

Simulating Spreading of COVID-19 in Zurich: An Application of EpiSim

Sebastiano Papini

Jingyan Li

ABMT HS2021 Project Report

January 2022

Simulating Spreading of COVID-19 in Zurich: An Application of EpiSim

Sebastiano Papini
Doctorate at D-MTEC
ETH Zürich
spapini@ethz.ch

Jingyan Li
Msc. Geomatics
ETH Zurich
jingyli@ethz.ch

January 2022

Abstract

The World Health Organization (WHO) declared COVID-19 a pandemic on 11 March 2020. There have been 1,833,571 confirmed cases and 12,142 deaths in Switzerland till 21 January 2022. The first wave in Switzerland began in the early March 2020. Since then, public health measures including testing, tracking, social distancing, the use of personal protective equipment and restrictions on activities were conducted to control the transmission of the virus. In this project, we utilized EpiSim, an agent-based model for epidemic simulation, to fit the first wave in Zurich between February 2020 and May 2020. Agent-based models, which captures mobility patterns of individuals, are regarded as a solution to the homogeneity issue in traditional epidemic models. It combines data-driven human mobility modelling with an infection model and a disease progression model to simulate the disease spreading. Interventions like restricted activities and masks can also be considered. We thus employed such a model to simulate the first wave, test different containment measures and the scaling capability of model.

Keywords

Agent-based Modeling, MATSim, EpiSim, Epidemic Models, COVID-19

Contents

List of Tables	1
List of Figures	2
1 Introduction	3
2 Literature Review	4
2.1 Epidemic Models	4
2.2 Agent-based Models for Epidemics	5
2.3 Modeling the Spreading of COVID-19	5
2.4 EpiSim	6
2.4.1 Contact Model	6
2.4.2 Infection Model	7
2.4.3 Disease Progression Model	8
3 Methodology and Model Extensions	8
3.1 Generating Activity Chains in Zurich Scenario	8
3.2 Simulation in EpiSim	9
3.3 Calibration	10
4 Results	11
4.1 Best fits compared to official data	11
4.2 Transmission Analysis	13
4.3 The Impact of Containment Measures	14
4.4 The Scaling Problem	15
5 Discussion	16
5.1 Limitations	16
5.2 Conclusion	18
6 References	19

List of Tables

1 Contact intensity parameters by activity type	9
2 Average Share of Infection Types	14

List of Figures

1	Restrictions from Google community mobility report	10
2	Hospitalisation: Simulation and Actual	12
3	Number of COVID-19 patients in intensive care (ICU): actual population in ICU (red line), different simulations (black line	13
4	Cumulative number of COVID-19 cases: actual cumulative positives (red line), different simulations (black line.)	13
5	Mean cases 7-day moving average: simulated positive cases (blue line), simulated positive cases with symptoms (black line) and positive tests from official data (red line).	13
6	Share of cases detected: the share of positive tests from official data over simulated positive cases (blue line), the share of positive tests from official data over simulated positives with symptoms (black line).	13
7	Transmission per Infected Person over all Simulations	15
8	Average Prediction and Confidence Interval for Counterfactual Confinement Measures	16
9	Average Prediction and Confidence Interval for Different Sample Sizes	17

1 Introduction

On 11 March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. Zurich faced the first wave from March 2020 to the end of April, 2020, during which around 200 people were hospitalized at the peak. The national government declared a series of measurements to slow down the spreading. Since 20 March, most shops, restaurants and schools were closed, and all gatherings of more than five people in public spaces were banned. The measures were gradually removed in several phases beginning in late April until June 2020 with the receding of the first wave.

Modeling the pandemic is one of the keys to understand the spreading of the virus and propose efficient measures to slow down its transmission. Generally, virus spreading can be captured by compartment models. Population are split to different compartments to represent different states of a person interacting with the disease, and a social network is established to capture the physical contact between people. Agent-based models (ABMs), which are designed to capture individual mobility, excels at capturing the complex human behavior with high granularity, thus could model virus transmission between individuals. EpiSim¹, as one of implementations of ABMs for epidemics, was developed based on a large-scale agent-based transport simulations (MATSim²). Müller *et al.* (2020a) and Penazzi and Axhausen (2021) have already shown promising simulations by EpiSim in Berlin and Switzerland since the outbreak of the pandemic.

In the project, we are going to utilize EpiSim to simulate the first wave (i.e., late February 2020 - May 2020) in Zurich and test different containment measures based on our simulation. In chapter 2, we introduced the traditional epidemic models, ABMs for epidemics and related studies to model COVID-19 since its outbreak. We specifically introduced the structure of EpiSim. Then we elaborated how we extended EpiSim to the Zurich scenario in chapter 3. The calibration results were shown in chapter 4. We also discussed the impact of different containment measures and the scaling problem of the model. In the last chapter, we discussed limitations of the project and envisioned the future work.

¹<https://github.com/matsim-org/matsim-episim-libs>

²<https://www.matsim.org/>

2 Literature Review

2.1 Epidemic Models

Epidemics can be modeled to simulate the spreading of disease. The widely used model is so called SIR model and its variants (e.g. SEIR, SI, etc.) (Kermack and McKendrick, 1932). These models splits the population into compartments to represent different states of a person interacting with the disease. In general, SIR models three major states: (1) S = *susceptible*, (S) portion of population is susceptible, (2) I = *infected/infectious* denotes the set of people who are infected and infectious to others, and (3) R = *recover* represents people who recover from the disease and become resistant to it. Notice that the basic SIR model does not consider the incubation period. SEIR model, as an extension, adds an intermediate stage E (E = *exposed*) to describe people who have been exposed to the disease but not yet get infected, making the incubation period possible in a model. Such epidemic models describe the transmissions of disease by nonlinear ordinary differential equations that calculate transition rates between different compartments of population.

