Estimating the impact of interventions against COVID-19: from lockdown to vaccination

James Thompson[1*], Stephen Wattam[2]

1 Dept. of Mathematics, University of Luxembourg, Esch-sur-Alzette, Luxembourg
2 W&P Academic Consultancy, Manchester, United Kingdom

March 21, 2021

* j.thompson@turing.ac.uk

Abstract

Coronavirus disease 2019 (COVID-19) is an infectious disease of humans caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Since the first case was identified in China in December 2019 the disease has spread worldwide, leading to an ongoing pandemic. In this article, we present a detailed agent-based model of COVID-19 in Luxembourg, and use it to estimate the impact, on cases and deaths, of interventions including testing, contact tracing, lockdown, curfew and vaccination.

Our model is based on collation, with agents performing activities and moving between locations accordingly. The model is highly heterogeneous, featuring spatial clustering, over 2000 behavioural types and a 10 minute time resolution. The model is validated against COVID-19 clinical monitoring data collected in Luxembourg in 2020.

Our model predicts far fewer cases and deaths than the equivalent equation-based SEIR model. In particular, with $R_0 = 2.45$, the SEIR model infects 87% of the resident population while our agent-based model results, on average, in only around 23% of the resident population infected. Our simulations suggest that testing and contract tracing reduce cases substantially, but are much less effective at reducing deaths. Lockdowns appear very effective although costly, while the impact of an 11pm-6am curfew is relatively small. When vaccinating against a future outbreak, our results suggest that herd immunity can be achieved at relatively low levels, with substantial levels of protection achieved with only 30% of the population immune. When vaccinating in midst of an outbreak, the challenge is more difficult. In this context, we investigate the impact of vaccine efficacy, capacity, hesitancy and strategy.

We conclude that, short of a permanent lockdown, vaccination is by far the most effective way to suppress and ultimately control the spread of COVID-19.

Introduction

The ongoing COVID-19 pandemic is among the most disruptive global events in modern history. At the time of writing, the SARS-CoV-2 virus has spread to almost every country in the world, resulting in over a hundred million infections and over two million deaths. It is of vital importance that we continue to build a rigorous understanding of how the SARS-CoV-2 virus spreads and predict the impact of interventions, to help policy makers formulate effective strategies that save lives while simultaneously balancing the economic and social impact.

Central to such a strategy is a recognition of heterogeneity and behavioural diversity. Indeed, the regional impact of COVID-19 has been extremely variable. For a given region, the impact of any infectious disease depends fundamentally on who lives in that region, how these individuals interact with one another and how this population connects with the populations of other regions. For a given individual, factors such as age, sex, ethnicity and the presence of underlying medical conditions might determine how that individual responds to an infection. Disease transmission is not only determined by the nature of the disease itself, but also by a multitude of factors relating to human behaviour. Such factors might include the time of day, the day of the week, the climate, seasonal effects and the prevailing culture of the region. These underlying variables result in correlations, producing an extremely complex and computationally irreducible system of social interactions and disease dynamics, beyond the scope of simple mathematical theory. Modelling the impact of public health policy, in the context of an infectious disease such as COVID-19, is therefore necessarily difficult and subject to unavoidable limitations.

One commonly used indicator of epidemic dynamics is the effective reproductive number R_t, defined roughly as the expected number of secondary infections caused by a typical infected individual at time t. This number aggregates the factors mentioned above by simultaneously averaging over individuals and individual behaviour. It is not therefore possible to measure the true R_t of a population (as opposed to that of a particular model). Additional simplifying assumptions, on the population and its mixing habits, are required in order that R_t be estimated. The most basic assumption supposes that all individuals are identical and mix with one another with equal probability. In a sufficiently large population with sufficiently many individuals infected, such mass action might be realistic, but in circumstances where the proportion of infected individuals is low, it neglects the unpredictable nature of interactions between small numbers of people. Nevertheless, such homogeneity assumptions give rise to a number of popular mathematical and computational models, including the equation-based compartmental models [1]. Such models typically use ordinary differential equations to keep track of how many individuals are in various health states at various times, sometimes stratified by age or households. The R_t associated to such a model can be fairly easily calculated, as well as certain other quantities of interest, for example limiting equilibria.

The equation-based approach to epidemic modelling could be considered the *top-down* approach, which postulates a set of equations whose solution, after appropriate configuration, is supposed to describe the system in question. Such an approach has the advantages of flexibility and speed, typically involving only a small number of parameters, but on the other hand is unable to capture the heterogeneity and granularity obtained using the *bottom-up* approach of an agent-based model. In an agent-based model, the simultaneous actions and interactions of multiple individuals, referred to as agents, are simulated in an attempt to re-create and predict the emergence of complex phenomena as a result of their collective behaviour.

Agent-based models are computationally intensive, and therefore have risen to prominence only in recent decades, with one of the earliest advances being John Conway's Game of Life [2]. Agent-based models have been applied across many areas of study, for example ecology [3], social science [4], macroeconomics and financial markets [5] and epidemiology. Agent-based models have been used extensively to study the spread of infectious diseases including COVID-19.

One approach to agent-based models is based on network theory, in which real-world contact networks are represented by graphs, with vertices representing individuals and edges connecting individuals who with some probability are able to interact with one another. Various mathematical tools from graph theory can be applied to such models, resulting in a topological or geometric analysis of the underlying network [6]. The results of this article, however, are obtained using an agent-based model operating on slightly different principles.

In particular, the contact network associated to our model is dynamic and based on collocation. At each moment, contacts are described by a partition of the total population, with each subset corresponding to a particular location, for example a house, restaurant or shop. These subsets describe who is in each location at each time, with homogeneous mixing occurring internally. As individuals move between locations, the subset of individuals present in a given location is updated accordingly. On top of this framework sits the disease model and a range of interventions.

Our model is custom-built, featuring numerous heterogeneous dimensions and substantial behavioural diversity. It is able to capture both spatial and temporal variations in disease dynamics. The model consists of four basic layers, described as follows:

- **Locations:** A procedurally generated random environment of locations.
- **Agents:** A heterogeneous population with daily and weekly routines defined on a 10 minute time resolution.
- **Disease model:** An age-dependent compartmental model featuring hospitalization and intensive care.
- **Interventions:** Implementations of a broad range of public health interventions.

Interventions are the means by which a policy maker can control or suppress an epidemic. Interventions are either pharmaceutical or non-pharmaceutical. The World Health Organization divides the latter into four categories [7]. First there are the personal protective measures, which includes improved hand hygiene, respiratory etiquette and face masks. Second are the environmental interventions of improved ventilation and surface and object cleaning. Third are the various physical distancing measures, including such things as quarantining, school closures, workplace measures, closure of businesses, cancelling of events, curfews and lockdowns. Fourth and finally are the travel-related measures, referring to travel advisories, entry and exit screening, internal travel restrictions and border closures. Various combinations of these interventions have been implemented by governments around the world in response to the COVID-19 pandemic. In addition, individuals themselves have responded to the pandemic with self-imposed changes of behaviour, avoiding crowds or other perceived high-risk situations. In addition, a key component of most control strategies is testing and contact tracing, with various COVID-19 tests having been developed that determine, with some degree of uncertainly, whether or not an individual is infected with the virus. The impact of testing results from how the information obtained is used, for example when implementing the other interventions. Accompanying the non-pharmaceutical inventions are the pharmaceutical inventions, in particular anti-viral therapies and, perhaps most importantly, vaccination.

Vaccination is generally considered the most effective method of preventing infectious diseases, with mass vaccination campaigns having achieved the global eradication of smallpox and the suppression of diseases such as polio, measles and tetanus from much of the world, thereby saving hundreds of millions of lives. Controlling COVID-19 on a global scale cannot be achieved using only the non-pharmaceutical interventions listed above, associated to which are enormous economic and social costs, and therefore mass vaccination against COVID-19 will form a central part of any successful COVID-19 control strategy. There are well known mathematical models of the relationship between vaccination and herd immunity, for example [8].

Several COVID-19 vaccines have been developed and tested (see, for example, [9–12]) and are now being distributed in a number of countries around the world. In most countries, vaccines are being administered according to a priority list, starting with either those individuals who most require immediate protection against the disease, or those individuals for whom reduced transmission will be of the greatest benefit from a public health

perspective. Besides the manufacturing and logistical challenges associated with mass vaccination, there is also the issue of vaccine hesitancy [13–15], which refers to the fact that significant numbers of people would prefer, for various reasons, not to get vaccinated. Assessing the impact of vaccination, against the backdrop of various overlapping non-pharmaceutical interventions, is therefore challenging.

The objective of this article is to use our computational model to compare interventions according to their epidemiological impact. We consider, in particular, the following questions:

- How do non-pharmaceutical interventions compare, in terms of their impact on cases and deaths?
- At what level is herd immunity achieved?
- To what extent does the success of a vaccination campaign depend on efficacy, daily capacity and hesitancy?
- How does a vaccination strategy that focusses on reducing deaths compare to one that focusses on reducing transmission?

In this article, we are not concerned with the economic or social costs of the interventions. Moreover we do not look for optimal strategies, this being instead a topic for future research. Calibration will focus on cases, hospitalizations and deaths, avoiding such things as the basic or effective reproduction numbers. We will measure the impact of interventions by comparing cases and deaths with the baseline scenario in which no interventions are active. We will suppose vaccination is implemented in a two-dose format, with an interval of time between doses, with limited daily availability and a priority scheme that administers doses to certain individuals before others, based on their age, living arrangements or place of work, and who potentially refuse the vaccine with a certain age-dependent probability.

Our model is configured to represent Luxembourg, a small western European country with a population on 1st January 2020 estimated at 626,108, together with populations of cross-border workers in the neighbouring countries of Belgium, France and Germany. Input data therefore comes from various institutions and surveys associated with Luxembourg. Consequently, this article investigates the impact of interventions specifically in Luxembourg, although the model itself is flexible and can be adapted to other regions. Luxembourg, however, is particularly interesting because, while being an independent nation with its own unique response to the COVID-19 pandemic, it has a population small enough to be within the reach of a computational agent-based model.

Our code and all output data generated by the code, used to plot figures or otherwise underlying the results presented in this article, will be made publicly available on GitHub. Census data, including data on age distribution and household structure, were obtained from STATEC, the government statistics service of Luxembourg, and are publicly accessible [16]. Public transport data came from Mobilitéit [19] and the Ministry of Mobility and Public Transport (MMTP) of the government of Luxembourg [20]. Population grid data came from the 2011 GEOSTAT study, organized by Eurostat [21]. Location counts came from STATEC and OpenStreetMap [18]. Behavioural and mobility data came from the 2014 Luxembourg Time Use Survey and the 2017 Luxmobil Survey, these datasets being the property of STATEC and MMTP, respectively, available for researchers who meet the criteria for access to confidential data. COVID-19 clinical monitoring data is the property of IGSS, the General Inspectorate of Social Security of Luxembourg, with the relevant dataset being available from IGSS for researchers who meet the access criteria [17]. Interventions are otherwise parametrized using public knowledge.

Chief among the unknown parameters in our model are the transmission probability, initial exposure count and asymptomatic probability. The calibration of these and other parameters is discussed in the model evaluation section.

The organization of the paper is as follows. In the next section, we briefly describe the state of the art, referencing only a small sample of articles from the immense body of research that has emerged since the start of the COVID-19 pandemic. In the section after we describe our model. This is followed by a section on model evaluation, in which we discuss the processes of verification and validation and the limitations of the model. After that we present and discuss our main results. Finally, in the last section, we draw conclusions, while making further remarks about the limitations of the study and directions for future research.

State of the Art

Since the start of the pandemic, models based on ordinary differential equations have been used to study the impact of interventions against COVID-19. In [22], the authors used an equation-based compartmental model to study the impact of vaccination and other interventions on the shape of epidemic curves in Luxembourg. Such a model was also applied to Luxembourg in [23], to study the interplay between the epidemiological and economic aspects of the COVID-19 pandemic. Multiple authors have used equation-based models to study optimal strategies for lifting restrictions [24] and vaccination [25, 26]. A approach utilising Bayesian techniques, and a game theoretical modelling of adherence to restrictions, has been applied in [27], while the use of game theory and social network models for decision making on vaccination programmes has been further emphasised in [28].

The article [29] presents an approach to modelling spatio-temporal vaccination strategies that uses stochastic differential equations. Therein, individuals move within a continuous space according to Brownian motion dynamics and, when they find themselves within a certain distance of one another, interact and potentially transmit the virus. The system of stochastic equations is then used to describe the number of individuals who are susceptible, exposed, infectious and recovered at each time. This is then used by the authors to derive a mean-field statistical model, from which they draw conclusions. Our model also features spatial dimensions, and therefore could be used to investigate spatial strategies, for example ring vaccination, however this is beyond the scope of the present study.

Moving beyond the equation-based models to the agent-based models, we draw attention to the following three open-source agent-based COVID-19 models: OpenABM-Covid19 [30], Covasim [31] and COMOKIT [32]. OpenABM-Covid19 and Covasim assume individuals mix homogeneously outside households, workplaces or schools, drawing the number of random connections an individual makes throughout a day from an over-dispersed negative binomial or a Poisson distribution. On the other hand, [32] is somewhat more similar to our own model, with a dynamic contact network developed via mobility and daily agendas. Some researchers have used these open source models, while others have developed their own. For example, Laurent Mombaerts and Atte Aalto have also developed an agent-based model for Luxembourg, somewhat different from our own, using social security data to construct a contact network. Their model has been used in the recently published article [33] to study the large-scale COVID-19 testing programme in Luxembourg.

The impact of vaccination on cases, hospitalisations and deaths has been studied using agent-based models in [34] and [35], these two articles focussing on areas in Canada and the United States, respectively. For each individual, these articles assume a static, empirically determined contact network and sample the number of daily contacts from a negative-binomial distribution. The authors of both articles assume a predetermined coverage rate achieved by the vaccination campaign and a specific vaccination rate of 30 individuals per 10,000 population per day, with efficacy against symptomatic infection set to 95%. Various levels of pre-existing immunity were also assumed, ranging from 5% to 20%, depending on the region. In the article [36], the authors use an agent-based model to study the optimal arrangement of drive-through vaccination stations. In the article [37], a dynamic contact network was constructed in order to study the optimal choice of

vaccination strategy under a partial or complete lockdown or without any non-pharmaceutical interventions active at all. Each of the individuals appearing in this network had a pre-assigned daily routine, specified on the resolution of 1 hour, with the routine determining the order in which the individuals move between different locations, such as workplaces, schools, public places, hospitals and homes. The effect of vaccination combined with non-pharmaceutical interventions including reduced mobility, school closure and face mask usage was also studied in [38], for the state of North Carolina. In that model, individuals interact only in locations such as the home, work and school and move between those locations in the morning and in the evening each day. The paper investigates scenarios under which vaccine efficacy takes the values of 50% or 90%.

This body of work is rapidly growing. Compared to models found in the existing literature, our model appears to have a more detailed and dynamic interaction system, containing a greater range of location types than in any of the works mentioned above, with an extremely fine time resolution of only 10 minutes and over 2000 behavioural types, allowing our model to capture the sort of brief encounters that take place outside of homes, work and schools. Our model contains a broad set of interventions, including vaccination, and is the first agent-based model to be applied directly to the study of mass vaccination against COVID-19 in Luxembourg.

Methods

Our model is written in Python. The code base is organized around a modular framework, in which components represent submodels. This has the advantage that new components, such as additional interventions, can easily be added while existing components can be quickly updated or replaced. A communications system handles messages sent between the various components, a crucial feature since many of the interventions are required to interact with one another, while a scheduling system handles the timing of events such as lockdowns and testing regimes. The code will soon be open source and available on GitHub.

All input data is found in a single configuration file separate from the rest of the code. Using this file we are able to configure the model to represent COVID-19 in Luxembourg or, given appropriate data, a different disease in a different region. The model is very flexible, but as with most agent-based models [39] has the limitation of long run times for large populations.

We will now present an overview of the various layers of the model, describing the key components and the generic parametrization of submodels. Scenario-specific parametrizations used for validation will be discussed in the model evaluation section, while experimental parametrizations will be discussed in the results section. A description of the model according to the ODD protocol [40] can be found in the appendix.

Location Types

The lowest layer of the model consists of a procedurally generated random environment, representing the region in question. The environment consists of locations, categorized by type and each assigned spatial coordinates. By assigning coordinates, our model is able to simulate the spatial dynamics of an epidemic. The list of location types includes:

- **Houses, Care Homes, Hotels, Primary Schools, Secondary Schools, Restaurants, Shops, Hospitals, Medical Clinics, Places of Worship, Indoor Sport Centres, Cinemas or Theatres, Museums or Zoos, Cars, Public Transport, Outdoors**.

The remainder of the list consists of other types of working location, categorized by sector:

- **Agriculture, Extraction, Manufacturing, Energy, Water, Construction, Trade, Transport, Catering and Accommodation, ICT, Finance, Real Estate, Technical, Administration, Education, Entertainment, Other Services**.

The model is configured to feature as many locations of each type as are present in the region in question, in our case Luxembourg. If the simulation population size is configured to be smaller than the true population size, then the numbers of locations appearing in the model are scaled down accordingly, together with other relevant quantities. Smaller populations are useful from the point of view of code testing, thanks to a reduced runtime.

In the case of Luxembourg, location counts are derived from a number of different sources. Table 1 lists the location counts for types for which we use data from OpenStreetMap (OSM), a collaborative project that aims to build a free editable map of the world.

Table 1. Estimated location counts in Luxembourg derived from OSM data, 2020.

Location type	Count
Care Home	52
Restaurant	2247
Shop	3136
Medical Clinic	125
Hospital	11
Hotel	213
Place of Worship	677
Indoor Sport	199
Cinema or Theatre	34
Museum or Zoo	77

The numbers of primary and secondary schools, as well as other working locations categorized according to sector, are estimated using data from STATEC, the government statistics service of Luxembourg. These numbers were published in the 2019 edition of their Répertoire des Entreprises Luxembourgoises [41]. Some care was taken to avoid overlap with working location types already listed above, the adjusted estimates being tabulated below in Table 2.

In addition, schools are divided into classrooms. In the case of Luxembourg, STATEC data indicates that, on average, each primary school consists of 17 classes while each secondary school consists of 34 classes. Modelling the classroom structure avoids excessive crowding in schools, but has the drawback of limiting interaction between students in different classes. In Luxembourg, however, most students remain in the same class for all subjects so in this case the assumption is perhaps reasonable.

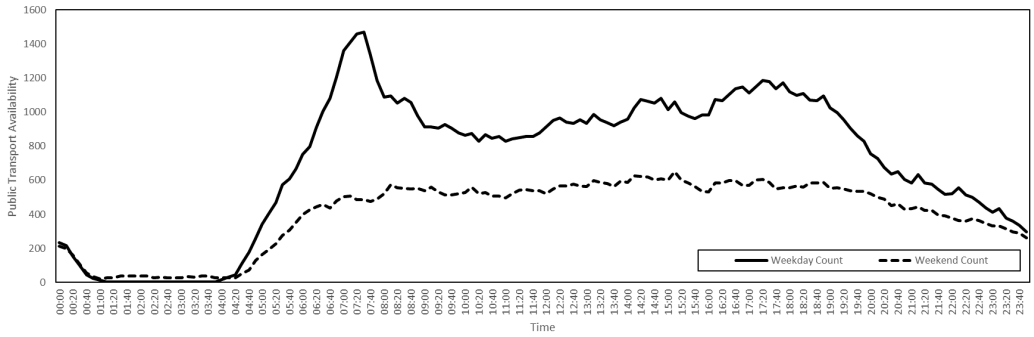
Some locations types do not appear in these tables and are subject to special treatment. For example, public transport is implemented in such a way as to produce a variable number of units of public transport at each time. A unit of public transport is defined to be either a bus or a carriage deck of a train or tram. A single-deck carriage consists of one unit, while a double-deck carriage consists of two units. The total number of buses and rail compartments operating in Luxembourg can be derived from publicly accessible timetable data published by Mobilitéit. We used data referring to the period starting on 4th November 2019 and ending on 14th December 2019. Estimating average units per train at 10, average daily public transport availability in Luxembourg can then be visualized as in Fig 1 and is used to configure the variable number of accessible locations of type **Public Transport**.

