to be healthy enough to proceed as usual. A further assumption is that 25% will seek medical care and spend the rest of the prodromal phase at home. The remaining quarter will also stay at home from work.

### Table 1
Features of smallpox during its six phases

<table>
<thead>
<tr>
<th>Phase</th>
<th>Time/time distribution</th>
<th>Health status</th>
<th>Infectiousness</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incubating 1</td>
<td>3 days</td>
<td>Healthy</td>
<td>None</td>
<td>Vaccination is effective</td>
</tr>
<tr>
<td>Incubating 2</td>
<td>4–16 days, distribution according to Figure 1</td>
<td>Healthy</td>
<td>None</td>
<td>Vaccination has no effect</td>
</tr>
<tr>
<td>Prodromal</td>
<td>3–5 days, uniform distribution</td>
<td>Influenza-like symptoms increasingly severe</td>
<td>25% during the last 2 days</td>
<td>Patients staying at home/visiting ER during this phase*</td>
</tr>
<tr>
<td>Symptomatic 1</td>
<td>4 days</td>
<td>Pox erupt</td>
<td>Full</td>
<td>Patient admitted to DID</td>
</tr>
<tr>
<td>Symptomatic 2</td>
<td>16 days</td>
<td>Pox dry out</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>Immune/ deceased</td>
<td>Death occurs 7–14 days into this phase in 30% of cases, uniform distribution</td>
<td>Recovered or deceased</td>
<td>None</td>
<td>Patient returns home</td>
</tr>
</tbody>
</table>

* Stage 1: day 1, 100% go to work, none visits ER. Stage 2: from day 2 to the day before the last day, 50% go to work, 50% are ill and stay at home (25% of these visit an ER). Stage 3: last day, those who have not visited an ER before do so now.
home but wait until the last day of the prodromal phase before seeking medical care. On the last day of the prodromal phase, those who continued their daily routines will visit an ER. The events of the symptomatic phase are determined by awareness of an imminent smallpox epidemic. Until three patients have been confirmed with smallpox, it is assumed that an infected individual will wait one day before visiting the ER. The next day, this individual will be transferred to a department of infectious diseases (DID) and stay there for the duration of infection. Once three smallpox patients have been confirmed, all those entering the symptomatic stage will go to the nearest ER immediately.

Modelling places

The places where contacts occur and transmission may take place are of two basic types: regular and random. Regular places are based on register data, and represent contacts for which empirical data are available. The homes and workplaces represented in Microsim are collected from the register data, including geographical coordinates. The resolution of the coordinate system (Cartesian) is 100 metres. Workplaces also include schools, day care centres, and hospitals. Schools and day care centres differ from each other in that the contact rate is higher in day care centres. Workplaces, including schools, are further divided into departments. We assume that a department consists of 25 people, roughly the size of a school class. In comparison, the actual mean (and median) size of a workplace in Sweden is 15, given that one-person companies are excluded from the model (since such workplaces play little or no part in spreading the disease).

Hospitals are special cases in that they will sometimes have an ER or a DID, and sometimes both. These are both represented as departments where people work, as in other workplaces, but also as places where symptomatic smallpox patients are present. Our register data do not connect children to schools and day care centres since the employment register applies only to adults. Instead, we assign children to schools on the basis of proximity and size – i.e. the number of employed adults. Random noise in this assignment accounts for children attending schools other than the one closest to their home.

The second basic type is comprised of places where people meet haphazardly – for example, brief contacts in such places as shopping centres and airports. We used two random place types: neighbourhood and travel. We chose a partition of Sweden into 81 regions, defined by workplace attachment, ‘local workforce region’ [24]. This partition is useful because it means that most travelling, to and from work, is done within these defined regions, and is thus modelled implicitly by the daily movement to and from work. Travel between regions (and to a small extent within the region) is defined as ‘travel’ in our model and is not connected with the workplace or school. We defined one travel destination for each of the 81 regions. This

putative place gathers everyone who travels within or to this region each day. The travel destination mimics the meetings that take place on public transportation such as trains, buses, and aeroplanes. The activity of travelling means one-day journeys in which the traveller is included in the list of travellers for this region, with the possibility of being infected by other travellers in the region as well as infecting them. The destination of a journey is determined on the basis of probability by using a gravitation model based on the number of people in the region and the distance from the dwelling of the traveller. Short trips are more likely than long ones, and trips to a densely populated region are more likely than to one that is sparsely populated. ‘Neighbourhood’ is a proxy for random encounters in the immediate vicinity. These encounters could take place at grocery stores, cinemas, on public transport, or in other places where many people meet. The underlying assumption for this is that it is more likely for a person to meet someone from his or her immediate area than from far away. When transmission has occurred in a dwelling, a neighbourhood list is created and filled with a number of individuals. The probability that an individual will be added to this list decreases with distance. For computational efficiency reasons, the lists are created once only for each neighbourhood and filled with 1,000 individuals from which the contacts are picked at random when transmission is simulated.