The compartment models have the major restriction that they regard population segments in each compartment homogeneous (Frias-Martinez *et al.*, 2011), thus interactions within each compartment are modeled randomly. The heterogeneity of infections is hard to simulate, especially for detecting some 'superspreaders' who have connections with a great number of people. Metapopulation models, as a solution to such a restriction, further split types of population within each epidemic state. People differentiate from each other based on vaccination status, being either symptomatic or not, mobility patterns and natural immunity to diseases (Balcan *et al.*, 2009; Leventhal *et al.*, 2015; Belik *et al.*, 2011). Then the model regards the population as a network, where persons are represented as vertices, and edges denote the connection between persons. The random interactions are modeled by the connectivity of edges. Belik *et al.* (2011) also considered the human movement in the metapopulation network: people have a home base and have several frequently visited places in the model. Although metapopulation models achieve good performance in describing the complexity of human behavior, they cannot deal with the heterogeneity between individuals.

2.2 Agent-based Models for Epidemics

Agent-based models (ABMs) that are designed to capture individual (agent) mobility excel at capturing the complex human behavior with high granularity (Frias-Martinez *et al.*, 2011), making it possible to model movements and interactions individually. It is then a solution to solve the limitations of capturing heterogeneous human behavior in compartment models.

ABMs for epidemics generally start from generating the social networks which model movements and interactions among population. Usually, the network is generated based on synthetic population from census data (Connell *et al.*, 2009; Apolloni *et al.*, 2009; Barrett *et al.*, 2008). Individuals are assigned to a home address, and their daily activities are modeled according to transport surveys. Frias-Martinez *et al.* (2011) also showed the possibility to generate real-time social network based on mobile phone data. They extracted individual activities and interactions between agents from call detail record during the epidemics, and the ABM thus could describe the disease spreading in time. After setting up the social networks, a disease model is introduced to model the disease transmission between agents and the disease progression for an agent (Connell *et al.*, 2009; Apolloni *et al.*, 2009; Barrett *et al.*, 2008; Frias-Martinez *et al.*, 2011). Besides, Apolloni *et al.* (2009) introduced intervention models to integrate the impact of public policy and the change of individual mobility behaviors.

2.3 Modeling the Spreading of COVID-19

Since the outbreak of COVID-19 in 2019, studies have been modeling the spreading of the epidemics and making contributions to public policies. Imperial College (Ferguson *et al.*, 2020) assessed the impact of non-pharmaceutical interventions on the disease spreading in the UK and US based on the simulation of an epidemic model. Their results showed that multiple interventions of public health measurements are required to have a substantial mitigation of transmission, which had a great impact on the policy making in the UK.

Specifically, several ABMs for COVID-19 have been implemented during the pandemics to describe the real-time human behavior and test public health policies. EpiSim (Müller *et al.*, 2020b) is an extended implementation based on MATSim (W Axhausen *et al.*, 2016), an ABM for transportation. Replaying the COVID-19 in Berlin by EpiSim showed the possibility of speeding up implementation of an ABM for epidemics based on an existing ABM. Besides, they tested the intervention of reducing activities and compulsive

masks. REINA (Tuomisto *et al.*, 2020) is another ABM for epidemics proposed to test the destructive policies in Helsinki. In addition to the macro-scale simulation, Cuevas (2020) zoomed the study area to facilities and utilized ABM to simulate the disease transmission within individual facility, as a tool to measure the protective policies within workplace. Rockett *et al.* (2020) employed ABM to evaluate the possibility of using genome sequencing to reveal unsuspected epidemiological links. They ran an ABM based on synthetic population in Australia with a disease transmission and progression model. The genome sequences in a sub-population of infected people were added as an attribute to corresponding agents, and were tracked during the simulation. The transmission trajectories were then compared with the real epidemiological links. They evidenced it possible to employ the genome sequencing of SARS-CoV-2 to clarify the probable source of infection in cases where epidemiological links could not be determined.

2.4 EpiSim

EpiSim³ is an epidemic simulation based on MATSim. It was implemented by Transport Systems Planning and Transport Telematics group of Technische Universität Berlin. The group has simulated COVID-19 since 2020 in Berlin (Müller *et al.*, 2020a) and the whole Germany (Müller *et al.*, 2020b). ETH IVT group (Penazzi and Axhausen, 2021) also worked on the Switzerland scenario since 2020.

In the implementation, EpiSim combines a human mobility model with an infection model and a disease progression model to simulate the disease spreading (Müller *et al.*, 2020a). There are mainly three sub-models to capture the whole process, namely contact model, infection model and disease progression model.

2.4.1 Contact Model

Contact model is taken from ABM for transportation (i.e., MATSim). Based on the synthetic population and activity chains from transport surveys, MATSim generates complete daily activity chains of agents, including the routes, transport modes, re-adjusted time and locations of activities. Activities are tagged with time, owner (i.e., agent) and locations. Once the simulation of MATSim converges, the trajectories of each agent can be then extracted from the output of MATSim, namely the output events file. EpiSim

³Link to the github repository: <https://github.com/matsim-org/matsim-episim-libs>

uses the output events file as the state of the system at the first day of simulation, extracts co-existence duration of agents at activities or vehicles and uses it as a clue to calculate the infection probability. Then the system iterates the same daily trajectories of all agents and adjusts the activity chains if there are interventions of activity participation.