There is also a single outdoor location **Outdoor**, in which we assume zero disease transmission, and a **Cemetery**, to which agents are moved after death. In the Luxembourg

Table 2. Estimated location counts in Luxembourg derived from STATEC data, 2020.

Location type	Count
Primary School	181
Secondary School	58
Agriculture	86
Extraction	11
Manufacturing	785
Energy	98
Water	71
Construction	4366
Trade	4684
Transport	1349
Catering and Accommodation	251
ICT	2752
Finance	1241
Real Estate	1458
Technical	8349
Administration	2433
Education	838
Entertainment	87
Other Services	1523

Fig 1. Public transport availability on a typical day in Luxembourg, 4th November 2019 to 15th December 2019.



implementation, there are also three border country locations, namely **Belgium**, **France** and **Germany**.

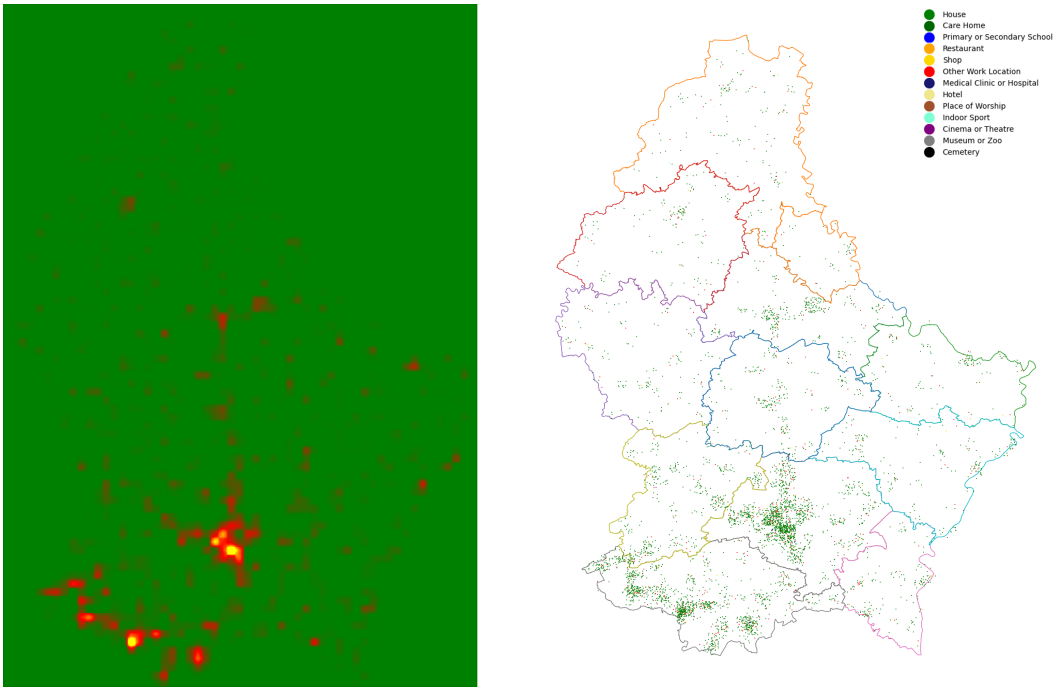
The number of locations of type **House** is determined by an algorithm that assigns agents to homes. This algorithm is described later. The number of locations of type **Car** is set equal to the number of houses, with each house being assigned one car. As with the units of public transport, the cars in our model are, for simplicity, static. The cars are simply locations in which agents are placed should they wish to use a car. In particular, agents living in the same house will use the same car, no matter their destination. If an agent chooses to use public transport, then a unit of public transport is randomly selected among all those available at the time.

Spatial Distribution

Locations are assigned spatial coordinates by randomly sampling the population distribution of the region. In the case of Luxembourg, the population distribution is described using population grid data collected by Eurostat's 2011 GEOSTAT initiative. This grid data specifies the number of people living inside each 1km square, with the grid format being that of the ETRS89 reference frame. Note that such grid data is available for countries across the European Union.

We also have the option of sub-sampling the grid data to produce a grid of finer resolution. For example, with a resolution factor of 2, each original square with edge length 1km is replaced by four smaller squares each of edge length 500m. Population is then distributed among the small squares by linearly interpolating, with the option of setting the population of a small square equal to zero if there was no population present in the original square. Our population distribution model for Luxembourg, obtained using a resolution factor 2 and areas of zero population preserved, is illustrated as a heat map below, in Fig 2, together with a sample distribution of locations.

Fig 2. Population distribution in Luxembourg, derived from interpolated grid data, 2011. On the left, a heat map of population in Luxembourg. On the right, 10000 locations of the various types appearing in our model, distributed randomly according to the weighting illustrated on the left and plotted against a map of Luxembourg and its 12 cantons.



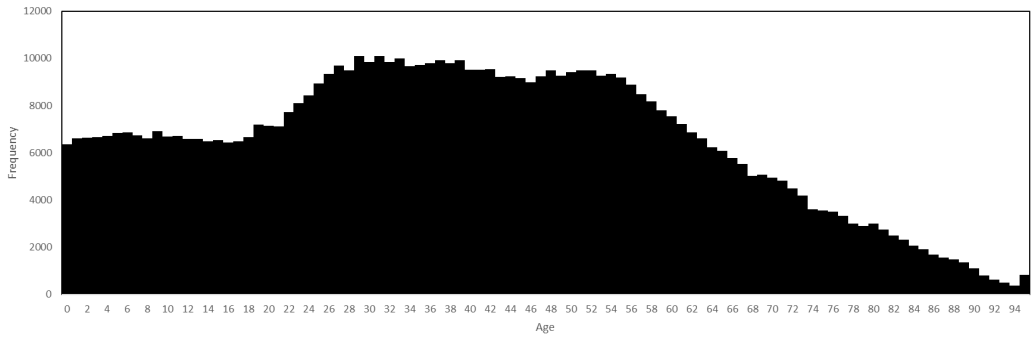
Since we set the spatial coordinates of a location by sampling the (interpolated) population distribution, we implicitly assume that all types of location are distributed as population is distributed. While this is approximately true, some location types are, in reality, subject to additional clustering. An improvement to the model would be therefore to assign coordinates using type specific spatial distributions, possibly achieved using additional OSM data, to produce a slightly more realistic environment.

Agents

Having generated a static environment of locations, the next step is to populate this virtual world with agents. The agents in our model represent individuals. Agents are assigned a country of residence and an age. We do not assign sex, ethnicity nor the presence of underlying medical conditions.

Age is distributed according to the population of the region in question. In the Luxembourg model, age is distributed as in Fig 3, this data having been collected by STATEC, representing a resident population of 626,108 on 1st January 2020. We have suppressed the age category 95+ to 95.

Fig 3. Age distribution in Luxembourg, 1st January 2020.



In addition to the resident population, we also generate populations of non-resident commuters who live in neighbouring countries. Luxembourg shares borders with Belgium, France and Germany and large numbers of people travel across these borders every day for a variety of reasons. We focus on those who cross the border for work, since these are the individuals who typically spend large amounts of time in the region and who travel on a regular basis. We assume that populations of cross-border workers consist only of adults, that the age of cross-border workers is distributed identically to that of adults in the resident population, and that cross-border workers travel to the region for work and for no other reason. According to STATEC, the numbers of cross-border workers travelling to Luxembourg are given in Table 3.

Table 3. Populations of cross-border workers employed in Luxembourg, 2019.

Country	Workers
Belgium	47173
France	104070
Germany	46863

We do not model air travel nor other long distance connections between regions.

Activity Choice

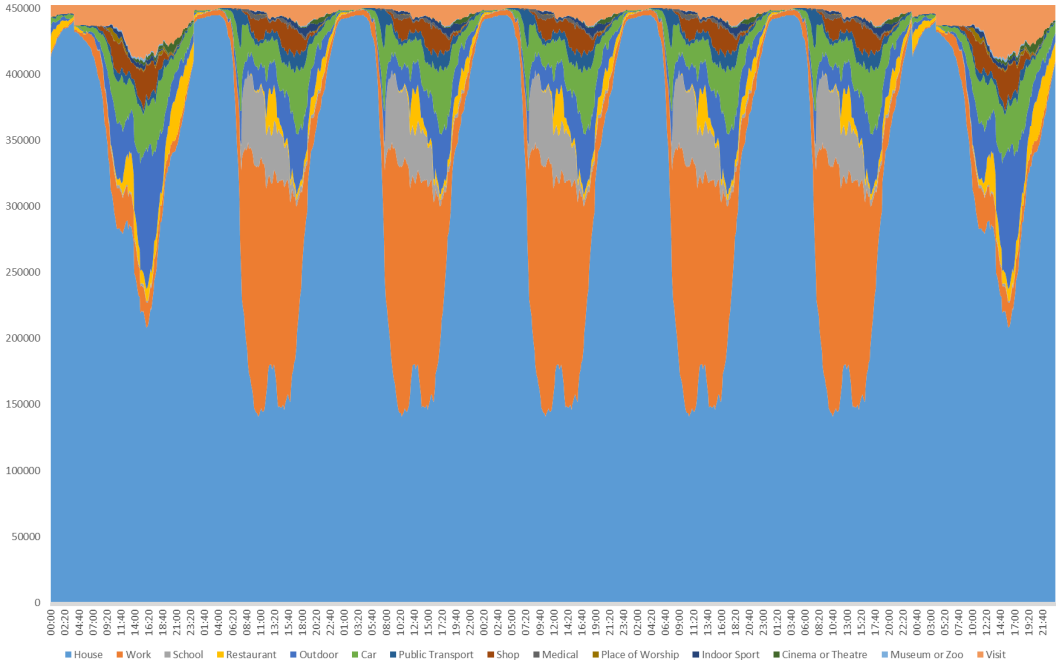
Agents are able to perform various activities. Activity selection is based on time use data. The Harmonised European Time Use Surveys (HETUS) [42] are national surveys conducted in European countries to quantify how much time people spend on various activities, including paid work, household chores and family care, personal care, voluntary work, social life, travel and leisure. Similar data are collected in other countries, such as the United States. Respondents to the European surveys were asked to record dairies of both a week day and a weekend day, with a time resolution of 10 minutes. In other words, for each

respondent, the time use data specifies what the respondent was doing during each 10 minute interval of each day. The list of activities recognised by the survey is long and therefore simplified for our purposes, resulting in the following list of activities appearing in our model:

- **Home, Visit, Work, School, Restaurant, Shopping, Outdoors, Car, Public Transport, Medical, Worship, Indoor Sport, Cinema or Theatre, Museum or Zoo.**

The activity **Home** refers to all domestic activities, such as cleaning, cooking and sleeping. The activity **Outdoors** includes such things as going for a walk, riding a bike or playing outdoor sports. The activity **Visit** refers to visits of family or friends in other houses or care homes. The activity **Medical** refers to medical activities not related to the epidemic, and places agents either in hospital or a medical clinic. The other activities are self explanatory. We construct weekly routines by concatenating 2 copies of the weekend dairy with 5 copies of the weekday diary for each respondent, with the week starting on a Sunday. We therefore do not distinguish between Saturday and Sunday nor between weekdays. In the Luxembourg implementation, data is derived from the 2014 Luxembourg Time Use Survey. The resulting distribution of activities performed each week is illustrated below in Fig 4. Differences between weekend and weekday behaviour are clear, as are features such as rush hour, lunch breaks and increased time spent outdoors at the weekend.

Fig 4. Distribution of activities performed each week. The week starts on a Sunday and the total magnitude of the vertical scale is equal to the total statistical weight of the time use dataset.



Since the age of respondents in the HETUS is known, we can assign agents weekly routines according to age. We do this by associating to each resident agent the routine of a respondent randomly selected from those of a similar age and according to the statistical weights attached to data. This results, in the Luxembourg implementation, in over 2000 unique behavioural types. The minimum and maximum ages of respondents to the HETUS

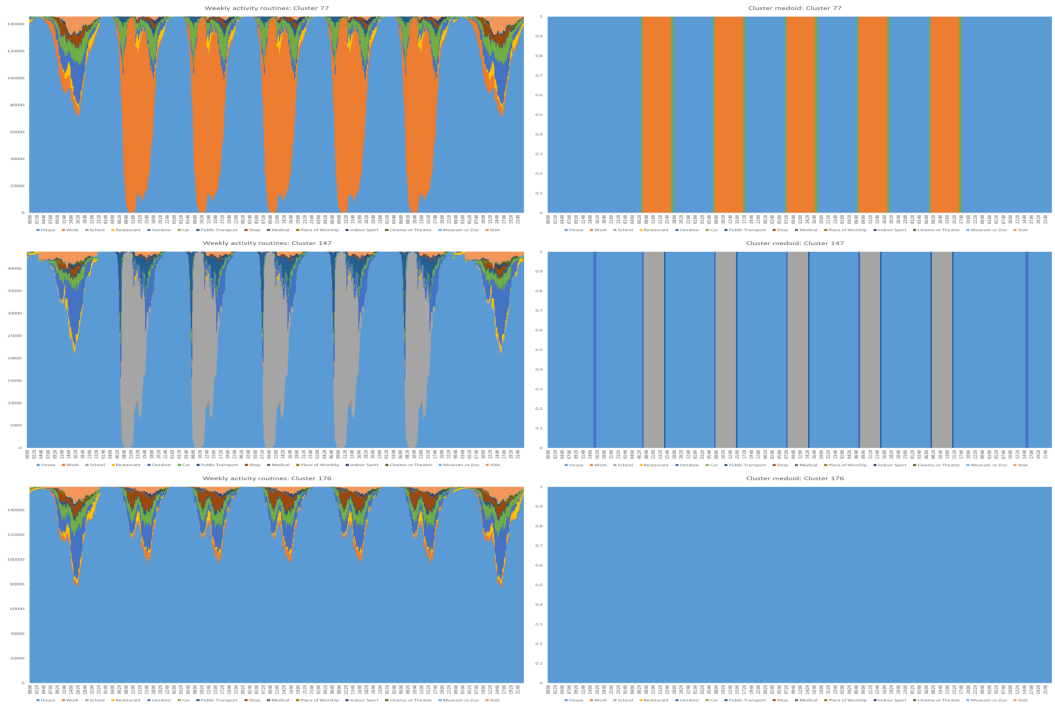
are 10 and 75, respectively, and we therefore introduce special rules for the very young and very old, in order to produce what we believe is a reasonable behavioural model covering agents of all ages.

Since the resolution of the time use data is 10 minutes, a weekly routine can be thought of as a vector of length 1080, with entries specifying which activity is to be performed at each corresponding time. For example:

$$[\mathbf{Home}, \mathbf{Home}, \mathbf{Work}, \mathbf{Work}, \cdots, \mathbf{Restaurant}, \mathbf{Home}].$$

Each agent is assigned such a vector. We can put a distance on the space of all such routines by summing the number of entries in which the activities of two routines differ. Doing so we can perform hierarchical clustering to determine if there exist naturally occurring behavioural types. A distance threshold of 250 yields a total of 358 clusters, the three largest of which, labelled 77, 147 and 176, are illustrated below in Fig 5.

Fig 5. Routine clusters. On the left, the weekly routines of the three largest clusters with, on the right, the corresponding medoid routines. The routines are constructed using data from the Luxembourg Time Use Survey, 2014.



Cross-border workers are assigned the canonical working routine given by the medoid of Cluster 77. This ensures that cross-border workers really do cross the border and go to work, since random sampling would have many of them performing other activities instead.

We also experimented with a more complicated activity model where agents choose activities randomly. This involved aggregating routines in such a way as to produce transition matrices and corresponding time-inhomogeneous Markov chains, the sampling of which generates infinitely many behavioural patterns. The drawback of this approach is the computational cost and the possibility of a sampling unrealistic routines, so for simplicity we decided to stick with the deterministic system described above, in which agents read off which activity to perform next using their given routine vector.

Having selected a preferred activity, an agent must then decide where to perform that activity. For example, if an agent decides to go **Shopping**, then the agent must choose a

Shop at which to do the shopping. Agents are grouped into households and assigned a place of work, together with sets of locations at which they can perform the other activities.

Households

The home of an agent is the location in which they perform the activity **Home**. Home assignment begins by populating care homes with the most elderly residents and by setting the home of non-residents to be their country of origin. We assume that each care home contains 38 residents. We will assume that no internal transmission occurs within the neighbouring countries, focussing instead on transmission within the central region only. Remaining resident agents are then assembled into households, with household composition for the Luxembourg model being determined using population structure data on families and households collected by STATEC for the 2001 census. Data on the numbers of children and retired individuals in houses of various sizes in Luxembourg is tabulated below, in Table 4.

Table 4. Household structure in Luxembourg, 2001.

		Size of Household						
		1	2	3	4	5	6	7+
Children aged 14 years or less	0	50384	46191	15885	9035	2514	784	276
	1	0	2382	12257	4972	2041	514	149
	2	0	0	1108	14021	2180	720	205
	3	0	0	0	253	4165	622	219
	4	0	0	0	0	38	737	150
	5+	0	0	0	0	0	5	146
Persons aged 65 years or more	0	33586	30677	25001	26510	9743	2706	821
	1	16798	6893	2388	1240	971	445	191
	2	0	11003	1656	483	207	214	112
	3	0	0	206	36	11	12	19
	4	0	0	0	12	4	3	0
	5+	0	0	0	0	1	2	2

Note that in our implementation, the categories 5+ and 7+ are suppressed as 5 and 7, respectively. The largest private household in our model of Luxembourg is therefore of size 7. Using only the data contained in these tables, we are able to construct a discrete probability distribution on household types. For a household of size n, a household type is a triple (c, a, r) where c, a and r denote the numbers of residents in the ages categories 0-14, 15-64 and 65+, respectively, with $c + a + r = n$. For example, a household of size 5 containing two children, two adults and one retired person would be encoded $(2, 2, 1)$. If N denotes the total number of households in the census data, with $C_n(c)$ and $R_n(r)$ the numbers of households of size n with c children and r retired, respectively, then we postulate that

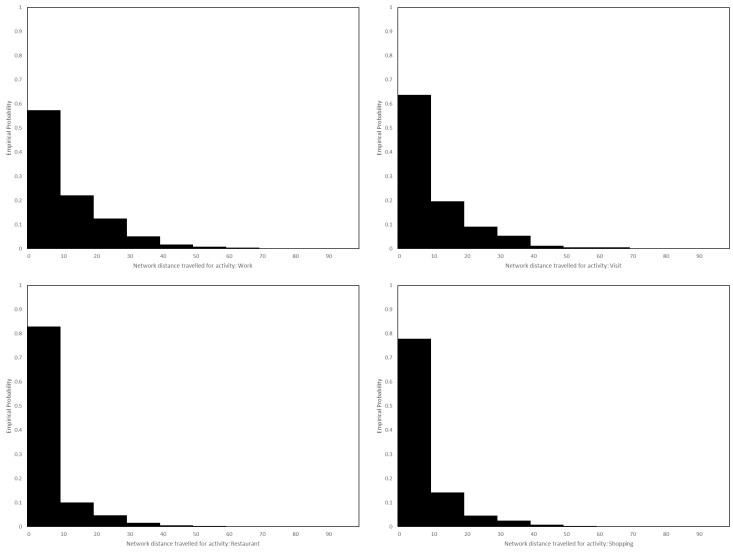
$$\mathbb{P}((c, a, r)) = \frac{C_n(c)R_n(r)}{2N}\left(\frac{1}{\sum_{i=0}^{n-r} C_n(i)} + \frac{1}{\sum_{j=0}^{n-c} R_n(j)}\right)$$

where $\mathbb{P}((c, a, r))$ denotes the probability of the profile (c, a, r) occurring. Note that this does indeed yield a discrete measure with unit total mass. During the initialization phase of our model, houses are generated with profiles sampled from this distribution and populated with appropriate numbers of agents taken randomly from the three age groups. Houses are spatially distributed as the other locations, according to interpolated population grid data. While this process of generating households could be improved with more detailed data on household composition, using only Table 4 our method appears sufficiently accurate.