Modelling transmission

In Microsim, individuals are assigned to workplaces and to homes in periods of eight and 16 hours, respectively. The risk of infection differs for contacts depending on where they take place, following an assumption about the closeness of contacts and duration of each individual contact. The closeness of contact is assumed to be highest at home, followed by day care centres, schools, and workplaces, in descending order. For ERs, DIDs, as well as for the places of neighbourhood and travel, the ordering of the risk of infection is not as intuitive. The risk was assumed to be quite high in ERs, motivated by the closeness to other people in a crowded waiting room and the long duration of contact when awaiting a doctor. In DIDs, the risk for transmission is much smaller since the risk awareness is high and the staff are likely to take precautions, such as wearing masks. The risks for neighbourhood and travel were obtained by calibration (Table 2) – that is, we tested different values and used the infectiousness values that produced the desired outcome in terms of number of infections from that type of place. In the past smallpox has spread between regions and countries, even though it is known that those with the infection are often very ill. Our way to represent this somehow contradictory behaviour is by setting a low risk for travelling when infectious but a high risk for infection when the infectious person does indeed travel. A highly infectious person might not feel ill when beginning a journey. Such a person might both develop symptoms while travelling, and expose many fellow travellers on a train or bus. Note that the infection risk
for homes is doubled in the simulations, since individuals are assumed to spend 16 hours at home and only eight at work. The risk of infection is much higher within a department than between departments of the same work place. We assume no prior immunity; thus everybody is susceptible to the infection at the outset.

Disease transmission is modelled to occur twice daily. The actual risk of contracting the disease for a susceptible individual during the duration of stay, the infection risk, or $IR$, is calculated for each department and home and is given in Equation 1. Every place type is associated with a basic per contact transmission risk, $pr$. The risks for different place types are listed in Table 2. The basic transmission risk is modified by the current disease phase of the infectious individual in the guise of a phase coefficient $p$ and, if applicable, a coefficient $d$, corresponding to the department to which the infectious individual belongs. The relative infectiousness in prodromal and symptomatic phases is presented in Table 1. Parameter $d$ equals 1 if the susceptible and the infectious individual belong to the same department, otherwise it is lower. The complement of the resulting risks is multiplied over the infectious individuals at that place to produce the complement of the infection risk.

$$1 - IR = \prod_{n=1}^{i} 1 - p_i d_j pr, \text{ for } 1, \ldots, i, j.$$

In Equation 1, above, $i$ designates the infectious individuals and $j$ the susceptible individual under consideration. The assignment of the base risk variables is not trivial. When values exist, they are contradictory or do not lend themselves to implementation models, especially at the micro-level. Initially, we used parameters taken from the model of Halloran et al. [16] as their model was conceptually the closest one to Micr osim. We then calibrated our model, adjusting the parameters to achieve a predetermined goal value of transmissibility, as well as reasonable results in terms of numbers of infected at the different place types. The estimations were based on experiences from previous epidemics, such as the smallpox outbreak in Stockholm in 1963 [25]. The values used for our experiments are found in Table 2. The place type distribution – places where infections took place (the majority in ERs and DIDs) – is shown in Table 3.

We calculated transmissibility using an algorithm [6], essentially starting a simulation with 500 randomly picked initially infected individuals in a totally susceptible population, and counting the number of secondary cases. We iterated 500 times, each time with a new set of 500 infected individuals. Interpreting our transmissibility values in the light of analyses of historical smallpox data, we note that historical data show $R_0$ to have a value of 3.5–6 [26,27]. In Sweden today, where every second household consists of a single person and half the population lives alone or with one other person, the social structure implies that we should end up well below the low end of this interval. Our transmissibility value of 2.25 was hence deemed reasonable.