2.4.2 Infection Model

Infection model is to calculate the probability of infection given two people have contact. Given M potential infectors at a facility at timestamp t , the probability for agent n in the same facility getting infected is defined by (Müller *et al.*, 2020a; Smieszek, 2009)

$$p(\text{infect} \mid \text{contact}) = 1 - \exp \left(-\Theta \sum_m sh_{m,t} \cdot ci_{nm,t} \cdot in_{n,t} \cdot \tau_{nm,t} \right) \quad (1)$$

- $m = 1, 2, \dots, M$, denotes one of all possible infectors within the facility,
- sh is the shedding rate (approximately equal to microbial load),
- ci is the contact intensity which corresponds with the proportion of infectious material spread by potential infector m that is actually ingested by agent n ,
- in is the intake that reduced by some measurements (e.g., masks),
- τ is the duration of interaction between agent n and agent m that is extracted from contact model, and
- Θ is the calibration parameter.

The contact intensity is determined by the room size rs and the air exchange rate ae according to

$$ci = \frac{1}{rs \times ae} \quad (2)$$

$$rs = \frac{N_{\text{personsAtFacility}}}{N_{\text{spacesPerFacility}}} \cdot fs \quad (3)$$

where fs is the floor space per person, $N_{\max}^{\text{personsAtFacility}}$ denotes the maximum number of persons at the facility, and $N^{\text{spacesPerFacility}}$ is the number of sub-spaces within each facility. This is to handle issues when the facility is very large and contains many floors and rooms. Agents that are active in the facility should be further separated into different sub-spaces. The contact intensities and statistics of $N^{\text{spacesPerFacility}}$ for different activities in Switzerland was listed in Table 1 by Penazzi and Axhausen (2021).

2.4.3 Disease Progression Model

Once a person gets infected, the disease progresses in several states until the recovery. The has states *exposed*, *infectious*, *showing symptoms*, *seriously sick* (= should be in hospital), *critical* (= needs intensive care), and *recovered* (Müller *et al.*, 2020a). The transmission probability between states is age-dependent since symptomatic cases among children have lower possibilities to develop to *critical* than among the elderly. The age-dependent transition probabilities in theory was listed in Table 2 by Müller *et al.* (2020a).

3 Methodology and Model Extensions

3.1 Generating Activity Chains in Zurich Scenario

The first step is to generate a realistic contact schedule for the whole population of Zurich. We run a standard MATSim-simulation. The goal is to find a dynamic stochastic user equilibrium. This means having a daily routine for each agent with all places visited including times and transportation modes in between. As is customary in agent based traffic simulations, agents are able to choose transport modes to some extent. Over several simulation steps, agents optimize their mode choice and converge on a global equilibrium. As input for this traffic simulation, we need detailed information about the population and its daily activities. We used a synthetic population that is created along the lines of Hörl and Balac (2021). The approach is to use survey data about daily schedules and infer information to generate a complete synthetic population. For this work, we used a cutout of agents conducting any activities in Zurich or its immediate surroundings from a synthetic population for all of Switzerland. The whole population file used includes around 1.2 million agents who passed through metropolitan area of Zurich. However, the bounding box of their active region is neither congruent with the population of the canton

of Zurich nor the city. If we do not otherwise indicated it, we work with a 25 % random sample of this cutout and scale outputs to the whole population again. This is due to obvious computational and time constraints. The entirety of these agent chosen plans is the input to the next step. Since we have the information, which agents share a bus or spend time at the same facility we can use it as a base for the infection model we saw in Chapter 2.

3.2 Simulation in EpiSim

The parameters of the infection model are taken over from Müller *et al.* (2021) if not otherwise indicated. We mentioned only the core parameter assumptions here.

Table 1 shows the contact intensity ci depending on the activity type relative to the category “other”. We can take the duration of interaction τ directly from the MATSim output. These two parameters define the infection model since we ignore variation of microbial load and intake. This is due to the assumption that just one virus strain was active during our investigation period in the case of microbial load. Mask wearing would be the changing factor for intake, since we do not have date on it we stay agnostic to it. These factors are therefore set to the default value in EpiSim (Müller *et al.*, 2020a). Since the calibration parameter Θ scales the infection probability as a whole, the two factors, intake and microbial load, are taking care of in the calibration procedure. Only the two parameters we have information about, namely contact intensity and duration of interaction, are taken over from the MATSim-output.

Table 1: Contact intensity parameters by activity type

Activity Types	pt interaction	work	leisure	education	shop	outside	home	quarantine_home	other
Contact Intensity	10.0	1.47	9.24	11.0	0.88	0.2	0.8	1.0	1.0