Location Choice

After home assignment, agents are then assigned a place of work, to which they will move if performing the activity **Work**. First, for each agent, a subset of all working locations is sampled uniformly at random. Working with only a subset reduces the computational cost of the next step, which involves assigning to each workplace in the sample a weight, obtained by multiplying together two subweights. The first is given by the expected number of workers at that location, configured for the Luxembourg model using STATEC data published in the 2019 version of their Répertoire des Entreprises Luxembourgoises. The second is determined using mobility data and the distance to the agent's house. In particular, we appeal to the 2017 Luxmobil Survey, in which respondents were asked to record how far they travelled (in terms of network distance) when doing so for various reasons. We have plotted aggregations of this data, for a selection of activities including **Work**, in Fig 6.

Fig 6. Distributions of travel distance. Distributions of network distance travelled in Luxembourg, in kilometres, for the activities **Work**, **Visit**, **Restaurant** and **Shopping**, according to the Luxmobil Survey, 2017.



Using this mobility data, and converting to Euclidean distance using a detour ratio formula [43], we are able to define, for several activities, a subweight that decreases the further away the location is from the agent's house. In the case of **Work**, the product of this and the other subweight yields a random choice function used to assign each agent with a place of work. For the activities **Shop**, **Restaurant** and **Visit**, the distance subweight alone determines the random choice function. Locations for some activities not specifically covered by the Luxmobil Survey, namely **Public Transport**, **Cinema or Theatre** and **Museum or Zoo**, are selected uniformly at random. Locations for activities **Schools**, **Medical**, **Worship** and **Indoor Sport**, are chosen based on household proximity. In the case of schools, there is a caveat that if a school is full then the next nearest school is selected instead, ensuring that classroom sizes are uniform across the region. Moreover we assume that children from the same household attend the same primary and secondary schools.

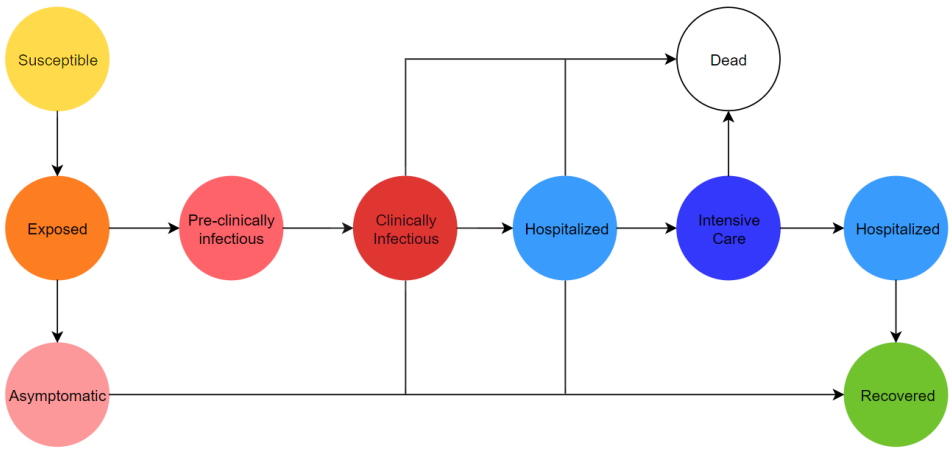
For large populations, it is too computationally costly to have the agents use the random choice functions during the simulation. Therefore, the choice functions are used beforehand to select, for each agent, a list of candidate locations of each type. Agents can then choose from this list, uniformly at random, when performing the relevant activity during the

simulation. Finally, we assume that agents only move to a new location when starting a new activity.

Disease and Transmission

Having modelled the population and its mixing patterns, we are then able to simulate an epidemic by attaching a disease and transmission model. Our disease model, which follows the SEIRD framework with additional compartments, is visualized below in Fig 7, where arrows illustrate possible state transitions.

Fig 7. Disease state diagram. The directed graph of health states.



The health states are characterized as follows:

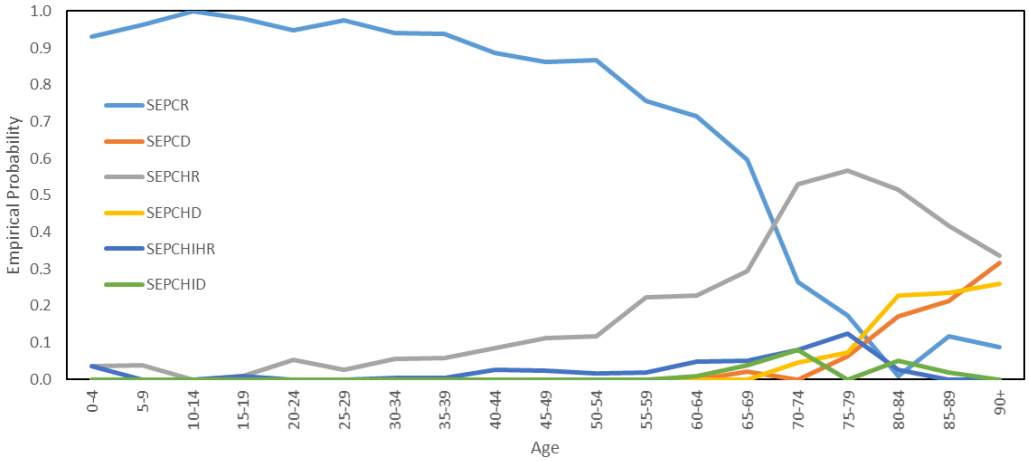
- **Susceptible**: The agent is able to catch the virus.
- **Exposed**: The agent has caught the virus but is not yet infectious.
- **Asymptomatic**: The agent is infectious but not symptomatic.
- **Pre-clinically Infectious**: The agent is infectious but not yet symptomatic.
- **Clinically Infectious**: The agent is infectious and symptomatic.
- **Hospitalized**: The agent should be in hospital but not intensive care.
- **Intensive Care**: The agent should be in intensive care.
- **Recovered**: The agent has survived the disease and is no longer infectious.
- **Dead**: The agent has died of the disease and should be moved to the cemetery.

Using the first letter in the names of each health state, we encode the possible trajectories through the above diagram as follows:

$$\mathbf{SEAR}, \mathbf{SEPCR}, \mathbf{SEPCD}, \mathbf{SEPCHR}, \mathbf{SEPCHD}, \mathbf{SEPCHIHR}, \mathbf{SEPCHID}$$

For example, the trajectory **SEPCD** describes an agent who having caught the virus passes through stages of pre-clinical and clinical infectiousness before dying from the disease outside of hospital. We assign to each agent a trajectory, with probabilities determined by age. For the model of Luxembourg, these probabilities are derived from COVID-19 surveillance data managed by the General Inspectorate of Social Security in Luxembourg, collected during the first wave of COVID-19 cases in 2020. The corresponding probability distributions for

Fig 8. Probabilities of symptomatic disease trajectories. The age-dependent distribution of symptomatic disease trajectories used in our model.



symptomatics are plotted in Fig 8. The probability that an agent follows the asymptomatic trajectory **SEAR** will be discussed later, in the subsection on model validation.

We do not assume limits on hospital and intensive care capacity, since we lack appropriate data. In particular, we have not tried to estimate the conditional probability of death given that the hospital or ICU is full.

We do not assume that time spent in a health state is geometrically distributed, as some other authors have done, for example [30]. Instead, we configure these durations according to the various distributions published in [44]. Denoting by $\Gamma(\alpha, \beta)$ the Gamma distribution with shape parameter α and scale parameter β and by $U(a, b)$ the uniform distribution on the integers $\{a, \ldots, b\}$, the distributions of time agents spent in each health state for each trajectory are then configured as in following diagram, in which the first and last states are ignored:

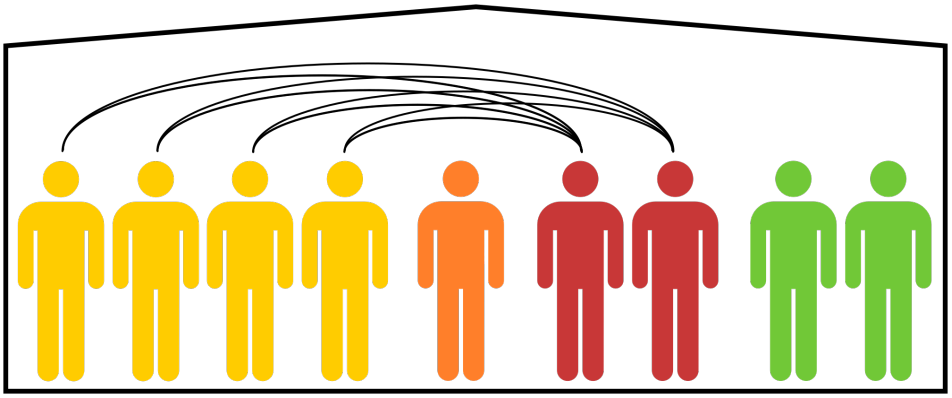
Table 5. Duration of time spent in each health state, ignoring the first and last in each sequence.

SEAR:	$\Gamma(4, 0.75) \to \Gamma(4, 1.250)$
SEPCR:	$\Gamma(4, 0.75) \to \Gamma(4, 0.525) \to \Gamma(4, 0.725)$
SEPCD:	$\Gamma(4, 0.75) \to \Gamma(4, 0.525) \to U(10, 15)$
SEPCHR:	$\Gamma(4, 0.75) \to \Gamma(4, 0.525) \to U(5, 9) \to U(8, 21)$
SEPCHD:	$\Gamma(4, 0.75) \to \Gamma(4, 0.525) \to U(5, 9) \to U(5, 16)$
SEPCHIHR:	$\Gamma(4, 0.75) \to \Gamma(4, 0.525) \to U(5, 9) \to U(3, 5) \to U(2, 12) \to U(2, 4)$
SEPCHID:	$\Gamma(4, 0.75) \to \Gamma(4, 0.525) \to U(5, 9) \to U(3, 5) \to U(7, 10)$

Our simulations begin with a number of agents infected with the virus. These agents are selected at random from among the resident population. Agents move between locations, and should a susceptible agent be in the same location as an infectious agent during the same 10 minute time interval, then with a certain probability a new infection will occur. More precisely, within each tick of the simulation clock, in each location, each symptomatic infectious agent transmits the virus to each susceptible agent with probability p. A susceptible agent is therefore infected if at least one infectious agent at the same location is successful in infecting them, the probability of this occurring being binomially distributed. For simplicity, we assume in the absence of personal protective measures that the

transmission probability is uniform across location types, except outdoors (which includes construction sites) and in the border countries where it is set to zero. An example of the transmission procedure is illustrated in Fig 9.

Fig 9. Transmission diagram. During each 10-minute interval, and in the absence of any interventions, an infectious agent (red) in the same location as a susceptible agent (yellow) infects them with probability p. Exposed agents (orange) are already infected but not infectious, while recovered agents (green) are assumed to be immune, having previously recovered from the disease. Some of these agents might be working in the given location, others only visiting temporarily.



We assume that asymptomatic and pre-clinically infectious agents are only 55% as infectious as the symptomatic infectious agents [45]. The number of new infections, at a given location during a given time interval, therefore follows a Poisson binomial distribution, an observation that allows for a certain amount of optimization.

Interventions

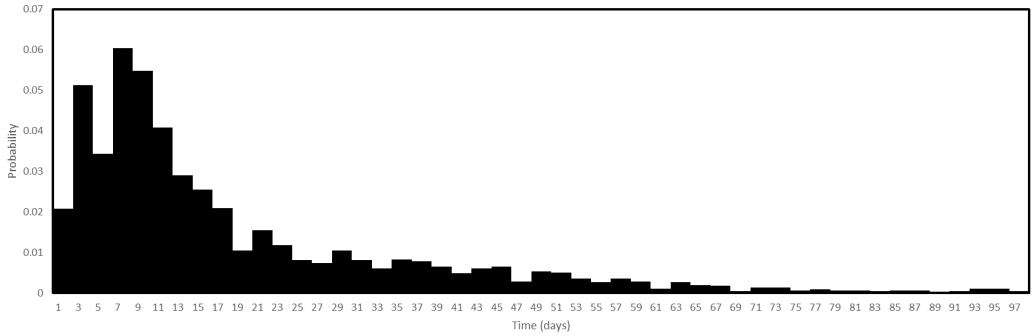
In this subsection, we describe briefly the various interventions featured in our model. Of course, we have not modelled all interventions, but only the most important ones. Firstly, agents in need of hospitalization are moved to a hospital for the duration of their required stay, and agents who have died are moved to the cemetery. We do not consider the impact of new anti-viral drugs or other treatments, instead assuming the hospital experience to remain constant. We assume that if an agent is directed by an intervention to behave in a certain way, for example to quarantine, then they will certainly do so, the only exceptions being face masks and vaccination. In the case of face masks, we assume that low face mask availability results in some agents not wearing the masks, while for vaccination we will consider the possibility that agents refuse the vaccine.

Testing

We split testing into a number of sub-processes. Firstly, there is a process representing large scale testing, which on particular dates distributes large numbers of test invitations. While this process is based on the system of large scale testing used in Luxembourg, where test invitations are not distributed randomly, we assume for simplicity that they are. We assume that there is a delay between agents receiving an invitation for large scale testing and the booking of the test. We assume this delay is distributed randomly as in Fig 10, the data for this having being collected by General Inspectorate of Social Security in Luxembourg in 2020.

Secondly, there is a process representing prescription testing, in which agents book a test one day after having developed symptoms. There is then a test booking system, which

Fig 10. Test booking delay. The probability distribution of time between an agent receiving an invitation for large scale testing and the agent booking the test.



handles these booking requests. We assume that if an agent has symptoms then the test takes place two days after the booking, while if an agent does not have symptoms then, given a lesser sense of urgency, it takes place four days after the booking. A laboratory process then performs the tests, returning results after two days with a 1% probability of a false negative. In addition, we assume that the laboratory is only able to perform a limited number of tests per day, the exact capacity being scenario-specific.

Contact Tracing

At the end of each day, an agent newly testing positive will have their contacts selected for testing and quarantine. Contacts are in this case defined to be those other agents who share a location with the given agent when performing the activities **House**, **Work** or **School**. These are the regular contacts of the agent, who the agent could be expected to identify through a manual search. Moreover, each day we limit the number of newly tested agents who are able to have their contacts traced, to model a limited scenario-specific capacity within the contact tracing system.

We also have a more sophisticated contact tracing system than this, which is more realistic and which operates over a rolling two day window of time, but at present this system is too computationally expensive to be implemented on large populations. We have also modelled the impact of a contact tracing app, namely Germany's Corona-Warn-App, but this is also too computationally expensive to simulate on large populations and therefore the subject of a future study on a smaller population.

Quarantining

Quarantining directs agents to perform all activities at their home location, for a default period of 14 days. Agents located in **Hospital** or the **Cemetery** are exempt from this directive. Should an agent obtain a negative test during their period of quarantine, then agents are able to leave quarantine restrictions after an additional 2 days.

Face Masks

According to the preprint [46], the effect of face masks is modelled by the mask transmission rate and mask absorption rate, which denote the proportion of viruses that are stopped by the mask during exhaling versus inhaling, respectively. We assume these proportions are equal, this value being denoted r. Then, given two agents in location l, one susceptible and one infectious, if p is the baseline transmission probability and q is the probability of an

individual wearing a mask, it follows that the modified transmission probability is

$$p((1-r)^2q^2 + 2q(1-r)(1-q) + (1-q)^2) = p(1-rq)^2$$

where moreover q can be expressed as the probability that an agent wears a mask given that the agent has a mask, multiplied by the probability that the agent has a mask. Following the authors of [46], we set $r = 0.7$.

Curfew

On 26th October 2020, an 11pm-6am curfew was imposed in Luxembourg. In our implementation, a curfew directs agents home between these hours unless they are located in **Hospital** or the **Cemetery**. While this implementation captures the essence of the curfew, it does not capture how a curfew in reality affects the behaviour of individuals earlier in the evening. On the one hand, individuals might cancel plans altogether to avoid breaking the curfew, while on the other they might simply perform the same activities but earlier. In this study, we do not consider such effects.

Location Closure

Location closures make locations of certain types inaccessible to agents between certain dates, with agents wishing to access such locations being instead directed home. Location closures can be used to model lockdowns, school closures and staggered closure or reopening of various sectors of the economy. In the special case of care home closures, we allow agents access if they work at the care home, meaning that in this case only visits are prohibited, while in the special case of shops we permit each shop to stay open with a certain probability, since in reality not all shops close during a lockdown. Typically shops selling food, drink or fuel will remain open.

Vaccination

In additional these non-pharmaceutical interventions, we also model vaccination. We assume a vaccine is administered in a two-dose format, with a fixed time between doses. We assume that the two doses successfully immunize the recipient with probabilities p_1 and p_2, respectively. The probability that the agent is protected against infection after the second dose is therefore $p_1 + (1-p_1)p_2$. For example, if this probability is set equal to 0.557, with p_1 set equal to 0.463, following [10], then we must set $p_2 = 0.175$. We assume that everyone who receives a first dose later receives a second dose. We assume that only a certain number of first doses of the vaccine can be administered each day and that agents are vaccinated in a particular order. The default scheme starts with care home residents and care home workers, followed by hospital workers, followed by everyone else, with each of these categories ordered by age, down to a minimum age of 16. We also assume that agents refuse vaccination with a certain probability, depending on their age. Such hesitancy is realized in our model by randomly selecting agents according and having these agents refuse the vaccine when it is offered to them during the simulation.

Our model of vaccination is relatively simple in that it assumes a successful dose completely protects against infection. In reality the situation is somewhat more complicated.

Model Evaluation

Crucial steps in the development of any computational model are model verification and model validation. We must have confidence that our model is internally and externally valid, that is, that is functions as it is supposed to and that it produces output relevant to the real

world. With this in mind, our code has been verified with a number of tests and will be made open source. With our code having passed these tests, we are confident that it functions correctly. It therefore remains to calibrate the model and address the uncertainties.

