



Social learning in a network model of Covid-19[☆]

Allan Davids ^{a,*}, Gideon du Rand ^b, Co-Pierre Georg ^c, Tina Koziol ^a, Joeri Schasfoort ^a

^a School of Economics, University of Cape Town, South Africa

^b Department of Economics, Stellenbosch University, South Africa

^c EDHEC Business School, France



ARTICLE INFO

JEL classification:

C63

118

Keywords:

Covid-19

Social learning

DeGroot model

Epidemiological network model

Vaccination strategy

ABSTRACT

This paper investigates the effects of social learning on the transmission of Covid-19 in a network model. We calibrate our model to detailed data for Cape Town, South Africa and show that the inclusion of social learning improves the prediction of excess fatalities, reducing the best-fit squared difference from 19.34 to 11.40. The inclusion of social learning both flattens and shortens the curves for infections, hospitalizations, and excess fatalities, which is qualitatively different from *flattening the curve* by reducing the contact rate or transmission probability through non-pharmaceutical interventions. While social learning reduces infections, this alone is not sufficient to curb the spread of the virus because learning is slower than the rate at which the disease spreads. We use our model to study the efficacy of different vaccination strategies and find that vaccinating vulnerable groups first leads to a 72% reduction in fatalities and a 5% increase in total infections compared to a random-order benchmark. By contrast, using a contact-based vaccination strategy reduces infections by only 0.9% but results in 42% more fatalities than the benchmark.

1. Introduction

While epidemiological models have been crucial in steering policy responses to the Covid-19 pandemic, their predictive performance has been poor (Ioannidis et al., 2020; Moein et al., 2021).¹ Economists have focused their efforts to improve the predictive power of epidemiological models towards adding more realistic human behaviour.² One way to do this is via social learning (Golub and Sadler, 2016). Recent empirical evidence (Bailey et al., 2021; Makridis and Wang, 2020) shows that individuals reduce their mobility more if contacts in their social networks live in areas more severely affected by Covid-19.

In Section 2, we introduce naive social learning (DeGroot, 1974) in an agent-based model (ABM) of Covid-19 transmission among physically interacting agents. In our model, agents are represented as nodes in a network and their interactions as time-varying edges.

[☆] We are grateful to seminar participants at George Mason University, Stellenbosch University, and the National COVID-19 Modelling Consortium of South Africa, the ABM4Policy seminar and in particular Krishna Dasaratha for helpful comments and discussion. Finally, we thank Peter Courtney for his excellent research assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The source code and replication files for this paper can be found [here](#). All errors are our own.

* Corresponding author.

E-mail address: allan.davids@uct.ac.za (A. Davids).

¹ See, for example, “The simulations driving the world’s response to COVID-19” - Nature, 2 April 2020

² See Verelst et al. (2016) for a survey of the epidemiological literature that features behavioural change. We discuss the rapidly growing literature of models that feature both disease transmission and some form of optimal agent behaviour in the literature review at the end of this section.

Nomenclature

a	Specific age group
\mathcal{A}	Set of all age groups
as	Asymptomatic
C	Critically ill status
D	Deceases status
E	Exposed but not infectious status
\mathcal{E}	Edges
$f_{i,j}$	Travel flows between two districts
F^{pop}	Observed district population
F^{age}	Observed district level age distribution
F^{hs}	Observed district household size distribution
F^{hc}	Observed age group household contacts
F^{oc}	Observed non-household contacts
F^{tv}	Observed travel matrix
F^{ca}	Observed cases per district
F^{in}	Observed degree of informality
\mathcal{G}	Undirected graph of interactions
h	Household
\mathcal{H}	Set of households
i	Informality level (applied to districts)
I_{as}	Infectious but asymptomatic status
I_s	Infectious and symptomatic status
j	An individual agent
\mathcal{J}_0^E	Set of agents that are infected at the start of the simulation
L	Health system capacity
m	Seed used for individual Monte Carlo simulation
M	Total number of Monte Carlo simulations
\mathcal{N}	Nodes in network, representing the total number of agents
N^{oc}	Number of non-household connections of agents
P	Epidemiological status
R	Recovered status
s	Symptomatic
S	Susceptible to infection status
t	Single time period representing a day
T	Total simulation time
w	Specific district (Ward)
w'	Travel destination district (Ward)
\mathcal{W}	Set of districts (known as wards in South Africa)
γ	Size of overlap between a ward and a region
β^{emp}	Set of empirical moments
β^{sim}	Set of simulated moments
δ_L	Health system overburdened multiplier
τ_E	Latency period
τ^{as}	Recovery period asymptomatic infection
τ^s	Transition period symptomatic infection
$\pi(P_i, X_i; P_j, X_j)$	Base probability of infecting another agent
π^{as}	Probability to enter asymptomatic state
$\pi^{C,s}$	Probability to enter critical state when symptomatic
$\pi^{D,s}$	Probability to enter deceased state when symptomatic
ω_1	Likelihood that an agent is aware that she has the virus
ω_2	Fraction of travel network connections that are active for each agent
ω_3	Maximum number of non-household contacts any agent is allowed to have each day
ω_4	Policy multiplier that can decrease the probability of infection