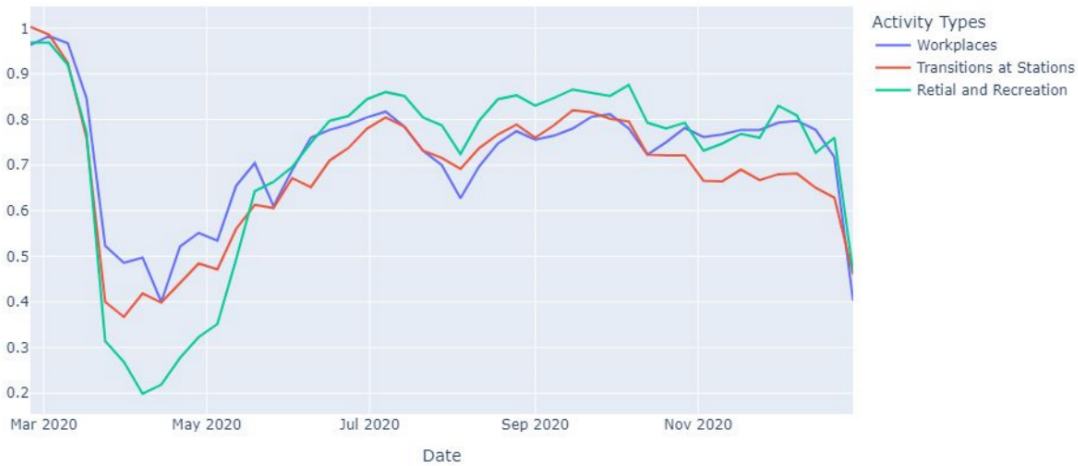
Basically, we imported the output event files of MATSim Zurich scenario as the system state of EpiSim in the first simulation day. Then we ran the EpiSim with the defined infection model and disease progression model. EpiSim also supports models of intervention like vaccinations. However, we did not consider integrating vaccination model since it was still not available in our study period.

We introduced dynamic data to reflect the restricted human mobility pattern during the pandemics. To model the behavior changed due to lockdown measures for shopping, workplace and the related public transport we used data from the Google community

mobility report ⁴, which records the frequencies of visits to facilities in 2020 in Zurich. We regarded the number of visits in early February as the baseline, and the restriction of activities in the following days is represented by the ratio between the number of visits today and corresponding number the baseline. The results are shown in Figure 1. By the beginning of April 2020, the number of visits in workplaces, recreation places and transitions at public stations had already decreased more than 50% compared to pre-pandemic period. Additionally, we also implement the closing of education facilities. Since schools and other education facilities were closed at the 16th of March, we reduced all education activities from this date onward by 90 %. By the 11th of May, schools in the canton of Zurich started to operate in half classes that iterated over the week and universities started to have some courses in person again. We increase the education activities to 50% from this date onward.

EpiSim has an interface to add such restrictions. The total visits of certain activity types were reduced to the appropriate level by randomly removing corresponding activities. Agents cannot infect or be infected during removed activity or the trip to and from the activity.

Figure 1: Restrictions from Google community mobility report



3.3 Calibration

To calibrate our model, we match the number of COVID-19 positive hospitalized patients in the canton of Zurich to the "*seriously sick*" agents in our simulation. Hospitalization

⁴<https://www.google.com/covid19/mobility/>

numbers have the advantage that they are available for a very early date (27th of February 2020). Additionally, the way they are measured has not substantially changed since. It is likely that patients with strong symptoms would go to the hospital, further it is likely that professionals diagnose their disease correctly. However, a possible shortcoming is that elderly patients in nursing homes never reach the hospital even though they would fall under the “*seriously sick*”-category.

We tune two parameters to calibrate the model. The calibration parameter Θ in the infection model is a linear scalar of the infection probability for each encounter between an infected agent with other agents. The second parameter we tune is the date of the first one hundred infections. We induced these initial infections into the model by randomly changing the infection status of one agent each day for one hundred days. We run a selective grid search with these two parameters.

Our model is stochastic and the underlying random sequence of numbers affects the outcome. The initially infected patients are chosen at random. Although the probability if a transmission takes place is deterministic, ultimately it is a stochastic decision. Therefore, we need to run each model with several different seeds. We decided on 12 seeds as the standard for most of our simulation procedures.

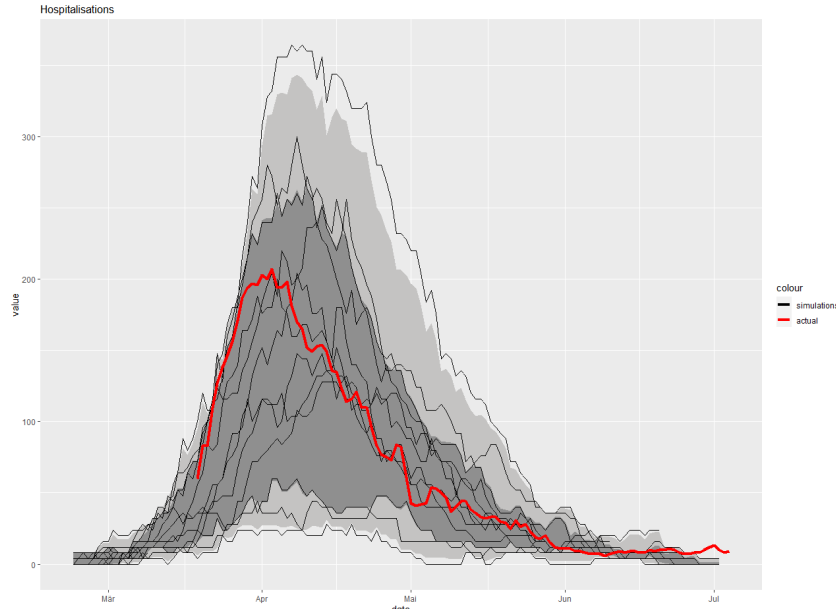
Figure 2 shows our best fit. This simulation is run with a calibration parameter of 9.5×10^{-6} and a start date for the initial infection sequence of the 4th of February 2020. Over the whole period of interest the actual hospitalization numbers lay well within the 95% confidence interval (light gray area) of our simulation and apart from the very right tail within the 65% confidence interval (dark gray area) and therefore closer than one standard deviation from the mean.