Our model is subject to certain limitations. It is not able to capture all the subtle complexities of population mixing and infectious disease dynamics. As mentioned in the introduction, this task is insurmountably difficult. Our objective was therefore to produce a reasonable approximation, capturing sufficiently many features that we are able to draw meaningful conclusions from our experiments. Nonetheless, a number of potentially important factors are not represented in our model. We do not model loss of immunity, nor the related impact of mutations to the virus, nor the introduction of new cases via long distance travel.

Incomplete or limited data is an obstacle that limits our understanding of the early states of the COVID-19 epidemic in Luxembourg. Very little testing took place, so the numbers against which we are calibrating are small. Nonetheless, our aim is to configure over the 122 day period from March 1st 2020 to 30th June 2020, covering the first wave of cases. Over a longer time horizon uncertainties would increase, due to factors not represented in our model becoming increasingly influential. For this reason, we will not make explicit quantitative predictions about the future, focussing instead on the relative impact of interventions.

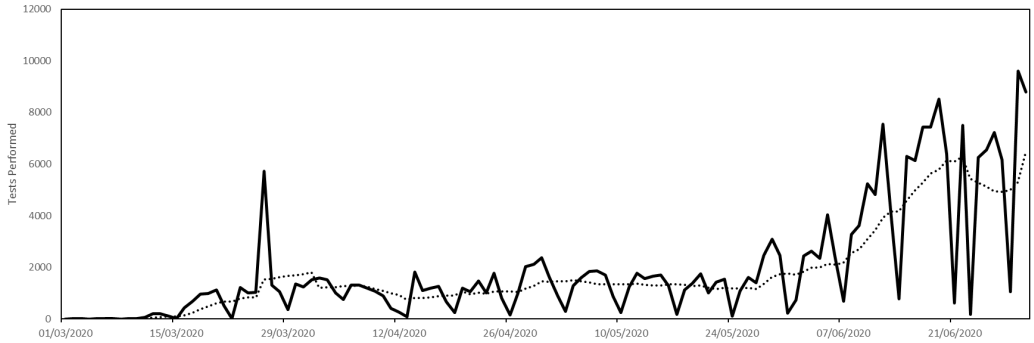
Parametrization of Interventions

The next step is to calibrate the interventions so as to reproduce the sequence of interventions that occurred in Luxembourg during the first four months of the epidemic. This is achieved using a scheduling system, which allows the interventions listed in the previous section to be enabled or disabled, and their parameters updated, on selected dates.

Testing

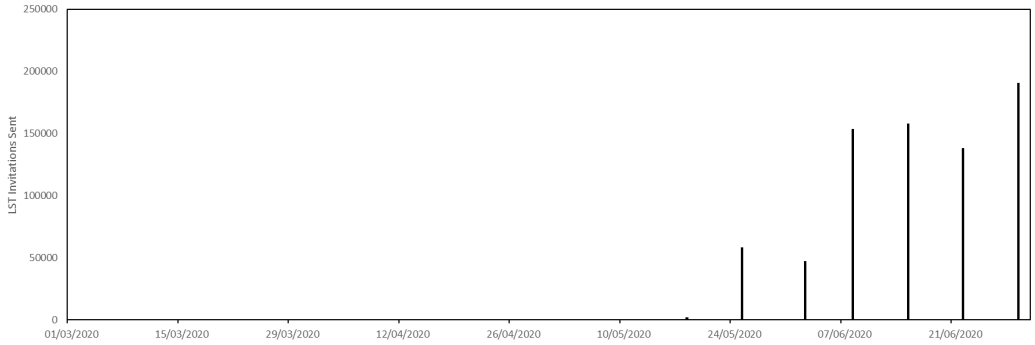
We assume that the capacity of the test laboratory is limited by the 7-day rolling average of the total number of tests recorded each day in Luxembourg. These daily totals, together with the trendline, are plotted below in Fig 11, between 1st March 2020 and 30th June 2020.

Fig 11. Parametrization of laboratory testing capacity. The capacity is given by the trend line.



The parametrization of large scale testing invitations is illustrated in Fig 12. This shows, approximately, the dates on which test invitations were sent in Luxembourg and the numbers of invitations sent on those dates. Recall that our agents respond to these invitations with a random delay.

Fig 12. Parametrization of large scale testing schedule.



Contact Tracing

We assume that contract tracing starts on 20th April 2020 with a capacity of 100. This means that as many as 100 agents testing positive each day can have their regular contacts traced. The capacity of the contract tracing system in Luxembourg subsequently increased, but not until much later.

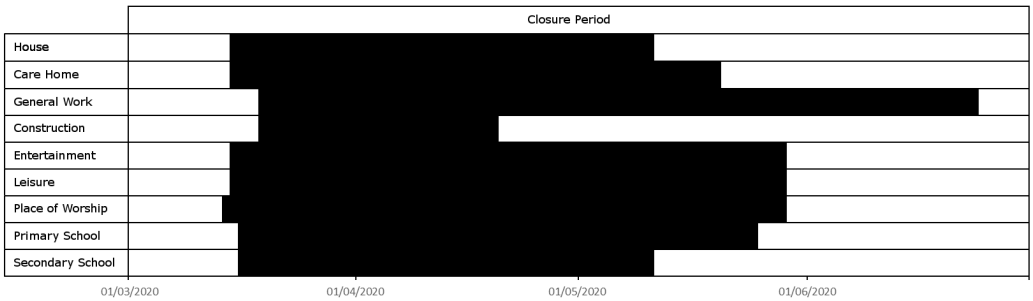
Face Masks

We assume that initially agents do not have access to face masks, the probability that they do increasing to 0.8 on 20th April 2020 and from 0.8 to 1.0 on 11th May 2020. We assume that the probability of a mask being worn, given that masks are available, depends of the type of location. We assume that this probability is 0.0 inside houses and cars and 1.0 inside public transport, shops, medical clinics, hotels, places of worship and museums and zoos. Elsewhere we assume that this probability is 0.2. These probabilities are only rough guesses. We assume moreover that face masks are always available in hospitals and medical clinics and that they are always worn.

Location Closure

Below, in Fig 13, we plot a time line indicating when locations of various types are assumed inaccessible during our validation simulations.

Fig 13. Schedule of location closures. A partial timeline of location closures in Luxembourg, from 1st March 2020 to 30th June 2020. The lockdown began on roughly 15th March 2020.



The category General Work appearing in Fig 13 refers to location types listed in Table 2,

except **Primary Schools**, **Secondary Schools**, **Construction** and **Entertainment**, which are listed separately. Leisure refers to locations of type **Indoor Sport**, **Cinema or Theatre**, **Museum or Zoo** and **Restaurants**. Closure of locations of type **House** or **Care Home** mean that agents are unable to access these locations while preforming the activity **Visit**.

In addition, we assume that 72% of shops close from 15th March 2020 to 11th May 2020, since according to [41] approximately this percentage of shops in Luxembourg do not sell either food, drink or fuel and were therefore subject to such restrictions.

Validation

With the interventions and other components configured, it remains to calibrate the transmission probability, initial infection count and the age-dependent probabilities of asymptomatic infection. This process involved a preliminary exploratory phase, followed by a systematic small grid search.

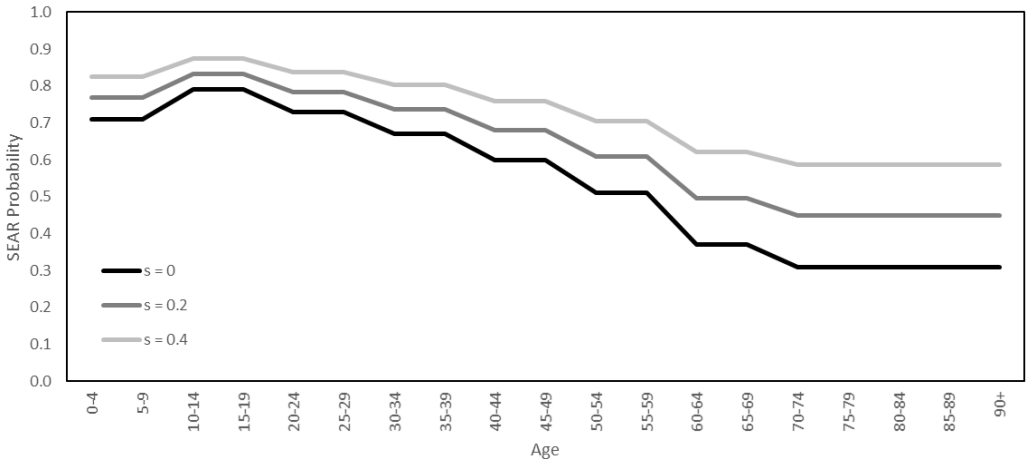
During the preliminary phase, we discovered that several features of our model needed developing or adjusting. For example, we discovered the importance of classroom structure in schools, in the absence of which large numbers of students gathered in schools would produce an unreasonably large number of infections. We also identified the role of care home parameters in determining overall deaths. In particular, we observed that care homes are, in most simulations, hotspots for both infection and death. We observed that a small number of large care homes results in considerably more deaths than a large number of small care homes. We were therefore careful to adjust the care home parameters to reflect the number and size of care homes in Luxembourg as best we could. We also adjusted care home closure restrictions to allow workers continued access to the care homes, since otherwise the extent to which care homes were isolated during lockdown was unrealistic.

Another key point relates to shops. In our model, we do not distinguish between different types of shop and originally configured the model to allow all shops to remain open during lockdown. However, this resulted in an unreasonably large number of infections occurring in shops during lockdown. We realized that we must try to more accurately reflect the fact that during the first lockdown in Luxembourg, shops selling food, drink or fuel were allowed to remain open while others had to close. We therefore decided to adjust the model so that only an appropriate percentage of shops remain open during lockdown periods. Finally, we observed that unreasonably large numbers of infections were occurring in construction sites, and we therefore set the transmission probability for these locations to be zero, as we had already done for the other outdoors location. Construction sites in Luxembourg were opened earlier than many other working locations as it was believed that the working environment of a construction site yields a relatively low transmission probability.

We then had to choose initial conditions. We decided to model the start of the outbreak by randomly selecting a number of residents as initial cases. Other approaches were possible, however for simplicity and clarity we chose to select randomly. We decided that the randomly selected initial cases should have their initial health state set equal to the first infectious state appearing in their assigned disease trajectory. This means, for example, that if an agent is selected to be one of the initial cases and has disease progression **SEPCR**, then their starting health state will be **Pre-clinically Infectious**. Setting the initial health states in this way appears preferable to the alternative in which the health states of the initial cases are set to **Exposed**, since it results in slightly more stable dynamics at start of the simulation. Although we are primarily interested in an interval of time starting on 1st March 2020, we ultimately decided to start our simulations a week earlier, on 23rd February. This gives the simulation an extra week in which to stabilize, before the start of interventions on 15th March. Of course, it will never been known exactly how many cases there were in Luxembourg on 23rd February, however, after some consultation, we settled on the number 320.

Infected agents are either symptomatic or asymptomatic. During initialization, we assign agents the asymptomatic progression **SEAR** with a probability that depends on their age. As a starting point for such probabilities, we take the numbers reported in [47]. Then, for each agent of age a, we have a probability $A(a)$ that the agent will be assigned **SEAR**. This is, however, a point of substantial uncertainty, since we do not know for sure what proportion of cases in Luxembourg were asymptomatic during the relevant time period. We therefore introduce a parameter $s \in [0, 1]$ to interpolate between these probabilities and the extreme case in which all agents are assigned **SEAR** with probability 1. Given an agent of age a, the probability that they are assigned **SEAR** is then $A(a)(1 - s) + s$, with the probability that they are assigned a particular one of the other sequences being $(1 - A(a))(1 - s)$ multiplied by the probability display in Fig 8. Using the parameter s we then have some control over the probabilities of hospitalization and death, without disrupting the distributions visualized in Fig 8, the data for which was carefully collected in Luxembourg by the General Inspectorate of Social Security. We plot the age-dependent asymptomatic probabilities in Fig 14 for the three values $s = 0$, $s = 0.2$ and $s = 0.4$. While $s = 0$ corresponds exactly to the probabilities quoted in [47], our simulations suggest these probabilities are too low, and therefore our calibration process will consider only $s = 0.2$ and $s = 0.4$.

Fig 14. Age-dependent asymptomatic probabilities. The probabilities of an agent being assigned the sequence **SEAR**. Being assigned this sequence means that if an agent is infected then they will, after exposure, become an asymptomatic case with a reduced transmission probably, before finally recovering.



Finally, we must set the transmission probability p. Recall that, given a 10 minute interval of time, and a pair of agents in the same location with one symptomatic and infectious and the other susceptible, p represents the probability of the infected agent successfully transmitting the virus to the susceptible agent. We consider the three values $p = 0.00015$, $p = 0.00025$ and $p = 0.00035$.

Table 6 shows the range of values of the pair (s, p) over which we now perform the small grid search. Preliminary investigations suggest that the pair best fitting clinical data sits somewhere in this range. A more sophisticated analysis is not possible at the present time, due to the computational burden of the agent-based model.

Due to computational and time constraints, we will preform all simulations at 0.25 scale. As explained earlier, this means that all relevant quantities are reduced to a quarter of their full size. Such quantities include population size, the number of locations and various

Table 6. Pairs (s, p) in the small grid search. Here s is the parameter that controls the probability of being asymptomatic while p is the transmission probability.

$s = 0.2, p = 0.00015$	$s = 0.4, p = 0.00015$
$s = 0.2, p = 0.00025$	$s = 0.4, p = 0.00025$
$s = 0.2, p = 0.00035$	$s = 0.4, p = 0.00035$

quantities relating to the interventions, such as testing and contact tracing capacity. We then rescale the output to full size, by multiplying by 4 all relevant quantitative output. This step is justified by the fact that increasing the scaling parameter does not push the model through thresholds but appears rather to yield a stable convergence, an expected result of the stochasticity of the model. At 0.25 scale our simulations each take around 5 hours.

We performed 10 simulations for each pair of parameter values appearing in Table 6. In Fig 15, we plot the corresponding numbers of resident deaths and hospitalizations for each simulation (grey and pink, respectively), together with their averages (solid black and red, respectively) and the numbers of deaths and hospitalizations recorded in Luxembourg over the same time period (dotted black and red, respectively). We calculate the number of hospitalizations in a simulation by adding the numbers of agents whose health state is either **Hospitalized** or **Intensive Care**.

We see that the pair $s = 0.4$, $p = 0.00035$ produces the closest fit. These are, therefore, the parameters that will be used in all subsequent simulations. The objective of this article is not to make precise quantitative predictions about the future, but rather to investigate the relative impact of interventions.

We observed that the total number of dead in a simulation is somewhat sensitive to the distribution of care homes, in the sense that the total number of dead increases by a non-trivial fraction for every care home hit by the epidemic. In additional to illustrating the sensitivity of our model with respect to the parameters s and p, Fig 15 also illustrates the extent to which the use of a pseudo-random number generator results in experimental uncertainty. Each random seed results in a slightly different environment with a slightly different epidemic. Since there has only been one COVID-19 pandemic affecting Luxembourg, it is difficult to know if is, in some sense, a typical one. Nonetheless, it is clear that the data collected in Luxembourg, displayed by the dotted curves in Fig 15, should serve as the calibration target.

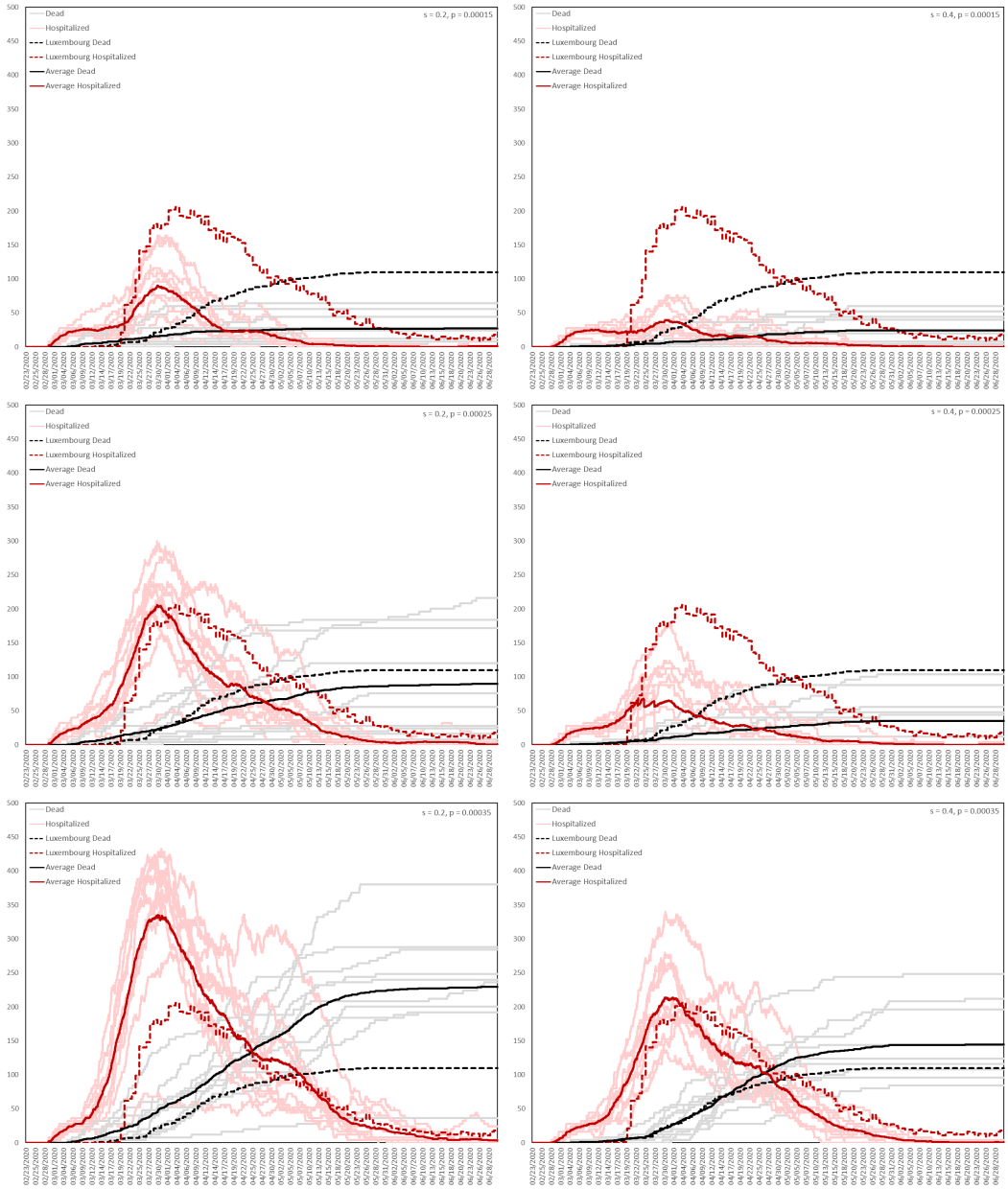
In Fig 16, we plot the average numbers, across the 10 simulations corresponding to the pair $s = 0.4$, $p = 0.00035$, of agents in the health states **Exposed**, **Asymptomatic**, **Pre-clinically Infectious**, **Clinically Infectious**, **Hospitalized**, **Intensive Care** and **Dead**.

In Figure we see how most new exposures occur during regular working hours on weekdays, with more towards the beginning of the week than the end. In particular, we clearly see the daily and weekly cycles resulting from the activity model and the use of time use data. In the next section, we will analyse the baseline scenario in more detail, using the extensive output of our model to look behind the scenes of the outbreak.