Modelling vaccination policies

We assumed that the vaccine grants immunity to 80% of those inoculated and we disregarded any adverse effects. The vaccination policies we set out to compare were the following:

- ring vaccination (Ring)
- targeted vaccination of medical care personnel at risk for exposure (Care)
- mass vaccination (Mass)
- pre-vaccination of medical care personnel at risk for exposure + ring vaccination (Combo)

The Combo policy was included because the National Board of Health and Welfare considered the scenario of an outbreak starting in neighbouring countries, with some time permitted to vaccinate medical care personnel in Sweden, as likely and thus of interest. The population in Sweden is nine million. Some 10,500 people work in ERs and DIDs, and are considered to be at high risk of exposure.

Ring vaccination involves tracing the contacts of infectious people as they are identified, including family

---

**Table 2**

The risk of infection with smallpox during a contact, for each place type

<table>
<thead>
<tr>
<th>Place type</th>
<th>Basic infection risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td>0.25</td>
</tr>
<tr>
<td>Day care (within group)</td>
<td>0.1</td>
</tr>
<tr>
<td>School (within class)</td>
<td>0.05</td>
</tr>
<tr>
<td>Work place (within department)</td>
<td>0.05</td>
</tr>
<tr>
<td>Between groups, classes and départements</td>
<td>0.001</td>
</tr>
<tr>
<td>Emergency room</td>
<td>0.2</td>
</tr>
<tr>
<td>Department of infectious diseases</td>
<td>0.01</td>
</tr>
<tr>
<td>Neighbourhood</td>
<td>0.02</td>
</tr>
<tr>
<td>Travel</td>
<td>0.2</td>
</tr>
</tbody>
</table>

* The duration of a contact is eight hours at day and 16 hours at night. The high infection risk at home is a combined result of the close type of contact and the duration. Travel risk includes car, bus, train, and flight travel. Some forms of travel are of long duration in small compartments, hence the relatively high risk assigned.

**Table 3**

Distribution of locations where transmissions of smallpox took place in the vaccination policy experiments

<table>
<thead>
<tr>
<th>Location</th>
<th>Base</th>
<th>Ring</th>
<th>Care</th>
<th>Mass</th>
<th>Combo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwelling</td>
<td>1,953</td>
<td>431</td>
<td>499</td>
<td>142</td>
<td>139</td>
</tr>
<tr>
<td>School/day care</td>
<td>36</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Office</td>
<td>44</td>
<td>7</td>
<td>16</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Travel</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ER/DID</td>
<td>4,700</td>
<td>852</td>
<td>636</td>
<td>429</td>
<td>55</td>
</tr>
</tbody>
</table>

DID: department of infectious diseases; ER: emergency room.
members and colleagues, which are readily available in the model. We assumed that this process is 100% successful and that immunity from vaccination is generated immediately. Both of these assumptions are optimistic, admittedly, and these assumptions should be subjected to sensitivity analysis in longer series of experiments.

In the mass vaccination strategy, we included a capacity at the hospitals and care centres that limited the number of vaccinations to be administered each day. We assumed that a tenth of the nurses could be assigned to the vaccination programme, each able to administer 80 doses a day. This equates to a theoretical maximum of 720,000 patients per day. With the exception of the pre-vaccination part in the Combo policy, which takes place at the start of the simulation, all programmes are launched after the first case has been identified at a DID.

**Experiments**

Baseline values were recorded by running the simulation without intervention. We made 500 runs with different random seeds. Each run had a single individual initially infected, also picked at random. A random seed determines a vector of random numbers that are used throughout the simulation run for all kinds of stochastic events in the model, such as if an individual will travel or not on a particular day. If the same seed is used in several runs, the same random numbers would be generated and the simulation would repeat itself. By using 500 distinct seeds we generated a spectrum of possible scenarios. Each scenario was run for 100 days, which was deemed sufficient for evaluating policies. Longer runs, at the time of these experiments, exceeded the computer’s memory capacity, hence a few outbreaks were not taken into full account because they had not finished by the 100th day. These computational complexity issues have since been fixed, and the model is currently optimised for 300-day runs, even if no more than 100 days are typically required. Of the 500 runs, 41 predicted the infection of 49 or more individuals. These runs were classified as outbreaks and their random seeds were recorded for further use in the policy comparison. A vaccination policy had to reduce the size of these 41 outbreaks to be considered effective.

On average, 172 individuals (family, and colleagues from the same office department) were vaccinated in the Ring vaccination policy. The Care policy vaccinated 10,530 individuals (the same number in each experiment). The efficacy of policies was compared in terms of the difference in numbers of individuals infected. We therefore conducted four further experiments seeded with the same seeds recorded from the 41 outbreaks. We also recorded and compared the per dose reduction in incidence.