4 Results

4.1 Best fits compared to official data

In the previous chapter, we obtained the best fit based on grid search. The calibration parameter $\Theta = 9.5 \times 10^{-6}$, and the start date for the initial infection is February 4, 2020. We ran the best fit with 12 random seeds during the period of interest (i.e., 150 days starting from February 4, 2020) and compared the result to official data.

Figure 2: Hospitalisation: Simulation and Actual



We firstly compared the absolute value between our simulations and the official data. As shown in Figure 3, statistical data was not published in time at the beginning of the pandemic, thus we can see the steep increasing of patients in intensive care unit (ICU) in late April. In our simulation, the curve for the number of patients in ICU goes to the peak in the middle April and steadily drops down in May. But due to the random initialization in EpiSim, different random seeds could reveal very different results. In terms of cumulative number of positive cases, we compared the actual cumulative positives with simulations of different random seeds. Around 100,000 people have the virus after the first wave averagely in our simulation (Figure 4), which takes around 8% of the total population in Zurich. The simulation result is far more than the official data. We suppose the lack of testing boxes and related infrastructure biased the official data during the first wave.

Besides, we compared the actual positives with the simulated positives and simulated positive cases with symptoms. Figure 5 shows the absolute value of the three scenarios by taking average over all random seeds. Cases with symptoms takes around 55% in all simulated positive cases. Even the number of simulated cases with symptoms is still far more than the real test result. We also calculated the share of actual test over positive cases in the simulation. As shown in Figure 4, testing during the first wave found approximately 3% of all actual cases and 6% of cases with symptoms.

Figure 3: Number of COVID-19 patients in intensive care (ICU): actual population in ICU (red line), different simulations (black line)

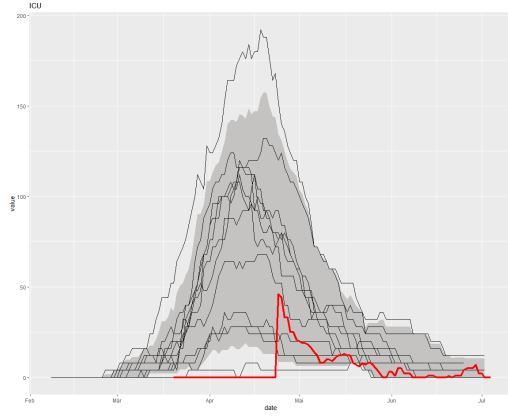


Figure 4: Cumulative number of COVID-19 cases: actual cumulative positives (red line), different simulations (black line.)

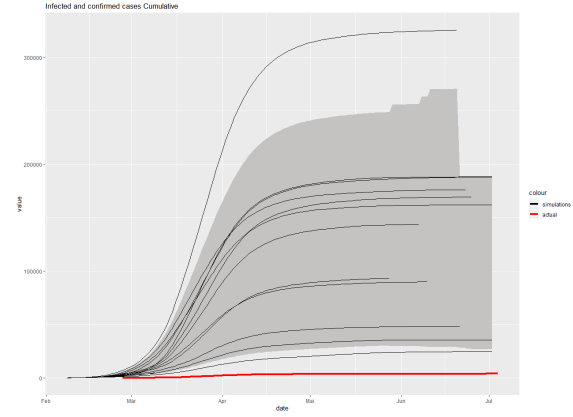


Figure 5: Mean cases 7-day moving average: simulated positive cases (blue line), simulated positive cases with symptoms (black line) and positive tests from official data (red line).

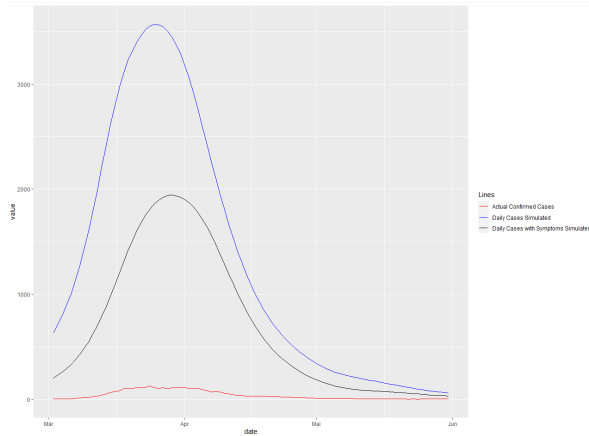
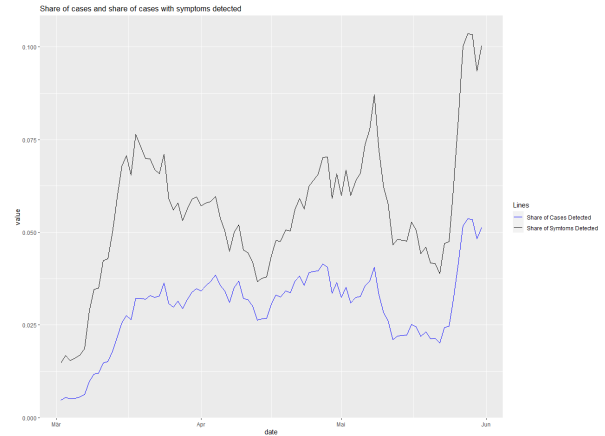


Figure 6: Share of cases detected: the share of positive tests from official data over simulated positive cases (blue line), the share of positive tests from official data over simulated positives with symptoms (black line).