Results

We now present our main results, simulating with the parametrization $s = 0.4$ and $p = 0.00035$ established in the previous section. This parametrization was found by fitting the model to the epidemic in Luxembourg observed from March to July 2020, and therefore refers to the strains of the virus found in Luxembourg at that time. We consider a number of different scenarios, with ten simulations performed for each scenario, with each simulation running over the same 129 day interval but with a different random seed. We use the same set of ten random seeds for each scenario. For experiments involving interventions, we

Fig 15. Results of the grid search.

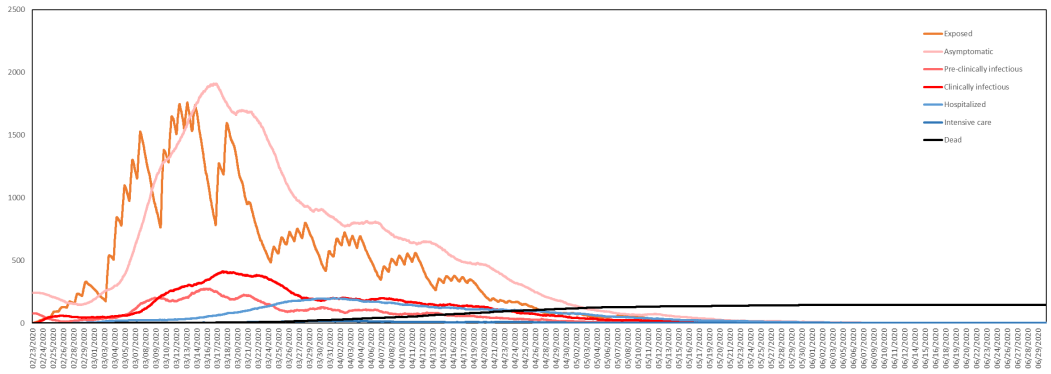


suppose that the interventions activate after exactly 3 weeks and continue until the end of the simulation. Before presenting the results of those experiments, we first establish the baseline scenario, in which no interventions are active. This scenario will act as the control, against which other scenarios can then be compared.

Baseline

In the baseline scenario, no interventions are active, meaning that agent behaviour does not change in response to the epidemic. In this case, we will compare the output of our agent-based model to that of the equation-based SEIR model. To make the comparison,

Fig 16. Numbers of resident agents in each health state. Averaged over 10 simulations with $s = 0.4$, $p = 0.00035$.



observe for the SEIR model that S stands for Susceptible and is equivalent to the health state **Susceptible**, E stands for Exposed and is equivalent to the health state **Exposed**, I stands for Infected and is equivalent to the set of health states **Asymptomatic**, **Pre-clinically Infectious**, **Clinically Infectious**, **Hospitalized** and **Intensive Care**, and R stands for Removed and is equivalent to the health states **Recovered** and **Dead**. For the agent-based model, we plot in Fig 17 the numbers Exposed and Infected in each of ten simulations of the baseline scenario, together with the corresponding means. Similarly, in Fig 18 we plot the numbers **Dead**.

Fig 17. Exposed and infected residents in the baseline scenario. Ten simulations of the baseline scenario with $s = 0.4$ and $p = 0.00035$. The dark curves plot the averages.

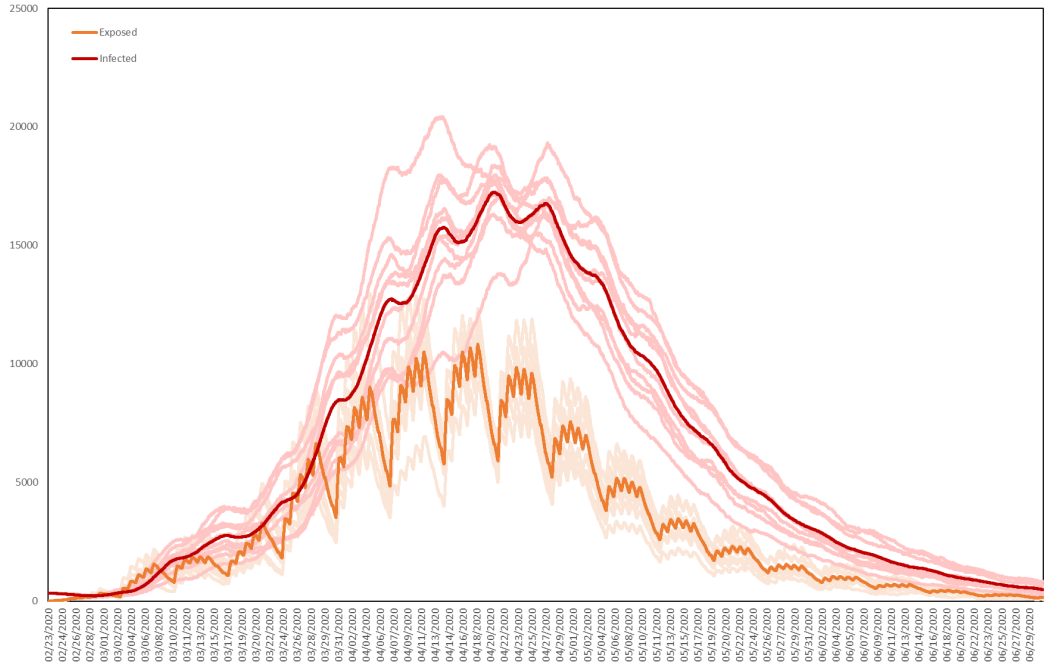
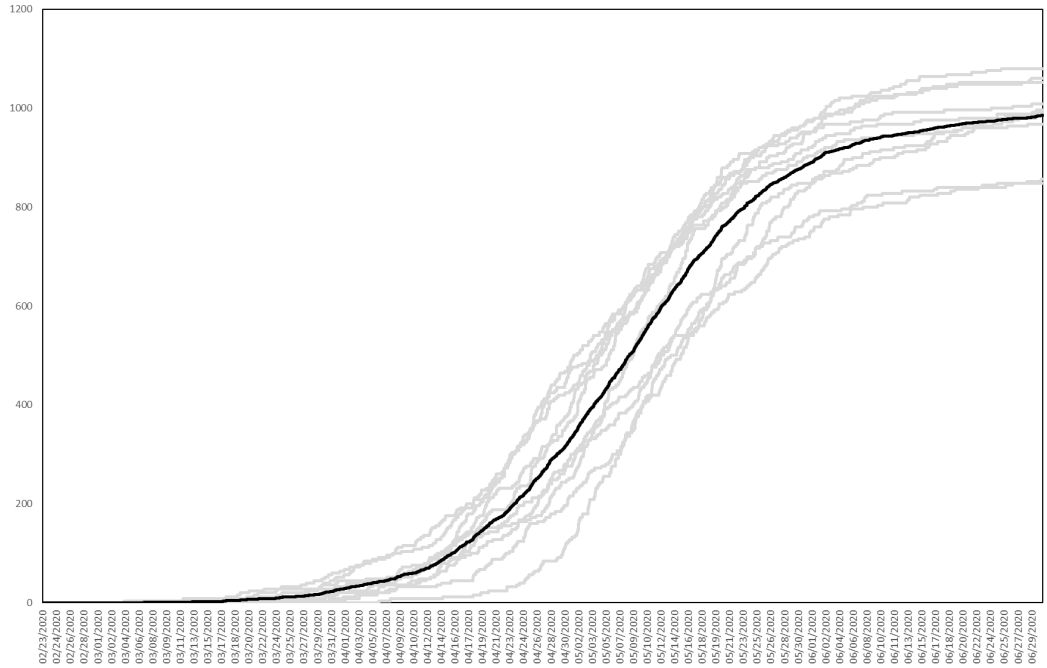


Fig 18. Dead residents in the baseline scenario. Ten simulations of the baseline scenario with $s = 0.4$ and $p = 0.00035$. The dark curve plots the average.



The baseline scenario results, after averaging over the 10 simulations, in approximately 985 deaths among the resident population of Luxembourg. This compares to a recorded 110 deaths over the same period, ending 30th June 2020, and 638 deaths by 28th February 2021, most of which occurred during a period of relaxed restrictions in November and December 2020.

Now consider the SEIR model given by system of ordinary differential equations

$$\begin{aligned}
\frac{d}{dt}S &= -\beta\frac{SI}{N} \\
\frac{d}{dt}E &= \beta\frac{SI}{N} - \alpha E \\
\frac{d}{dt}I &= \alpha E - \gamma I \\
\frac{d}{dt}R &= \gamma I
\end{aligned}$$

with initial conditions

$$(S(0), E(0), I(0), R(0)) = (625920, 0, 320, 0).$$

For such a model it is assumed that the incubation and infectious periods are exponentially distributed with mean durations α^{-1} and γ^{-1}, respectively. We set

$$\alpha^{-1} = 6.0512 \text{ days}, \quad \gamma^{-1} = 3.0020 \text{ days}$$

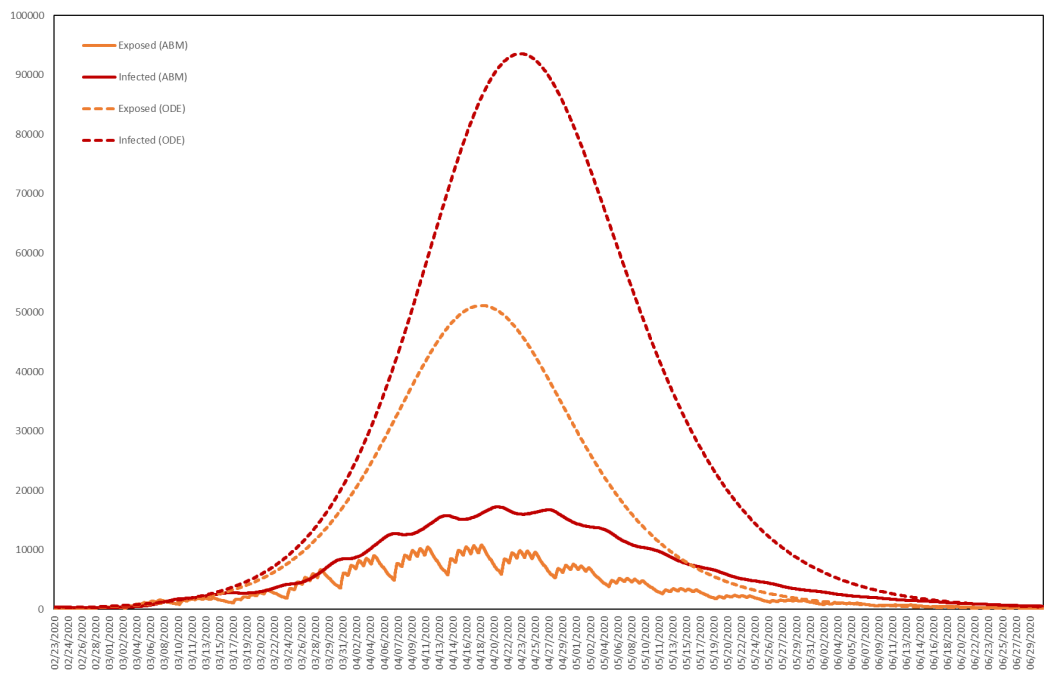
since these are the average incubation and infectious periods among residents in the agent-based model. The basic reproduction number of the SEIR model, denoted R_0, is given

by the ratio

$$R_0 = \frac{\beta}{\gamma}.$$

Choosing R_0 therefore determines β. To be precise, β is the average number of contacts per person per day, multiplied by the probability of disease transmission in a contact between a susceptible and an infectious individual. We observe that setting $R_0 = 2.45$, and therefore $\beta = 0.4049$ days^{-1}, yields a solution that peaks at roughly the same time as the epidemic produced by the agent-based model with $p = 0.00035$. For the two models, we plot the numbers Exposed and Infected in Fig 19. We observe that the agent-based model predicts an epidemic with considerably fewer cases than is predicted by the SEIR model. In particular, out of a total population of 625920, the SEIR model resulted in 554673 infections by the end of the 129 day period, representing 87% of the total population, whereas the agent-based model resulted, on average, in only 143162 infections, representing only 23% of the total population.

Fig 19. Numbers of resident agents exposed and infected. The average numbers exposed and infected in the agent-based model (ABM) with $p = 0.00035$ and in the SEIR ordinary differential equation (ODE) model with $R_0 = 2.45$.

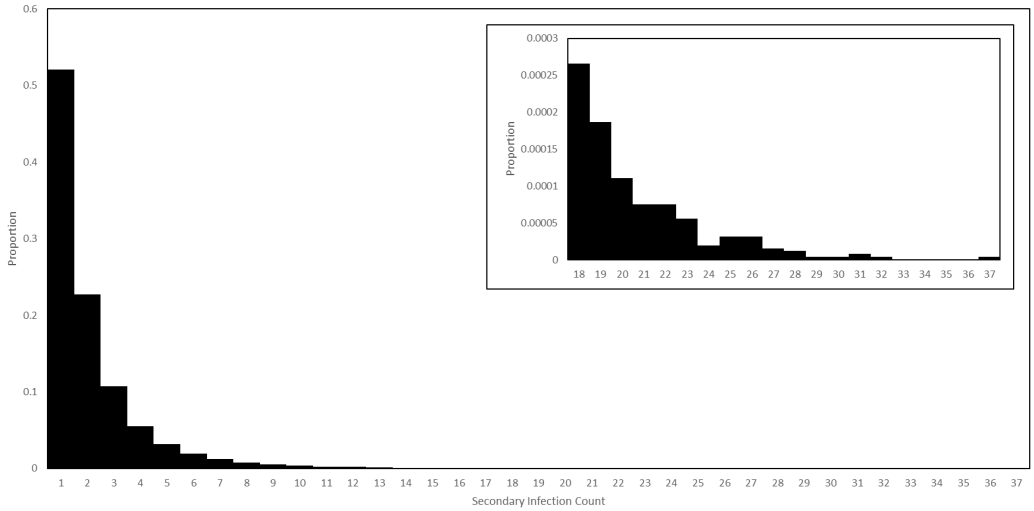


If alternatively β is configured so that the final state of the equation-based model agrees with that of the agent-based model, then the epidemic curves resulting from the equation-based model would be considerably wider and flatter than those of the agent-based model. Therefore, our agent-based model makes predictions that are quantitatively very different from those of the corresponding SEIR model, a result of the numerous heterogeneities present in our model. For example, clustering along the spatial dimensions limits the reach of infected individuals while the daily and weekly routines result in a fragmentation of the underlying contact network at night and during weekends. These features are not captured by the simple equation-based model. Our model suggests that if no action had been taken during the early stages of the pandemic then the death toll would

have been high, but not as high as predicted by some of the simpler epidemiological models.

Our model records not only the numbers of agents in each health state at each time, but also data on transmission events. After simulating the baseline scenario, we found that approximately 12% of all agents caused secondary infections. Among those who did, the probability distribution of the number of secondary infections is displayed in Fig 20. While the majority of agents who caused secondary infections caused only 1 or 2, a few caused as many as 37, with these agents therefore playing the role of super spreaders. The majority of infections caused by these super spreaders occurred at work. Among all agents, the average number of secondary infections was 0.27 while among only those who caused at least one secondary infection the average was 2.14.

Fig 20. Secondary infection counts in the baseline scenario. The probability distribution of secondary infections counts among those agents who caused at least one, in a simulation of the baseline scenario. The tail of the distribution is displayed in the inset plot, on a different scale.

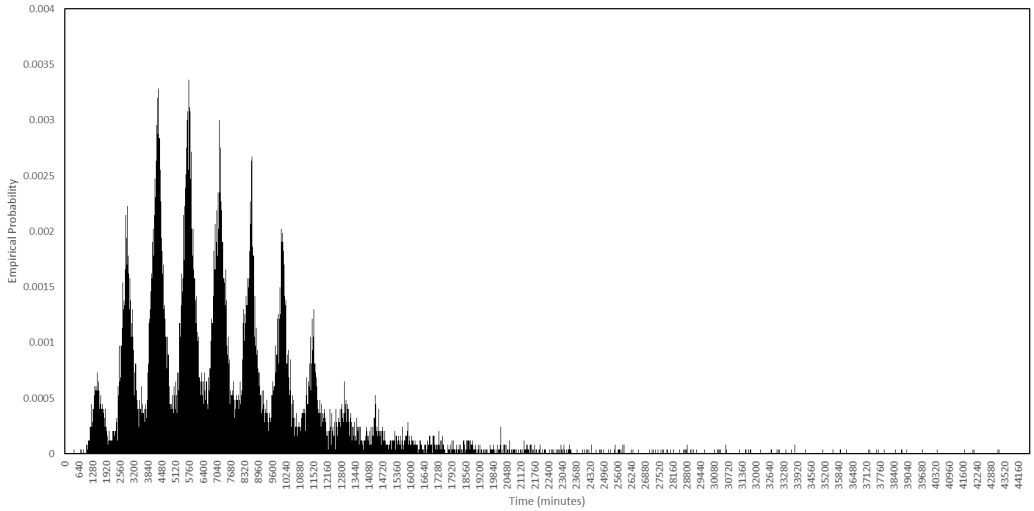


We can also get a handle on the serial interval. Simulating the baseline scenario, we found that a total of 53019 transmission events occurred, with 24683 of these agents going on to infect someone else. For each of these 24683 agents, we calculated the time between these agents catching the virus and the first time they transmitted it to someone else. The maximum such interval was 44680 minutes, or approximately 31 days, while the mean was 7154 minutes, or approximately 5 days. We plot the full probability distribution in Fig 21.

Notice in Fig 21 that the distribution is concentrated around multiples of 24 hours after infection, suggesting that in this baseline scenario agents are most likely to transmit the virus at the same time and type of place that they caught it. As regards deaths, in the baseline scenario we observed many deaths occurring in care homes, particularly towards the beginning of the epidemic.

Details such as these would be difficult to capture using an equation-based approach. That being said, the SEIR model referred to above is among the simplest of the compartmental models. A more complex mixing structure can be introduced with additional equations, resulting in output progressively closer to that of the agent-based model. Indeed, an agent-based model could always be formulated using a system of differential equations, however the number of equations would be enormous.

Fig 21. Serial intervals in the baseline scenario. Among those agents who passed on the virus, this is the probability distribution of the length of time between them being infected and them passing it on for the first time.



Individual Interventions

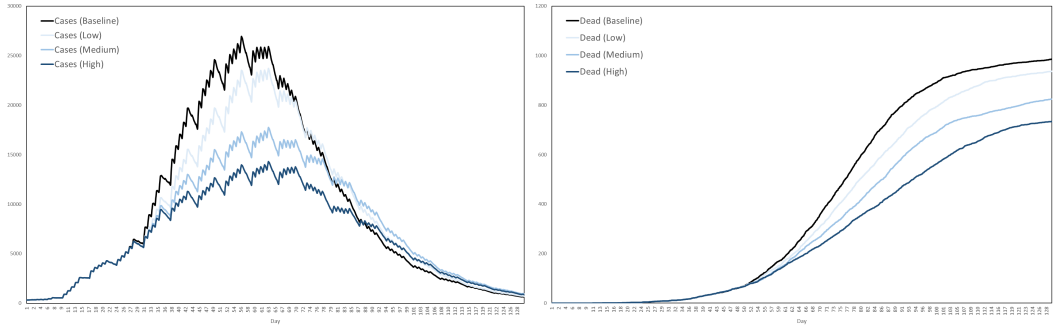
Now that we have established the baseline scenario, we can simulate interventions and assess their impact by comparison with the baseline. We start with those interventions that act on the level of the individual. In particular, we consider the collective impact of different levels of prescription testing, large scale testing and contact tracing, looking at low, medium and high intensities. In each of these three scenarios the test booking and laboratory systems are active, together with the quarantine intervention. We do not here consider the impact of face masks. Recalling that the model represents a total resident population of 625920, the four scenarios are as follows:

- **Baseline:** Agents behave as normal.
- **Low:** A daily testing capacity of 1000, with 800 invitations for large scale testing sent each day, and a contact tracing capacity of 100.
- **Medium:** A daily testing capacity of 5000, with 4000 invitations for large scale testing sent each day, and a contact tracing capacity of 300.
- **High:** A daily testing capacity of 10000, with 8000 invitations for large scale testing sent each day, and a contact tracing capacity of 500.