**Results**

In order to demonstrate the viability of our microsimulation model, our prototypical experiment set-up yielded the following results.

Figure 5 shows 99 base simulations, illustrating the variation in outbreak magnitude when no interventions were applied (range 0 to 357 infections). In Figure 6 the numbers of infections in each simulation run are shown for the different policies. The runs are sorted from the largest to the smallest number of cases, and the same random seed is used for the four different policy simulations. The results of our intervention experiments are shown in Table 4.

All strategies reduced the numbers of infected from base line values significantly: ring vaccination by 84%,
the care policy by 86%, mass vaccination by 93%, and the combo policy by 97%. The outcomes of the care policy and ring vaccination were not statistically different. Mass vaccination was significantly better than both, but the policy of combining ring vaccination with a pre-vaccination of the care personnel at most risk for exposure (combo) was significantly better than vaccinating the whole population (mass). This assumes that a vaccination of the care personnel is started after the first identified case and that the logistic restrictions described earlier apply.

Further comparisons can be made by examining the vaccination efficacy in terms of the numbers of doses required to prevent one case. It is evident that the combination policy is far more effective than mass vaccination. That the extra doses required to carry out ring vaccination were well spent is indicated by comparing the combined policy to the care policy. In terms of per dose effectiveness, ring vaccination is by far the most effective.

One motivation for vaccinating the highly exposed medical care personnel is the high rate of transmission assumed to occur at ERs and IDUs (where our model includes the personal protective equipment of staff only indirectly). Tables 3 and 4 illustrate how this assumption is represented in our model and explains the success of these strategies in terms of numbers infected and vaccination dosage.

Discussion
An outbreak simulation must take into account not only the numbers of those infected and their mortality but also the costs of vaccine doses and their distribution and the high-risk environment for medical care personnel. Note that we have considered neither adverse effects nor their consequences in our model. We have endeavoured to demonstrate here the general utility of our model and, although the results are subject to an array of assumptions and provisions as far as the parameter values are concerned, these results reflect those of other smallpox simulation studies [11,13,14,16,28].

In order to compare resources required, we calculated the per dose incidence reduction. Ring vaccination is the most effective in this sense. This was expected as only those who have been exposed to the index case are vaccinated. But since vaccine effectiveness is not 100%, and since there is no immunity outside the circle of contacts of the index case, the epidemic is allowed to continue. It is a feature of the model that it allows for interactive testing of different thresholds, that is, for the percentage of contacts that must be found for the policy to be effective.

When running our experiments, we saw that the results are very sensitive to the underlying assumptions. Here, a central variable is transmissibility, and the value of this seemingly simple variable is hard to determine. It is very difficult to assign exact probabilities to risk for disease transmission during a contact, since there are no data on number and nature of contacts. Historical records of outbreaks are of little help, since the reported values are a result of both the agent’s inherent properties as well as external factors, such as the density of the population and factors such as healthcare and social structure. A related complicating factor challenging our assumption about a fully susceptible population is immunologic memory – that is, the possible presence of residual antibodies after vaccination [29,30].

Other important assumptions are those concerning an infected individual’s behaviour, such as going to work or staying at home when ill, or whether an individual will visit an ER or not. To make explicit these assumptions, which are indeed central to the results, we added a graphical interface to the simulation programme. Through this, a user may easily set transmission rates and other variables. Further, a user can select whether a simulation should be run for the whole country or for a certain region only. Also, the number of repetitions and policy interventions may be selected. These features also make our model easier to adapt to the environments presented in other countries, or for use in a limited geographical region, such as a particular city or an island. That said, the availability of register data varies immensely between countries, and only systematic validation of experiment results can determine the utility of a model such as ours for other countries or regions.

Acknowledgements
The project, including all authors except Magnus Boman and Anders Tegnell, was funded by the Swedish Emergency Management Agency (SEMA). All authors except Boman and Tegnell had full access to the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis. Lisa Brouwers, Martin Camitz, and Kalle Mäkilä contributed to the design, the implementation, and the simulation experiments. Mäkilä prepared the datasets and Tegnell designed intervention strategies. Camitz provided the statistical analysis. The authors would like to thank Anette Hulth for helpful comments on earlier drafts.

* Erratum: The title of Table 4 was corrected after publication of the article, on 3 September 2010.

References