4.2 Transmission Analysis

Our simulation allows looking into the mechanics of how the virus spreads. In table 2 we see the share of infections by activity type. These shares are averages over all simulated seeds. With 44 % the "home" environment stands out as the single type where most infections take place. "Education" facilities (20%), "Leisure" (16%) and "Work" (14%) follow with similar and still substantial shares. "Outside", "Shop" and "Other" make up the rest of infections and they are nearly negligible as infection types. Each transmissions

at home need one member of the household to carry the virus. Necessarily this must be preceded by an infection somewhere outside. We would therefore expect that school closures and closure of leisure and work facilities would have the most effect to contain the spread of the virus. We will verify this in the next chapter.

In figure 7 we can see the distribution of infection counts per infected agent starting at one. Clearly most patients only infect a small number of others. The maximum number is 16 over all simulations. Since these are simulations with a sample of 25% we could multiply these numbers by four. This shows that the right tail of the distribution hosts infectors with huge transmission numbers. Since most households consist of no more than 4 agent, it is likely that many of the infections to the left are of the home type. This makes it likely that these few occurrences in the right tail are centrally located nodes in the network of infections. If this is the case, they could have a large influence on the total spread. The exact impact of these 'superspreaders' remains unclear.

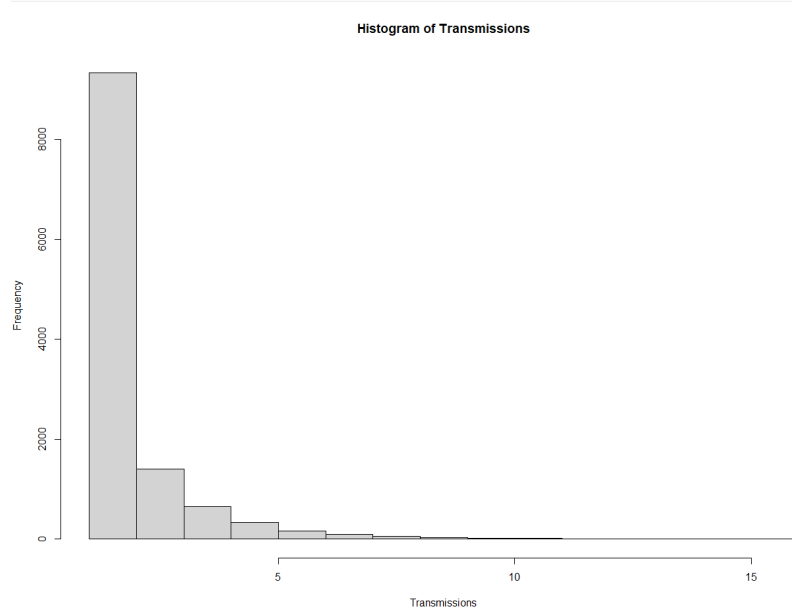
Table 2: Average Share of Infection Types

Infection Type	Share of Infections in %
Home	43.85
Education	19.73
Leisure	16.04
Work	14.17
Outside	5.98
Shop	0.19
Other	0.05

4.3 The Impact of Containment Measures

Our approach enables us to look at counterfactual outcomes of the first wave. We look at how the different containment measures impacted the pandemic in its early stages. We differentiate between four scenarios. The first scenario is that there no containment measures were taken whatsoever. This would mean that all agents behave as before. Not only are no measures enforced agents also do not alter their daily schedule at all. Further, we look at the separate impacts of school closures, on the one hand side, and the partial closure of shopping, work and leisure facilities on the other. School closures and other restrictions are implemented in the simulation as described in chapter 3.2 . Figure 8 displays the resulting mean numbers of seriously sick agents for all four scenarios and their 65% confidence intervals. The scenario with both restrictions is the baseline (blue)