Recall that the contact tracing capacity refers to the number of agents each day who having tested positive can have their regular contacts traced for testing and quarantine. For each scenario we performed ten simulations, using the transmission and asymptomatic probabilities of the baseline scenario, but with the interventions activating after exactly 3 weeks. The average numbers of cases and dead in the three scenarios are plotted in Fig 22, together with the baseline for comparison, and where by a case we mean any agent either exposed or infected.

From the plot we see that while medium or high levels of testing and contact tracing have a significant impact on reducing cases, their impact on reducing deaths is considerably smaller. Indeed, testing and contact tracing systems, at least as implemented in our model,

Fig 22. Impact of testing and contact tracing. The impact of a combined testing and contact tracing strategy, at three levels of intensity, on cases and deaths.



do not specifically address the needs of vulnerable individuals. For example, while large numbers of deaths occur in care homes, a resident of a care home will only be targeted by contact tracing if another resident or worker at the same care home tests positive and is able to be processed by the contact tracing system. Even then, quarantining such vulnerable individuals at home does little to reduce their chance of catching the virus, since they would typically spend most of their time at home anyway. A more directed use of testing and contact tracing could improve the efficiency of these interventions.

Location Interventions

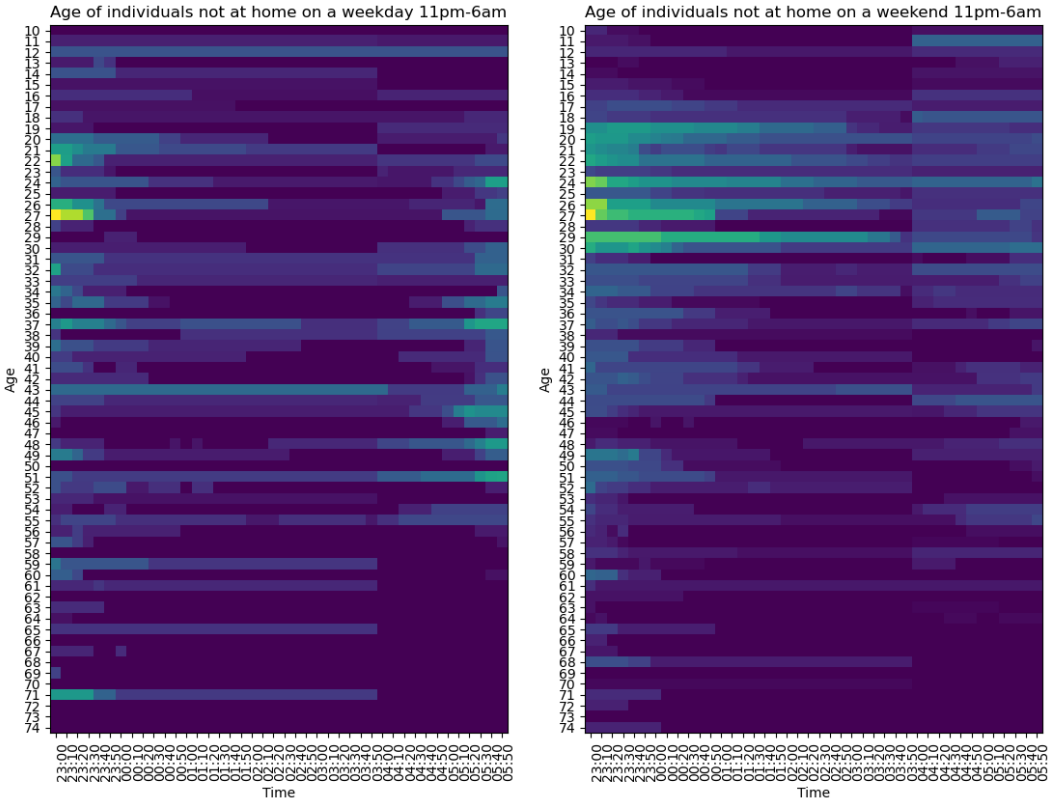
In this subsection, we look at the impact of interventions that act on locations, rather than agents. We compare the following four scenarios, the last of which is hypothetical:

- **Baseline:** Agents behave as normal.
- **Curfew:** Agents must stay at home between 11pm and 6am unless they are in hospital.
- **Lockdown:** Agents must stay at home unless their destination is a hospital, a care home at which they work or one of the 38% of shops selling food, drink or fuel.
- **Targeted Lockdown:** Agents belonging to households containing at least one person over the age of 65 must stay at home, unless their destination is a hospital, care home or one of the 38% of shops selling food, drink or fuel.

In each case, the interventions activate 3 weeks into the simulation and continue until the end. We expected the curfew to have only a small impact. Indeed, according to the Luxembourg time use data, aggregated and displayed in Fig 4, we see that during the relevant hours the vast majority of people are typically at home anyway. Moreover, Fig 23 shows that mainly young people are out between these hours, except on weekday mornings when small numbers of adults of a broader range of ages are not at home, mostly commuting or starting work.

We expected the lockdown to have the biggest impact in reducing cases and deaths, while we expected the targeted lockdown to retain a substantial impact on deaths, but less so on cases. The targeted lockdown focusses on those agents most at risk of death, while allowing large numbers of other agents to continue with work. In Fig 24 we illustrate how cases and deaths compare across the four scenarios, where for each scenario we plot the average output of ten simulations, using the disease and transmission parameters of the baseline scenario and the same set of random seeds used elsewhere. With respect to the

Fig 23. Age distribution of people not at home from 11pm to 6am, under normal circumstances, Luxembourg 2014. On the left, a weekday night and morning, for example Monday night and Tuesday morning. On the right, Saturday night and Sunday morning.



baseline scenario, the curfew, targeted lockdown and lockdown reduced deaths by 2.4%, 46.7% and 85.1%, respectively. In particular, the impact of the lockdown is enormous. It could, however, be argued that our estimate of the impact of the curfew is on the low side, since we do not consider the higher transmission levels present in bars and restaurants. However, the impact predicted by our simulations is so low that even with a higher local transmission probability the impact would still be relatively small. While the targeted lockdown has only a mild impact on total cases, its impact on deaths is much more substantial. The targeted lockdown could no doubt be improved with further refinements.

To assess the disruption caused by these interventions, in Fig 25 we plot the distribution of agents across location types over the 2 week period from day 15 to day 28, illustrating the impact of these interventions on these distributions. Observe that the lockdown has a dramatic impact on the numbers of agents working and going to school, while the impact of the targeted lockdown on the workforce is noticeable but much milder. The impact of the curfew is also visible but very small. Much of what is achieved by the full lockdown is also achieved by the targeted lockdown, but with a considerably smaller economic and social cost. Such targeted lockdowns could in reality represent a compromise between doing nothing and implementing a full lockdown.

Among agents who caused at least one secondary infection, while in the baseline scenario the mean number of secondary infections was 2.14, under the curfew the mean becomes 2.15, under the targeted lockdown 2.17 and under the full lockdown 2.42. Interestingly, therefore,

Fig 24. The impacts of the curfew, targeted lockdown and lockdown on cases and deaths. Averaged over 10 simulations and with the interventions activating on day 21.

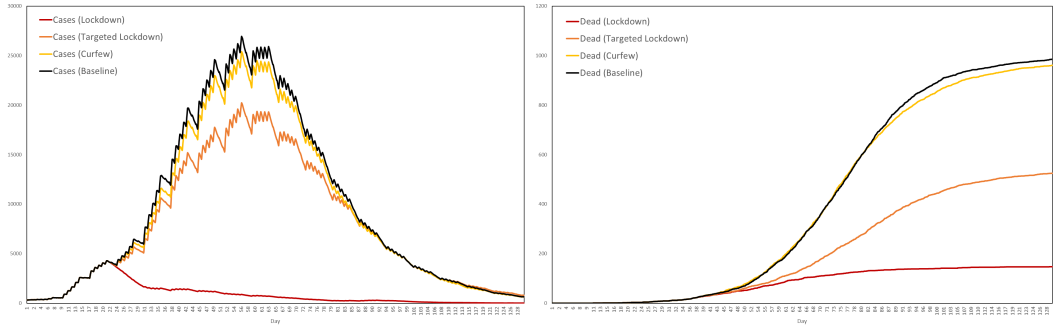
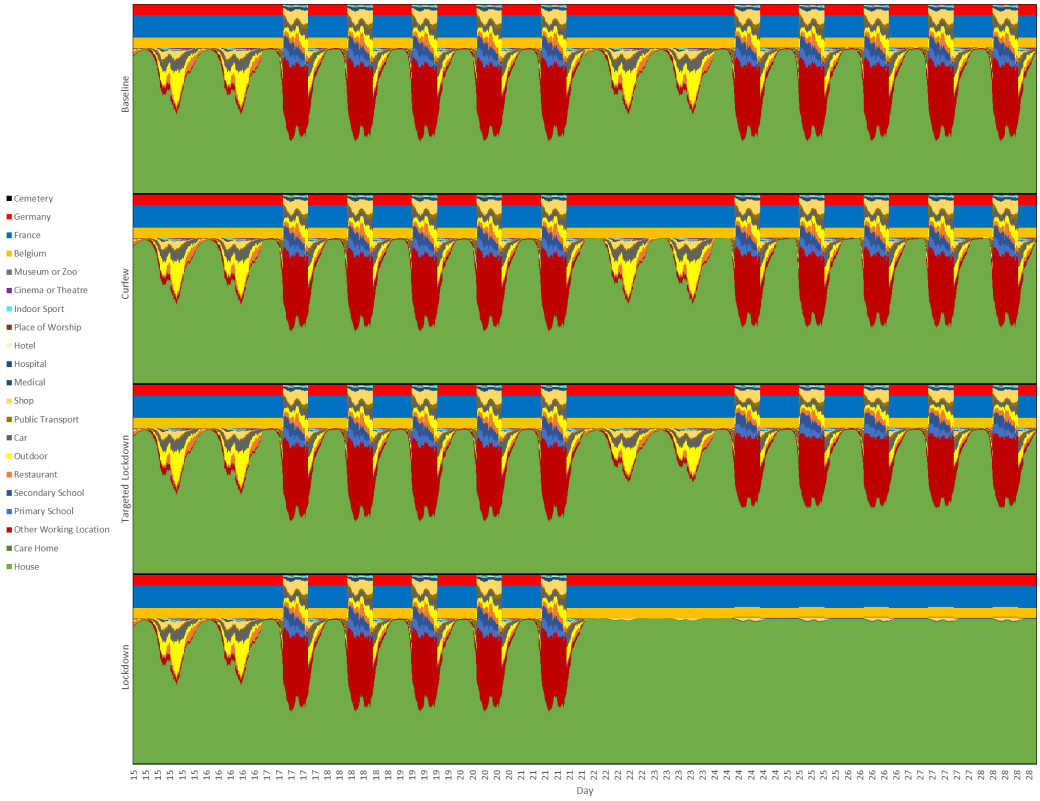


Fig 25. Occupancy of different types of location under the baseline, curfew, targeted lockdown and lockdown scenarios. The interventions activate 3 weeks into the simulation, and we therefore plot the 2 week period from day 15 to day 28. The data comes from four simulations, one for each scenario and all using the same random seed, to illustrate how the interventions impact the proportion of agents in locations of different types.



these interventions have the effect of increasing this average, even while reducing the total number of infections. Indeed, averaging over all agents, while in the baseline scenario the mean number of secondary infections was 0.27, under the curfew the mean becomes 0.26, under the targeted lockdown 0.23 and under the full lockdown 0.02. This highlights the fact

that when calculating averages, one must be careful with the choice of denominator.

Vaccination

We now consider several scenarios relating to vaccination. We investigate herd immunity, efficacy, capacity, hesitancy and strategy. For each of these five dimensions we construct several scenarios and perform simulations.

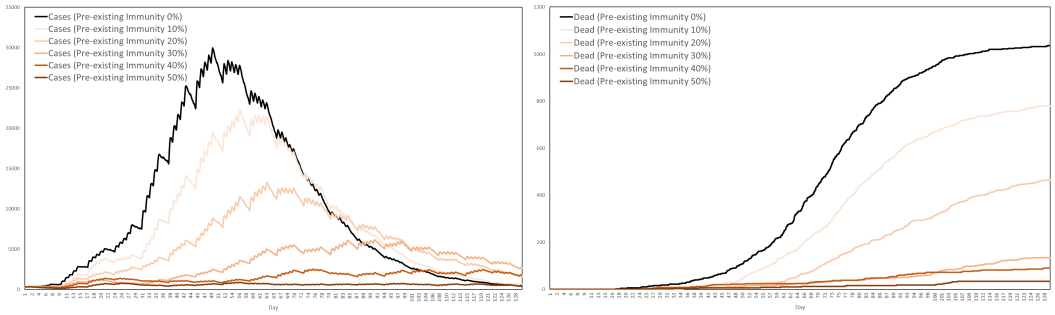
Herd Immunity

According to the World Health Organization [48]:

"'Herd immunity', also known as 'population immunity', is the indirect protection from an infectious disease that happens when a population is immune either through vaccination or immunity developed through previous infection."

Calculating the expression $1 - 1/R_0$ with $R_0 = 2.45$ implies a level of 59%. However, our model suggests that much lower levels of immunity provide the population with substantial protection against a future outbreak. Other studies have reached similar conclusions, for example [49]. We performed several simulations in which we assumed that a certain percentage of the population had pre-existing immunity. We selected these agents uniformly at random. In addition to two instances of the baseline scenario, where pre-existing immunity is 0%, we performed ten experiments in five pairs corresponding to levels of pre-exisiting immunity set at 10%, 20%, 30%, 40% and 50%. The simulations were otherwise parametrized as in the baseline scenario. For each pair, we averaged the two sets of outputs and the resulting numbers of cases and deaths are plotted in Fig 26.

Fig 26. Impact of pre-existing immunity on cases and deaths. The impact of 0% up to 50% of the population having pre-existing immunity.

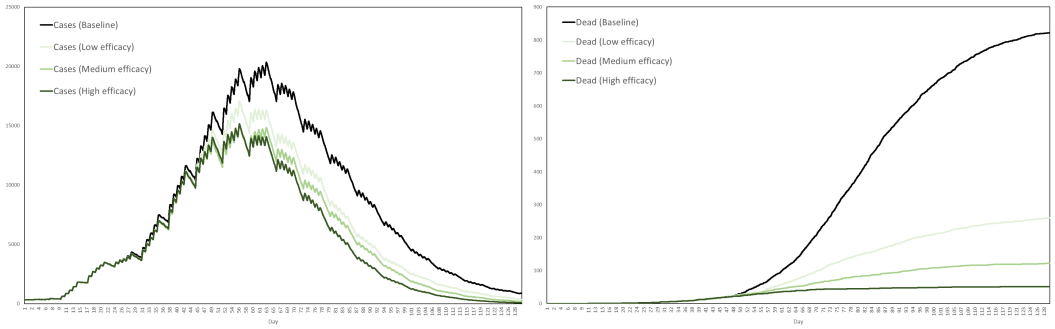


Recalling that the baseline scenario results, on average, in around 23% of all agents infected, much lower than the 87% predicted by the SEIR model, we see from Fig 26 that pre-existing immunity of only 30% already has a dramatic impact on reducing total cases and deaths. This suggests that relatively low levels of coverage can adequately protect a population from future outbreaks. A different situation is the one in which an epidemic is already under way, with vaccination occurring in response to it. This is the situation that will be considered next. Also, with a view towards COVID-19 vaccination programmes starting in early 2021, such as in Luxembourg where a significant proportion of the population is already immune having been previously exposed to the disease, we will assume for all subsequent experiments that 10% of the population have pre-existing immunity. It was therefore necessary to perform ten additional simulations of the baseline scenario, with 10% pre-existing immunity, with this new baseline being the one appearing in all subsequent figures.

Efficacy

We now consider the situation where vaccination begins 3 weeks into the epidemic. We assume that vaccines are distributed in a particular order. The order prioritizes care home residents and workers, followed by hospital workers, followed by all other agents down to a minimum age of 16. We will assume no vaccine hesitancy and that the number of first doses available each day is equivalent to 0.6% of the total population. In the Luxembourg implementation, this yields a constant daily capacity of 4864 first doses. We will assume that each vaccine is administered in two doses, precisely 3 weeks apart. We will investigate three vaccines, of low, medium and high efficacy, for which we assume that after the first dose these vaccines have efficacies 0.450, 0.675 and 0.900, respectively, with these efficacies increasing after the second dose to 0.55, 0.75 and 0.95, respectively. If p_1 and p_2 denote the probabilities that the first and second doses successfully protect against infection, then the values of the pair (p_1, p_2) corresponding to the low, medium and high efficacies are therefore $(0.450, 0.182)$, $(0.675, 0.231)$ and $(0.900, 0.500)$ since, according to our simple model of vaccination, if administered as a single dose the vaccines have efficacy p_1 while after two doses the efficacy increases to $p_1 + (1 - p_1)p_2$. For each of the three vaccines we performed ten simulations and averaged the resulting numbers of cases and deaths, plotting the results in Fig 27.

Fig 27. Impact of vaccine efficacy on cases and deaths. Low, medium and high efficacies correspond to probabilities of protection against infection after two doses of 0.55, 0.75 and 0.95, respectively.



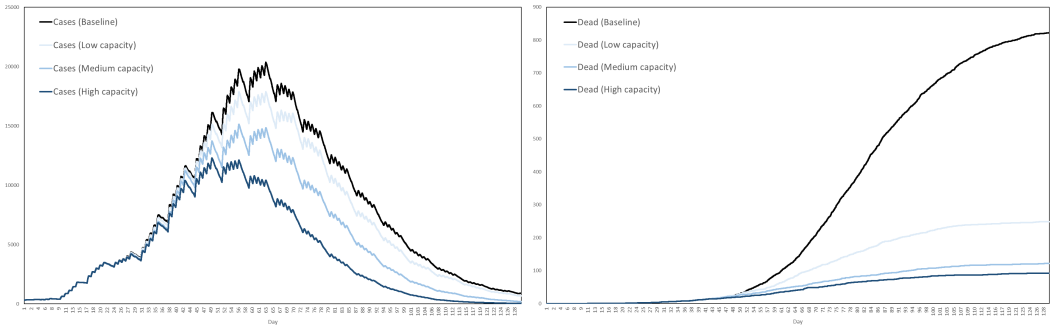
We see from Fig 27 that, vaccinating in the midst of an outbreak, the impact on cases is small, but the impact on deaths is high, even for the low efficacy vaccine. In particular, while on an individual basis the high efficacy vaccine is approximately 73% more likely to prevent infection, the high efficacy vaccine reduced deaths by only 38% more than the low efficacy vaccine, relative to the baseline scenario.

Capacity

We now look at the impact of lower and higher daily capacity. We take the medium efficacy vaccine, administer it according to the same strategy and assume no vaccine hesitancy. We set low, medium and high daily first dose availability equivalent to 0.2%, 0.6% and 1.0% of the total population, respectively, resulting in the Luxembourg implementation at daily first dose capacities of 1621, 4864 and 8107, respectively. Performing ten simulations for each scenario, we average cases and deaths and plot the results in Fig 28.

We see from Fig 28 that even a low daily first dose capacity has a significant impact on reducing deaths. As with efficacy, we see that the impact of capacity on cases is relatively small in comparison to the impact on deaths.

Fig 28. Impact of daily first dose capacity on cases and deaths. Low, medium and high capacities correspond to daily first dose availabilities equivalent to 0.2%, 0.6% and 1.0% of the total population, respectively.



Hesitancy

For the medium efficacy vaccine with the medium daily capacity, administered according to the same strategy, we now consider the impact of low, medium and high levels of vaccine hesitancy. In particular, we assume that with a certain probability agents refuse the vaccine when offered it. We assume that these probabilities are age dependent and that they remain constant throughout the simulation. An online survey conducted by science.lu in Luxembourg in December 2021 [50] suggested that vaccine hesitancy level were fairly high in Luxembourg, with only 55% of participants being likely or very likely to get a COVID-19 vaccine. Breaking down by age, the survey suggested that in the age group 13-34, only 48% were likely or very likely to get vaccinated, 57% in the age group 35-64 and 80% in the age group 65+.

For our simulations, we decompose according to the same age groups 16-34, 35-64 and 65+ with low, medium and high vaccine hesitancy levels for each age group parametrized as in Table 7. For example, for the low hesitancy scenario, we assume that agents aged 65+ refuse the vaccine with probability 0.10, while for the high hesitancy scenario agents aged 16-34 refuse the vaccine with probability 0.75, representing the two extremes. The medium scenario corresponds roughly to the data collected in the Luxembourg survey, while the probabilities for the low and high scenarios are obtained by interpolating half way between the medium scenario and the two extreme cases of zero and total hesitancy.

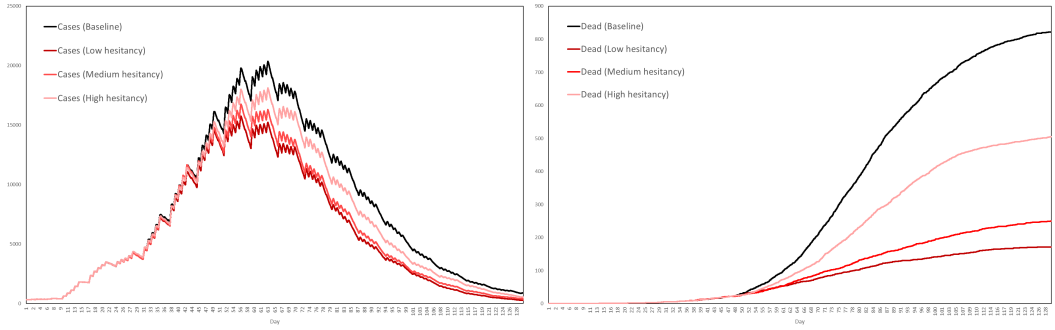
Table 7. Probabilities of vaccine refusal in low, medium and high hesitancy scenarios.

		Hesitancy		
		Low	Medium	High
Age	16-34	0.25	0.50	0.75
	35-64	0.20	0.40	0.70
	65+	0.10	0.20	0.60

Performing 10 simulations for each of the three scenarios, we plot the average numbers of cases and dead in Fig 29 together with those of the baseline.

We see from Fig 29 that high levels of hesitancy result in considerably more deaths. That being said, the levels of hesitancy corresponding to our high hesitancy scenario are in some sense very high. We assumed hesitancy levels to be constant throughout the simulation, although in reality hesitancy levels can change over time. For example, as more people are vaccinated, hesitancy levels might decrease as familiarity with the vaccine increases. On the other hand, as more people are vaccinated the likelihood of somebody experiencing unusual

Fig 29. Impact of vaccine hesitancy on cases and deaths. Low, medium and high hesitancy levels are assumed to be age-dependent.



side effects of the vaccine increases, with news of this potentially increasing hesitancy levels. While we have assumed a model of vaccine hesitancy that acts on the level of the individual, hesitancy can also manifest itself at a higher level, with policy makers themselves hesitant to implement the vaccine. Moreover, we have only simulated the use of a single vaccine. A future experiment would have several being administered simultaneously, starting on different dates, with different properties and with potentially different levels of hesitancy associated to them. Such considerations were beyond the scope of the present study.

Strategy

Finally, for the medium efficacy vaccine with medium daily capacity and no hesitancy, we now consider three different allocation strategies. The first, a simplified version of the priority scheme used in the other experiments, first allocates vaccines to the age group 65+ and then to the age group 16-64, proceeding in a random order within each group. The second distributes vaccines randomly to the entire age group 16+. The third starts with 16-64 and then moves onto 65+, the opposite of the first strategy. We expected that the strategy that prioritizes young people would lead to the biggest reduction in cases, while the strategy that prioritizes old people would lead to the biggest reduction in deaths. For each scenario, we performed 10 simulations and plot the average numbers of cases and dead in Fig 30, comparing to the baseline.

Fig 30. Impact of vaccine strategy on cases and deaths. We consider the three cases in which vaccines are distributed to old then young, young and old together, and finally young then old.

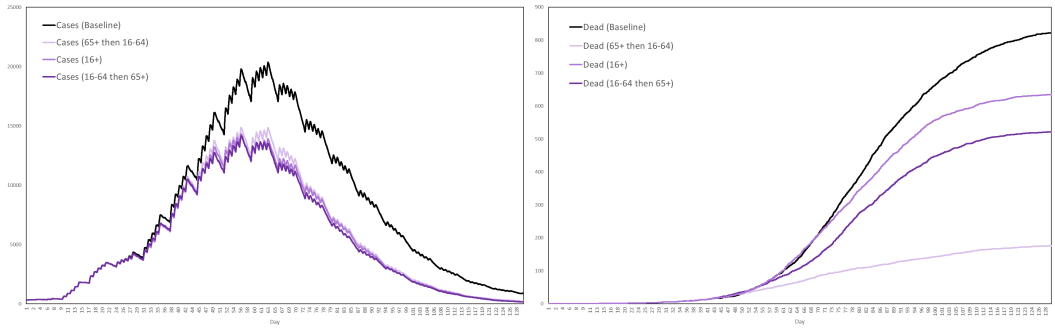


Fig 30 suggests that vaccinating younger people in an attempt to reduce transmission

and therefore deaths is not as effective as simply vaccinating the elderly first, since it leads to a much smaller reduction in deaths while resulting in only a very minor improvement in case numbers.

Conclusion

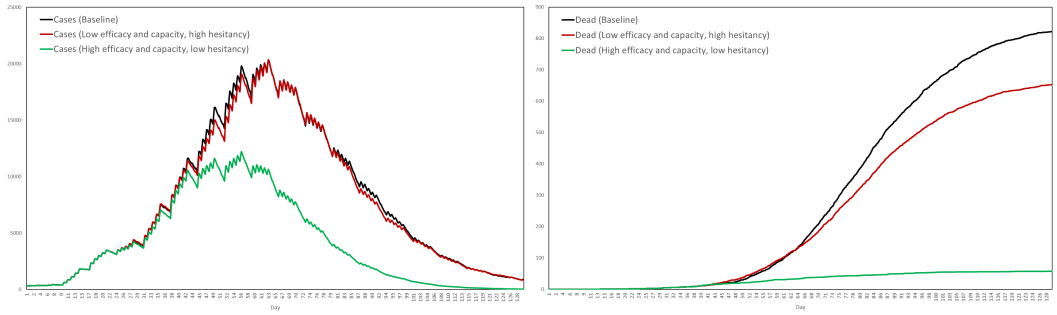
Based on the results presented and discussed in the previous section, we now draw several conclusions. We do so keeping in mind the limitations of our model, and the assumptions on which it is based. Our basic conclusions we list as follows:

- Our agent-based model predicts far fewer cases than the basic SEIR model. The latter assumes homogeneous mixing and therefore represents only an upper bound, with the heterogeneities captured by our model explaining the difference. Under generic assumptions, our model predicts only around 25% as many cases as the SEIR model.
- Testing and contract tracing reduce cases substantially, but are not very effective at reducing deaths.
- A full lockdown, although economically and socially very costly, dramatically reduces both cases and deaths. Alternatives to the full lockdown are also available, not as effective but less costly in terms of their economic disruption. The impact of an 11pm-6am curfew is relatively small.
- When vaccinating against a future outbreak, herd immunity is achieved at levels much lower than those predicted by the simple SEIR model. Under certain assumptions, our model predicts that substantial levels of protection are achieved with only 30% of the population immune.
- When vaccinating in midst of an outbreak, the task is more difficult. In this context, the impact of vaccination on total cases is reduced, however the impact on deaths remains high. In terms of total deaths, a low efficacy vaccine is almost as good as a high efficacy vaccine. As regards daily capacity, even with only a low number of doses administered each day the impact on deaths can be relatively high, so long as these doses are targeted at the most vulnerable individuals. High vaccine hesitancy results in considerably more deaths than would occur with low vaccine hesitancy and is the most serious challenge to a successful vaccination programme.

While in the previous section we considered independent variations in vaccine efficacy, daily capacity and hesitancy, in order to assess their individual impact, it is also worth considering the impact of a mixed variation of these parameters. In particular, we consider also the best and worse case scenarios, with the best case corresponding to high efficacy, high capacity and low hesitancy and the worse case corresponding to low efficacy, low capacity and high hesitancy. Performing ten simulations for each scenario, starting the vaccinations 3 weeks into the outbreak as before, we plot the average cases and deaths in Fig 31, as well as the averages for the baseline scenario in which no vaccination occurs. What we conclude from this is that in the worst case scenario the vaccination programme essentially fails, while in the best case scenario the vaccination programme is extremely successful at reducing deaths, the main factor here being the low vaccine hesitancy, with efficacy and capacity being nonetheless significant. Even in the best case scenario, when vaccinating in the midst of an explosive outbreak, there will still be large numbers of new cases many weeks after the start of the vaccination programme, however the peak will be smaller and occur sooner.

Let us finish with some final remarks about the limitations of the model and directions for future research. Firstly, confidence in our results would be further improved if we were to validate our model against other countries besides only Luxembourg. We fitted our

Fig 31. Best and worst case vaccination scenarios. The best case corresponds to high efficacy, high capacity and low hesitancy. The worse case corresponds to low efficacy, low capacity and high hesitancy. For each scenario we performed 10 simulations and plotted the mean numbers of cases and deaths, together with the baseline.



model to curves recorded in Luxembourg in 2020, but it is difficult to know how representative these curves are of a typical outbreak of COVID-19 in Luxembourg. Simulating the pandemic in other countries or regions would no doubt reveal more. Obtaining the data necessary to do this is a non-trivial task, and was therefore deemed beyond the scope of the present work. Doing so, however, we could then assess the impact of population distribution and also culture, with cultural differences realized through different distributions of daily and weekly routines. Suitable time use data has been collected by a number of countries, including all member states of the European Union, the United Kingdom and the United States. Moreover, while we used these activity routines to construct a model of mobility, we should note that other sets of mobility data could be used instead and implemented directly inside the location choice functions. Generally speaking, our model could be improved were we to find a way to capture more of the correlations in behaviour between familiar individuals and the way that agent behaviour changes automatically in response to an event such as the COVID-19 pandemic.

New strains of COVID-19 present new challenges, but we have not simulated the impact of different strains, nor attempted to model competition between strains. We speculate that social distancing and testing exert an evolutionary pressure on the virus that increases the reward for any mutation that makes the virus more transmissible or less easily detected. The simulation of such a competitive system is an objective for future research, with uncertainties surrounding the strains a major reason why we have not made any concrete predictions about the future. Moreover, since we are a part of the system that we are trying to model, and therefore not independent from it, to a certain extent we would be doomed to fail anyway.

Nonetheless, our results reinforce the widely held view that vaccination is the most effective intervention against COVID-19. Lockdowns are extremely costly, both socially and economically, with other non-pharmaceutical interventions having only a limited impact. Vaccination represents the best hope we have to free ourselves from this deadly virus, the implication being that a positive and progressive approach to vaccination is essential.

Appendix

In this appendix we describe our model according to the ODD protocol. The generic parametrization of submodels is described in the methods section, while parametrizations specific to particular scenarios are described in the model evaluation and results sections, so we will not repeat those details here.

Purpose and Patterns

The purpose of this model is to explore the impact of interventions, in particular vaccination, on cases and deaths due to COVID-19. The intention is to help decision makers understand the relative strengths of interventions when used in combination with one another. The model has been configured to represent Luxembourg and therefore the patterns that the model has been assessed against were observed in Luxembourg during the first few months of the pandemic. This includes, in particular, the drops in cases and deaths seen after multiple strict measures were introduced in March 2020.

Entities, State Variables and Scales

The basic entities in our model are agents and locations:

- **Agents**: The agents represent individuals living or working in a given region. They are assigned age, health state, nationality and lists of locations at which they are able to perform various activities. In addition to these state variables, agents are assigned a behavioural routine describing which activities they perform and when they perform them, the time resolution being 10 minutes.
- **Locations**: The locations represent places where the agents can perform activities. Locations are assigned spatial coordinates and a type, with the possible types of location listed in Tables 1 and 2. Coordinates are assigned by sampling population grid data. The grid data has a resolution of 1 kilometre, with the coordinates sampled from this in WGS84 format at a resolution of 1 meter.

Process overview and scheduling

Our model is configured to run for a fixed number of iterations, with each iteration representing a 10 minute interval of time. During each iteration of this main loop, interventions are updated according to a schedule and internal message and telemetry buses are notified of world updates occurring since the last tick of the simulation clock. Components are notified of the new time, to which they might then respond. For example, it might be time for the movement model to request that an agent moves to visit a care home, but a lockdown intervention listening to such requests overrides the request, requesting instead that the agent returns home. The disease model loops over all locations and determines if any new infections take place, requesting health state updates if so. Once these requests have been resolved via the message bus, world updates are enacted and the simulation moves onto the next tick, with the simulation finally ending after the predetermined number of ticks.

Design concepts

Basic principles

The model implements a conventional compartmental disease description within the bottom-up approach of an agent-based model. The compartmental disease model is familiar and easily understood, while the agent-based approach provides a more detailed and flexible model of social interactions than can be achieved with the equation-based approach. In particular, the agent-based approach allows for an intuitive and realistic implementation of interventions. This is much more difficult to achieve at the aggregated level of a small system of differential equations. Another basic principle of our model, and one that influenced its design, is adaptability. Our model is built on a modular framework, with components communicating with one another via a message bus, having the advantage that

components can be easily added or replaced, transforming the model with ease to describe new regions, diseases or interventions.

Emergence

Emergence is a concept that sits at the heart of our approach to modelling. Behaviour is described on an individual basis, with routines sampled from a pool of over 2000 possibilities, yielding an extremely complex system of collation and movement. By simulating an infectious disease spreading within such as system, we observe the resulting epidemic as an emergent phenomenon. The set of all possible sequences of interactions between agents is extremely large, with certain sequences having a dramatic effect on the total number of deaths. A chain of interactions ending in a care home might, for example, be of this type.

Adaptation

Agents in our model do not adapt their routines willingly. If a routine is disrupted, it is because an intervention has over-ridden it. In other words, in the absence of interventions agents will behave as if everything was normal. Adaptive routines, based on learning objectives and prediction might enhance the model, but would be very difficult to parametrize.

Sensing

Components, such as the disease model and the interventions, collect data on the world and respond accordingly. This is achieved via the message bus, the system of information exchange to which components can subscribe and publish events. The stream of communications between the components results from the interactions of the agents and the disease model, and therefore represents an emergent collection of events.

Interaction

If two agents occupy the same location for the same 10 minute interval, then it is assumed that an interaction occurs that with some probability results in disease transmission. The nature of this interaction is assumed to be uniform across all location types. While in reality location type or activity might be important factors in determining the probability of transmission, in the absence of relevant data we make no such hypotheses, assuming uniformity of interactions for simplicity.

Stochasticity

Stochasticity is used throughout our model, during both initialization and simulation. The world is procedurally generated, with locations distributed and populated by sampling probability distributions. For each agent, movement is determined by the random selection of locations belonging to certain lists, while disease transmission is also the result of random, binomial, sampling. Via repeated sampling, stochasticity washes away outliers that may arise form a particular configuration. Much care was taken to ensure that our experiments can be repeated and the results replicated, by keeping track of the random seeds used by the psuedo-random number generators appearing in our code.

Collectives

Agents routines are sampled from a finite pool, and therefore there are agents who behave similarly. In addition, agents living in the same house will tend to visit similar nearby locations. These correlations, however, are not the result of emergent collective behaviour, being instead consequences of the configuration process.

Observation

A telemetry system observes and collects data on each simulation. The systems consists of reporters, each of which looks at a different aspect of the simulation. The reporters are as follows:

- **Health State Counts**: This reporter records, at each tick, the numbers of resident agents in each health state.
- **Activity Counts**: This reporter records, at each tick, the numbers of resident agents performing each activity.
- **Location Type Counts**: This reporter records, at each tick, the numbers of resident agents in each type of location.
- **Testing Counts**: This reporter records, at the end of each day, how many tests and positive tests were performed that day, distinguishing between residents and non-residents.
- **Testing Events**: Each time a test occurs, this reporter records the date and time, the test result, the agent's age and health state, the residency status of the agent and the coordinates of their home.
- **Quarantine Counts**: This reporter records, at the end of each day, how many agents are in quarantine. It also calculates the average age of these agents and breaks them down by health state.
- **Exposure Events**: Each time a new infection occurs, this reporter records the date and time, the type of location and who infected who. It records the ages of the two agents and which activities they were each performing at the time.
- **Death Events**: Each time a agent dies, this reporter records the date and time, their age, whether they live in a house or a care home, and information on their place of work.
- **Vaccination Events**: Each time a first dose of a vaccine is administered, this reporter records information about the agent in question, including age, health state and household composition.
- **Secondary Infection Counts**: Throughout the simulation, this reporter counts how many infections each agents causes. At the end of this simulation, it then calculates a histogram, illustrating the distribution of secondary infection counts, from which a mean can then be derived.

Initialization

Initialization begins by creating a map of the region. This includes a model of population density. This is followed by the creation of the world, based on the map, which involves distributing locations and populating them with agents. Having created the world and a clock object, to keep track of time, the remaining components of the model are then initialized. These components are the disease model, the activity model, the movement model and the interventions. For example, during the initialization phase it is determined who will die if infected, who will work night shifts and who will refuse a vaccine. With the initialization phase completed, the simulation begins.

More precisely, having constructed the map object, the world is built in the following order:

- Resident agents are created and assigned an age and nationality.
- Locations are created and assigned coordinates.
- Resident agents are assigned homes, with the most elderly being assigned care homes. The mechanism by which agents are grouped into households reflects an expected distribution of ages derived from STATEC census data.
- Neighbouring countries and populations of cross-border workers are created, with these adults being assigned an age and nationality. These agents will perform all activities other than work in their home country.
- Agents are assigned a place of work, to which they will move if performing the work activity.
- Resident agents are assigned a number of homes, shops and restaurants that they may visit during the simulation. These are sampled in terms of the distance to the agent's home.
- Resident agents are assigned a number of cinemas or theatres and museums or zoos that they may visit during the simulation. These are sampled randomly from all such locations in the region.
- Resident agents are assigned primary and secondary schools, to which they move if performing the school activity, and also a medical clinic, place of worship and indoor sports center. Locations of these types are assigned based on proximity, unless the location has already been assigned its fair share of agents, in which case the next nearest available location is chosen. This is to avoid overcrowding, ensuring that a balanced number of agents visit these locations.
- Resident agents are assigned cars, with households being given one car each.