Figure 7: Transmission per Infected Person over all Simulations

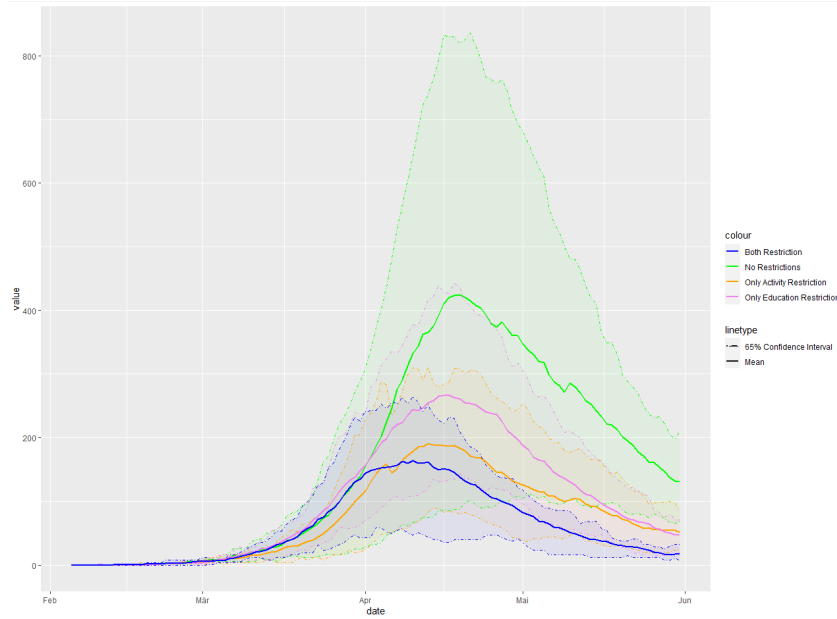


and represents our best approximation of the real data. The mean outcome peaks at 164 seriously sick agents. The scenario with no restrictions (green) peaks slightly later and reaches 424 on average. It also takes much longer for the wave to ebb away. Simply by comparing these two, we can note that the containment measures cut the peak of hospitalization containment measures we are able to separate. It is evident that school closures (violet) alone have a much smaller impact than only the other measures (orange). With a peak of 190 the combined work, leisure and shopping restrictions nearly achieve the same impact as the scenario with both restrictions. Just the closure of education facilities would bring the peak down to 266 hospitalizations. Given the relatively high social costs of school closures documented by (Bayrakdar and Guveli, 2020) and others it is questionable if closing schools was well justified.

4.4 The Scaling Problem

Figure 9 shows how the simulation changes if we use different sample sizes of the population. We use the parameters that were the outcome of our calibration procedure with the 25%-sample to simulate our model with a 50% and a 75% population sample. Figure 9 displays the mean and the 65% confidence intervals of "*seriously sick*" agents for every simulated day. The scenario with the 75% population sample is only presented until the end of April, since it is computationally elaborate to simulate this scenario we do not have enough

Figure 8: Average Prediction and Confidence Interval for Counterfactual Confinement Measures



support for the later stages.

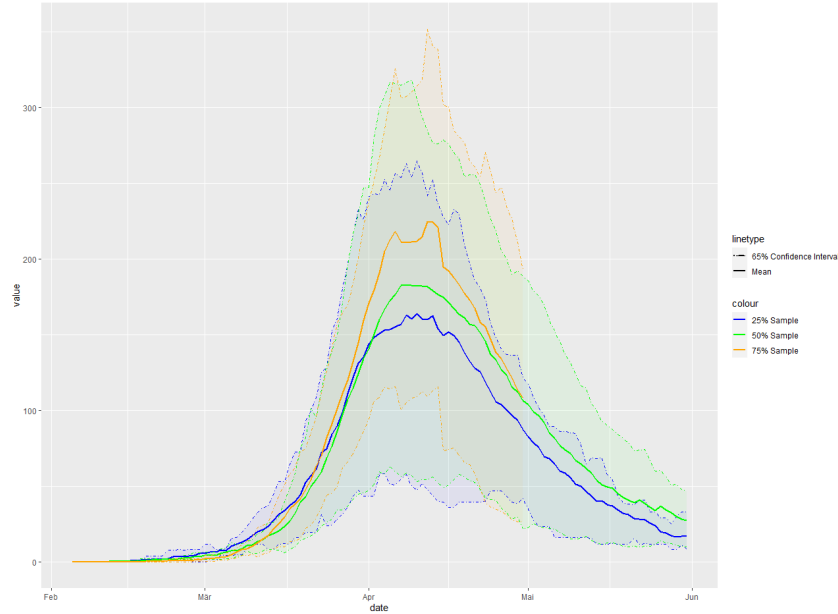
As we would expect the higher the share of the population we used the higher is the peak of the wave. This is because facilities are more crowded the higher the share of the population we look at. This simply increases the amount of encounters between agents and therefore opportunities to infect other agents. The effect is not balanced out just by linearly scaling all outputs. Interestingly this does not lead to an earlier tipping point. The curves are all relatively similar until the beginning of April. The variance is slightly higher for larger population samples. Overall, the curvature of the different samples is similar.

5 Discussion

5.1 Limitations

We want to point out some limitations of this work and present opportunities for future research. As was pointed out, the population that was used for the MATSim-input does not match up exactly with the population of the canton of Zurich. Since we calibrated the

Figure 9: Average Prediction and Confidence Interval for Different Sample Sizes



model with the hospitalization numbers for the canton, this could induce some bias. All the commuters from other Swiss regions that work in Zurich are included in this cutout. This implies that Zurich is a prime place that can easily import infections from other regions since it has a strong exchange of people. This is completely ignored in our model. Likewise, on a higher level Zurich is also an internationally connected economic hub. We also ignored these imported cases. This is probably the reason why after the first wave, that matches real numbers relatively well, no second wave occurs. We consider that a finding that it is imported case that induce further waves.

Our infection model is relatively simple. We ignored microbial loads due to different virus strains since there is no good information on them in these early stages of the pandemic. We were unable to model different masks wearing in any way since we are lack of statistics (e.g., share of people wearing FFP2 or surgical mask). We also missed any additional measures taken to reduce transmission of the virus such as widely adopted hand washing.

We tried to model the changed activity and transport behavior of agents but we were not able to reach realistic level. The Google community mobility report only offers a small number of categories. We were not able to distinguish between shops that were closed such as all non-food sellers compared to shops for daily necessities that stayed open may even have been frequented even more. The same is true for measures like the home-office duty. We did not observe if a work facility is an office or a construction site

that never shut during the lockdown. Additionally, we made the assumption that mode choice behavior is not affected by the pandemic. This is clearly not the case since we know that especially public transport was avoided and other modes such as bikes and cars increased their mode share.

Besides, we initialized the first 100 infections randomly by EpiSim because there is no specific information on the infected cases. However, in some regions (e.g. Hong Kong, South Korea) where the government regularly reports the confirmed cases and their rough home address, it is possible to locate initial infection places to make more accurate simulation. The results from different random seeds can be therefore more similar to each other. Meanwhile, the simulations with spatially registered cases could reveal information like hotspots of infection.