The procedure described above therefore assigns to each agent and for each activity a list of locations from which the agent can randomly choose when performing that activity during the simulation. It therefore remains to initialize the aforementioned components:

- The disease model assigns to each agent a disease profile, describing the trajectory of health states through which the agent will pass should they be infected, and an associated list of durations, indicating how long the agent will spend in those states. A number of resident agents are randomly infected and their health state set accordingly. These will be the initial cases that get the epidemic started.
- The activity model assigns to each agent a weekly routine, sampled from over 2000 such routines with a 10 minute resolution. These routines are built from data collected by STATEC and distinguish between weekdays and weekends. The initial activity of each agent is set accordingly, together with an initial location.
- The contact tracing system initializes, determining for each agent a list of regular contacts. This is a list of other agents who live, work or go to school with the given agent. These contacts will be subject to quarantine and testing should the agent test positive during the simulation.
- The test laboratory, test booking and prescription testing systems initialize, collecting information on health states from the disease model. The large scale testing intervention assigns to each agent a period of time that the agent will wait before responding to a test invitation, should such an invitation be received.

- The location closure interventions initialize. In the case of care homes, this involves creating lists of agents working in each care home.
- The vaccination intervention constructs an ordered list of agents to be vaccinated during the simulation. During this initialization phase, it is determined which agents will refuse vaccination and therefore be omitted from the list.
- The curfew and hospitalization interventions initialize, although do not require any detailed procedures.

Input data

The model uses several sources of input data. Some are used to configure time varying processes. The activity routines, assigned during the initialization phase, describe the sequence of activities performed by each agent, constructed from time use data obtained by STATEC [16]. The numbers of trains, buses and trams operating through the day is variable and configured within the movement model, using data obtained by Mobilitéit [19]. Moreover, each intervention operates according to a schedule, consisting of dates on which to enable or disable the intervention or on which to update the values of certain parameters. This uses COVID-19 surveillance data, derived from a national database managed by the General Inspectorate of Social Security in Luxembourg.

Submodels

The model includes a number of submodels, the most important of which are listed as follows (some of which are described in more detail in the methods section):

- **Map Factory**: The map factory compiles population grid data to produce a distribution from which location coordinates can be sampled. It includes a subsystem that refines this distribution via linear interpolation, improving the resolution beyond the default 1 kilometre.
- **World Factory**: The world factory creates agents and locations and for each agent assigns for each activity a list of locations to which the agent can move during the simulation. These lists are determined beforehand since otherwise the computational cost would be too great when dealing with large populations.
- **Message Bus**: The message bus allows components to communicate through a shared set of interfaces. Communications are either requests or notifications. Requests are made to, for example, begin a new activity, move to a new location or book a test. Other components might cancel these requests, issuing their own requests in response. Once such disagreements are all resolved, with the state of the world updated accordingly, notifications are sent through the message bus informing components of these changes. The message bus was implemented to account for the fact that interventions must interact with one another when several are simultaneously active. There is also a telemetry system, operating on the same principles as the message bus, that collects and saves data from the simulation for analysis.
- **Clock**: The clock keeps track of the time, both in terms of ticks and in ISO 8601 format. In the default configuration, a tick of the clock represents an interval of length 10 minutes. Components keep track of the current time via the message bus.
- **Deferred Event Pool**: This object stores events due to occur at a later time in the simulation. For example, once an agent has received their first dose of the vaccine, the administration of the second dose is added to the deferred event pool, as an event due

to take place on a particular date several weeks after the first. On that date, the system will then issue a request to the message bus, triggering the vaccination system to actually perform the second dose.

- **Scheduler**: The scheduler is a system that parses input data on dates and parameter values to produce for each intervention an implementation that varies over time. This is necessary since model validation requires the reproduction of measures introduced in Luxembourg during the first months of the COVID-19 pandemic, with various quantities associated to these measures being variable. For example, daily testing was variable, while places of work, schools and other locations were closed on certain dates and reopened on others.

- **Disease Model**: The disease model was designed according to the familiar compartmental framework but in such as way that avoids geometrically distributed periods of time spent in each health state. Rather than using stochasticity on each tick to decide who moves into the next health state, disease progression for each agent is determined during initialization, allowing for a richer and more realistic variety of patterns. On each tick, the transmission model loops through all locations and determines who, if anyone, is to be newly exposed. More precisely, it counts how many infectious agents are in a given location, distinguishing between symptomatics and asymptomatics, and loops through the susceptible agents in that location, sampling binomial distributions to determine if those agents are to be infected. If infections occur, the system then decides, via random selection, who exactly caused each infection. The algorithm is so ordered to optimize runtime, with the identification of the infecting agent needed only for telemetry and testing purposes.

- **Activity Model**: The activity model was designed to give agents interesting, varied and realistic daily and weekly routines. Assigning these routines during initialization lowers the computational cost, versus a system that for each agent chooses activities stochastically. Such a system, based on Markov chains, was previously implemented in our model, but was replaced due to the computational burden and the fact that, after repeated testing, did not appear to be sufficiently advantageous.

- **Movement Model**: As stated above, the world factory determines lists of locations that agents might visit. In the event that that an agent starts a new activity, the movement model simply selects a location at random from the appropriate list.

- **Hospitalization**: This hospitalization intervention moves agents to hospital if their health transitions to a state demanding hospitalization. This intervention is relatively simple, and does feature hospital or ICU capacity, a feature was omitted due to uncertainties in how to parametrize such a system. The hospitalization intervention also takes care of agents who have died, moving them to the cemetery. Dead agents are moved to the cemetery to avoid them being erroneously counted as inhabiting other locations.

- **Test Booking System**: The testing system is quite large and therefore divided into several subsystems. The test booking system handles requests to get tested. The test events themselves are scheduled via the deferred event pool.

- **Testing Laboratory**: The laboratory system performs the tests, handling deferred test event requests, published through the message bus. If the daily limit of tests has been reached, then subsequent tests that day are simply not performed. If a test takes place, the result of the test is published to the message bus for other components to see.

- **Prescription Testing**: Tests are booked in our model for one of two reasons. The first is that an agent has developed symptoms, detected if a health state transition has been published to the message bus in which an agent is symptomatic having not previously been so.

- **Large Scale Testing**: The other circumstance in which an agent books a test is after they have been invited to do so by the large scale testing system. For simplicity, our implementation of this system distributed tests at random. Once an invitation has been received, agents respond by booking a test after a delay. It was important to include this delay since data collected in Luxembourg shows that this period of time is often quite substantial.

- **Contact Tracing**: The contact tracing system responds to newly published test results. If the result of a test is positive, the system issues test booking and quarantine requests to regular contacts of the relevant agent. More detailed implementations of contact tracing are possible, and were tested, however the system described seems to provide a good balance between realism and runtime when simulating very large numbers of individuals.

- **Quarantine**: The quarantine model holds a list of agents who are subject to quarantine restrictions. Agents are added to the list if a quarantine request is made, which occurs either via the contact tracing system or if an agent tests positive. Agents are removed form the list once their period of quarantine is over, a period which can be reduced if the agent should happen to get a negative test result. The quarantine system interacts with the movement model by overriding requests to leave home if an agent is in the quarantine list. In particular, we assume that agents completely adhere to the quarantine rules.

- **Location Closure**: The location closure system interacts with the movement model in a way that is similar to the quarantine system. If an agent requests to move to a location that is, as determined by the scheduler, currently off limits, then that request is denied with the agent being sent home instead. The only exception here is care homes, with agents still permitted access to a care home if they happen to work there.

- **Curfew**: The curfew system is very similar to the location closure system, acting on list of disallowed locations which in this case includes everything except hospitals and the cemetery. The difference is that the curfew, on days when it is enabled, is only active between certain hours.

- **Vaccination**: This system incorporates several features that were deemed to be of most importance. One such feature is vaccine hesitancy, representing the fact that not everybody wants to get vaccinated. The probability of refusal is determined by age. Another is variable efficacy, representing an increased efficacy after the second dose of a two-dose vaccine. The system also features a priority list, representing systems in which limited supplies of vaccine are allocated to certain individuals before others, according to age, residency or place of work. Vaccinations run on a daily cycle, with a deferred event pool and the message bus being used to schedule the administration of second doses. The vaccination model was designed to encompass such a level of detail since an examination of the impact of vaccination was one of the main objectives of the study.

Acknowledgements

This project was funded by the COVID-19 Fast-Track program of the Fonds National de la Recherche Luxembourg. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The reference number for this project is:

COVID-19/2020-2/14858807/ABMLUX

The authors would also like to thank Dr. Mikołaj J. Kasprzak for his help in drafting the grant proposal, communications with STATEC, preparing the timeline used for Fig 13 and for pointing the authors towards several useful references.

Author Contributions

Conceptualization:	James Thompson
Data Curation:	James Thompson, Stephen Wattam
Formal Analysis:	James Thompson
Funding Acquisition:	James Thompson
Investigation:	James Thompson, Stephen Wattam
Methodology:	James Thompson, Stephen Wattam
Project Administration:	James Thompson
Resources:	Stephen Wattam
Software:	Stephen Wattam, James Thompson
Supervision:	James Thompson
Validation:	James Thompson, Stephen Wattam
Visualization:	James Thompson, Stephen Wattam
Writing – Original Draft Preparation:	James Thompson
Writing – Review & Editing:	James Thompson, Stephen Wattam

References

1. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond A. 1927;115(772):700–721. doi:10.1098/rspa.1927.0118.

2. Gardner M. Mathematical Games - The Fantastic Combinations of John Conway's New Solitaire Game 'Life'. Scientific American. 1970;223(4):120–123.

3. Grimm V, Railsback SF. Individual-based Modeling and Ecology:. Stu - student edition ed. Princeton University Press; 2005. Available from: `http://www.jstor.org/stable/j.ctt5hhnk8`.

4. Epstein JM, Axtell R. Growing Artificial Societies: Social Science from the Bottom Up. vol. 1. 1st ed. The MIT Press; 1996. Available from: `https://EconPapers.repec.org/RePEc:mtp:titles:0262550253`.

5. Hommes C, LeBaron B. Introduction to the Handbook of Computational Economics, Volume 4, Heterogeneous Agent Modeling. In: Hommes C, LeBaron B, editors. Handbook of Computational Economics. vol. 4 of Handbook of Computational Economics. Elsevier; 2018. p. xxi–xxix. Available from: `https://www.sciencedirect.com/science/article/pii/S1574002118300236`.

6. Keeling MJ, Eames KTD. Networks and epidemic models. J R Soc Interface. 2005;2:295–307.

7. World Health Organization Regional Office for the Western Pacific. Calibrating long-term non-pharmaceutical interventions for COVID-19 : principles and facilitation tools. Manila : WHO Regional Office for the Western Pacific; 2020-05-16.

8. Anderson RM, May RM. Vaccination and herd immunity to infectious diseases. Nature. 1985;318(6044):323–329. doi:10.1038/318323a0.

9. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. New England Journal of Medicine. 2020;383(20):1920–1931. doi:10.1056/nejmoa2022483.

10. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. The Lancet. 2020; p. 99–111. doi:10.1016/S0140-6736(20)32661-1.

11. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine. 2020;383(27):2603–2615. doi:10.1056/nejmoa2034577.

12. Logunov DY, Dolzhikova IV, Shcheblyakov DV, Tukhvatulin AI, Zubkova OV, Dzharullaeva AS, et al. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. The Lancet. 2021;doi:10.1016/S0140-6736(21)00234-8.

13. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nature Medicine. 2020;doi:10.1038/s41591-020-1124-9.

14. Murphy J, Vallières F, Bentall RP, Shevlin M, McBride O, Hartman TK, et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. Nature Communications. 2021;12(1):29. doi:10.1038/s41467-020-20226-9.

15. Neumann-Böhme S, Varghese NE, Sabat I, Barros PP, Brouwer W, van Exel J, et al. Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. The European Journal of Health Economics. 2020;21(7):977–982. doi:10.1007/s10198-020-01208-6.

16. STATEC - Statistics Portal of the Grand Duchy of Luxembourg;. `https://statistiques.public.lu`.

17. IGSS - General Inspectorate of Social Security of the Grand Duchy of Luxembourg;. `https://igss.gouvernement.lu/en.html`.

18. OpenStreetMap Foundation - Open Street Map;. `https://www.openstreetmap.org`.

19. Ministry of Mobility and Public Works of the Grand Duchy of Luxembourg, Department of Mobility and Transport - Mobilitéit;. `https://www.mobiliteit.lu/en/`.

20. Ministry of Mobility and Public Works of the Grand Duchy of Luxembourg, Department of Mobility and Transport `https://data.public.lu/en/datasets/horaires-et-arrets-des-transport-publics-netex/`.

21. Eurostat - GEOSTAT initiative;. `https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat`.

22. Kemp F, Proverbio D, Aalto A, Mombaerts L, d'Hérouël AF, Husch A, et al. Stages of COVID-19 pandemic and paths to herd immunity by vaccination: dynamical model comparing Austria, Luxembourg and Sweden. medRxiv. 2021;doi:10.1101/2020.12.31.20249088.

23. Burzynski M, Machado J, Aalto A, Beine M, Haas T, Kemp F, et al. COVID-19 Crisis Management in Luxembourg: Insights from an Epidemionomic Approach; 2020. Available from: `https://ssrn.com/abstract=3651024`.

24. Ghamizi S, Rwemalika R, Cordy M, Veiber L, Bissyandé TF, Papadakis M, et al. In: Data-Driven Simulation and Optimization for Covid-19 Exit Strategies. New York, NY, USA: Association for Computing Machinery; 2020. p. 3434–3442.

25. Tang B, Liu P, Yang J, Wu J, Yanni X, Tang S. The challenges of the coming mass vaccination and exit strategy in prevention and control of COVID-19, a modelling study. medRxiv. 2020;doi:10.1101/2020.12.18.20248478.

26. Volpert V, Banerjee M, Sharma S. Epidemic Progression and Vaccination in a Heterogeneous Population. Application to the Covid-19 epidemic. medRxiv. 2020;doi:10.1101/2020.12.06.20244731.

27. Jentsch P, Anand M, Bauch CT. Prioritising COVID-19 vaccination in changing social and epidemiological landscapes. medRxiv. 2020;doi:10.1101/2020.09.25.20201889.

28. Piraveenan M, Sawleshwarkar S, Walsh M, Zablotska I, Bhattacharyya S, Farooqui HH, et al. Optimal governance and implementation of vaccination programs to contain the COVID-19 pandemic. arXiv. 2020; p. 1–15.

29. Grauer J, Löwen H, Liebchen B. Strategic spatiotemporal vaccine distribution increases the survival rate in an infectious disease like Covid-19. Scientific Reports. 2020;10(1):1–10. doi:10.1038/s41598-020-78447-3.

30. Hinch R, Probert WJM, Nurtay A, Kendall M, Wymatt C, Hall M, et al. OpenABM-Covid19 - an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. medRxiv. 2020;doi:10.1101/2020.09.16.20195925.

31. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Hart G, Rosenfeld K, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. medRxiv. 2020;doi:10.1101/2020.05.10.20097469.

32. Gaudou B, Huynh NQ, Philippon D, Brugière A, Chapuis K, Taillandier P, et al. COMOKIT: A Modeling Kit to Understand, Analyze, and Compare the Impacts of Mitigation Policies Against the COVID-19 Epidemic at the Scale of a City. Frontiers in Public Health. 2020;8:587. doi:10.3389/fpubh.2020.563247.

33. Wilmes P, Zimmer J, Schulz J, Glod F, Veiber L, Mombaerts L, et al. SARS-CoV-2 transmission risk from asymptomatic carriers: Results from a mass screening programme in Luxembourg. The Lancet Regional Health - Europe. 2021;4:100056. doi:https://doi.org/10.1016/j.lanepe.2021.100056.

34. Vilches TN, Zhang K, Van Exan R, Langley JM, Moghadas SM. Projecting the impact of a two-dose COVID-19 vaccination campaign in Ontario, Canada. medRxiv. 2020;doi:10.1101/2020.12.10.20246827.

35. Moghadas SM, Vilches TN, Zhang K, Wells CR, Shoukat A, Singer BH, et al. The impact of vaccination on COVID-19 outbreaks in the United States. medRxiv. 2021;doi:10.1101/2020.11.27.20240051.

36. Asgary A, Najafabadi MM, Karsseboom R, Wu J. A Drive-through Simulation Tool for Mass Vaccination during COVID-19 Pandemic. Healthcare. 2020;8(4):469. doi:10.3390/healthcare8040469.

37. Goldenbogen B, Adler SO, Bodeit O, Wodke JA, Escalera-Fanjul X, Korman A, et al. Optimality in COVID-19 vaccination strategies determined by heterogeneity in human-human interaction networks. medRxiv. 2020;doi:10.1101/2020.12.16.20248301.

38. Patel MD, Rosenstrom E, Ivy JS, Mayorga ME, Keskinocak P, Boyce RM, et al. The Joint Impact of COVID-19 Vaccination and Non-Pharmaceutical Interventions on Infections, Hospitalizations, and Mortality: An Agent-Based Simulation. medRxiv. 2021;doi:10.1101/2020.12.30.20248888.

39. Parry HR, Bithell M. In: Heppenstall AJ, Crooks AT, See LM, Batty M, editors. Large Scale Agent-Based Modelling: A Review and Guidelines for Model Scaling. Dordrecht: Springer Netherlands; 2012. p. 271–308. Available from: `https://doi.org/10.1007/978-90-481-8927-4_14`.

40. Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: A review and first update. Ecological Modelling. 2010;221(23):2760 – 2768. doi:https://doi.org/10.1016/j.ecolmodel.2010.08.019.

41. STATEC - Répertoire des Entreprises Luxembourgoises 2019;. https://statistiques.public.lu/fr/publications/series/repertoire-entreprises/2019/repertoire-2019/index.html.

42. Eurostat - HETUS - The Harmonised European Time Use Surveys;. `https://ec.europa.eu/eurostat/web/time-use-surveys`.

43. Yang H, Ke J, Ye J. A universal distribution law of network detour ratios. Transportation Research Part C: Emerging Technologies. 2018;96:22–37. doi:https://doi.org/10.1016/j.trc.2018.09.012.

44. Rees EM, Nightingale ES, Jafari Y, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. BMC Med. 2020;18(270).

45. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368(6490):489–493. doi:10.1126/science.abb3221.abb3221.

46. Kai D, Goldstein GP, Morgunov A, Nangalia V, Rotkirch A. Universal Masking is Urgent in the COVID-19 Pandemic: SEIR and Agent Based Models, Empirical Validation, Policy Recommendations; 2020.

47. Davies NG, Klepac P, Liu Y, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med. 2020;26:1205–1211.

48. World Health Organization - Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19;. `https://www.who.int/news-room/q-a-detail/herd-immunity-lockdowns-and-covid-19`.

49. Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science. 2020;369(6505):846–849. doi:10.1126/science.abc6810.

50. Science.lu - Survey on willingness to be vaccinated;. `https://science.lu`.