5.2 Conclusion

In this work, we utilized EpiSim, an epidemic simulation engine based on MATSim, to simulate the first wave of the COVID-19 pandemic in Zurich. EpiSim was extended by adding the Zurich scenario. We firstly generated complete activity chains in MATSim according to the Zurich scenario and used it as the input to EpiSim simulation. The model was then calibrated by grid searching parameters. To reflect the dynamic human mobility during the pandemic, we added restrictions to different types of activities during the study period and tested the impact of different restrictions on the disease spreading. Due to the limitation of time and computation, our experiments were based on 25% sub-population samples. We also tested how the model was scaled given different sizes of sub-population samples.

Our simulation reveals a large number of positive cases compared to the reported positives. There is only 3% of total simulated positives were reported. Since we are fitting the first wave, we suppose the lack of testing boxes at the early stage of pandemic may cause bias in official data. Meanwhile, the simplicity of our model also leads to bias in the simulation. We also emphasize the imported cases are crucial for later outbreaks. Even our simulation shows much larger number of positive cases during the first wave, COVID-19 just vanished in June. In the first wave simulation, around 44% of transmissions occur at home. Education facilities, leisure and work follows with similar and smaller share of transmissions compared to home. Transmissions at outdoor or shops take the lowest share. The containment measures, especially the one restricting activities of work, leisure and shopping, have large impact. Finally, using larger population samples increases the

peak of hospitalization numbers but the upticks are nearly the same.

6 References

- Apolloni, A., V. A. Kumar, M. V. Marathe and S. Swarup (2009) Computational epidemiology in a connected world, *Computer*, **42** (12) 83–86.
- Balcan, D., H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck *et al.* (2009) Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility, *BMC medicine*, **7** (1) 1–12.
- Barrett, C. L., K. R. Bisset, S. G. Eubank, X. Feng and M. V. Marathe (2008) Episim-demics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks, paper presented at the *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, 1–12.
- Bayrakdar, S. and A. Guveli (2020) Inequalities in home learning and schools' provision of distance teaching during school closure of covid-19 lockdown in the uk, *ISER Working Paper Series*, **2020-09**, Colchester.
- Belik, V., T. Geisel and D. Brockmann (2011) Natural human mobility patterns and spatial spread of infectious diseases, *Physical Review X*, **1** (1) 011001.
- Connell, R., P. Dawson and A. Skvortsov (2009) Comparison of an agent-based model of disease propagation with the generalised sir epidemic model, *Technical Report*, DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION VICTORIA (AUSTRALIA) AIR
- Cuevas, E. (2020) An agent-based model to evaluate the covid-19 transmission risks in facilities, *Computers in biology and medicine*, **121**, 103827.
- Ferguson, N., D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg *et al.* (2020) Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand.
- Frias-Martinez, E., G. Williamson and V. Frias-Martinez (2011) An agent-based model of

- epidemic spread using human mobility and social network information, paper presented at the *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, 57–64.
- Hörl, S. and M. Balac (2021) Introducing the eqasim pipeline: From raw data to agent-based transport simulation, *Procedia Computer Science*, **184**, 712–719, ISSN 1877-0509. The 12th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 4th International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops.
- Kermack, W. O. and A. G. McKendrick (1932) Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity, *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, **138** (834) 55–83.
- Leventhal, G. E., A. L. Hill, M. A. Nowak and S. Bonhoeffer (2015) Evolution and emergence of infectious diseases in theoretical and real-world networks, *Nature communications*, **6** (1) 1–11.
- Müller, S. A., M. Balmer, W. Charlton, R. Ewert, A. Neumann, C. Rakow, T. Schlenther and K. Nagel (2020a) A realistic agent-based simulation model for covid-19 based on a traffic simulation and mobile phone data, *arXiv preprint arXiv:2011.11453*.
- Müller, S. A., M. Balmer, A. Neumann and K. Nagel (2020b) Mobility traces and spreading of covid-19, *MedRxiv*.
- Müller, S. A., M. Balmer, W. Charlton, R. Ewert, A. Neumann, C. Rakow, T. Schlenther and K. Nagel (2021) Predicting the effects of covid-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data, *PLOS ONE*, **16** (10) 1–32, 10 2021.
- Penazzi, S. and K. W. Axhausen (2021) Episim switzerland scenario, *Arbeitsberichte Verkehrs-und Raumplanung*, **1700**.
- Rockett, R. J., A. Arnott, C. Lam, R. Sadsad, V. Timms, K.-A. Gray, J.-S. Eden, S. Chang, M. Gall, J. Draper *et al.* (2020) Revealing covid-19 transmission in australia by sars-cov-2 genome sequencing and agent-based modeling, *Nature medicine*, **26** (9) 1398–1404.
- Smieszek, T. (2009) A mechanistic model of infection: why duration and intensity of

contacts should be included in models of disease spread, *Theoretical Biology and Medical Modelling*, **6** (1) 1–10.

Tuomisto, J. T., J. Yrjölä, M. Kolehmainen, J. Bonsdorff, J. Pekkanen and T. Tikkanen (2020) An agent-based epidemic model reina for covid-19 to identify destructive policies, *MedRxiv*.

W Axhausen, K., A. Horni and K. Nagel (2016) *The multi-agent transport simulation MATSim*, Ubiquity Press.