Each agent has an epidemiological state corresponding to the Susceptible, Exposed, Infected, Recovered (SEIR) compartments in the tradition of [Kermack and McKendrick \(1927\)](#). As in SEIR models, the spread of the virus, measured by the reproductive number R , is determined by two factors: (1) the contact rate, which is the frequency with which different agents interact; and (2) the transmission probability, which is the likelihood an infection takes place if two agents interact. In line with other state-of-the-art economic Covid-19 models ([Ellison, 2020](#)), our model features a heterogeneous contact rate that depends on agent age and location in the city, calibrated using demographic data on age distributions, the composition of households, age-based contact matrices, and travel surveys. The transmission probability is fixed while the contact rate can be directly influenced by the government through non-pharmaceutical interventions (*lockdown regulations*).³

In the existing literature, such *lockdown regulations* are usually modelled as restrictions on agents' ability to interact, i.e. as restrictions on the set of time-varying edges. In contrast, our model allows a much richer endogenous behavioural response, where individuals treat government regulations as one signal of the need for physical isolation, but choose their own degree of isolation—above or below that suggested by government—by learning from the extent of infections as well as the behaviour of those in their network. This allows for more endogenous feedback channels to affect the spread of the virus, which we show improves the predictive power of the model.

We model behaviour as a form of naive social learning that closely follows [Dasaratha et al. \(2020\)](#) and earlier work by [Golub and Jackson \(2010\)](#). Agents choose their degree of isolation based on a private and a social signal. The private signal is composed of a public signal—the government suggested degree of isolation—and a private opinion term, which models individual disagreement with the government's assessment. The social signal is composed of the observed isolation behaviour of neighbours. Agents are aware that neither their own nor the government's opinion is necessarily correct, given the uncertainty about the novel virus. This is a key channel of social learning in networks ([Golub and Sadler, 2016](#)), consistent with recent empirical evidence on the role of social networks during pandemics ([Bailey et al., 2021](#)).

Using this model, we study the impact of social learning on the spread of Covid-19 in Cape Town, South Africa. Cape Town is particularly interesting because all South African epidemiological models overestimated both the projected number of infections at the peak of the first wave as well as its duration, some significantly so.⁴ This was surprising since the South African government, due to economic considerations, started relaxing the stringent national lockdown policies while infections were still rising.⁵ Therefore, most models predicted an exponential increase in infections in South Africa.⁶ Despite the relaxing of the lockdown during this period, the peak of the excess fatality curve—generally considered the most reliable measure of Covid-19—was surprisingly low and it ended quicker than expected.⁷ Our model provides an explanation for this via endogenous and responsive behaviour.

To initialise our model, we use demographic data from the South African National Census, contact matrices from [Prem et al. \(2017\)](#), the National Household Travel Survey, initial detected cases from the Western Cape government, information about the government signal from the Oxford Stringency Index ([Hale et al., 2020](#)), information about observed mobility from Google, and fatality statistics recorded in South African hospitals. For epidemiological parameters we use estimates found in the extant literature. The remaining free parameters of our model, that we could not calibrate using these sources, are the transmission probability and the number of initial infections, as well as two parameters related to the weight and distribution of social learning.⁸ We discuss all parameters in detail in [Section 3](#).

Results are shown in [Section 4](#). We, first, jointly estimate the free parameters for both a version of the model with- and one without social learning, in each version finding the parameter choices to minimise a loss function as in [Nelder and Mead \(1965\)](#). In the model with naive social learning, we obtain a best-fit squared difference between the model prediction and the empirically observed excess fatality curve of 11.40, while for the model without social learning the best-fit squared difference is 19.34. It should be noted that just adding a learning model could mean more degrees of freedom, which could by itself explain the improved fit. However, a more extreme social learning model, which we named 'lexicographic learning', was not able to outperform the no-learning model using the same fitting algorithm, indicating that just adding any learning model is not enough. It should also be noted that there are models which exhibit a better fit ([Ambrosio and Aziz-Alaoui, 2020](#); [Acuña-Zegarra et al., 2020](#)). However, in these cases a functional form is chosen for the reproductive rate as the pandemic progresses, implying that either the contact rate or transmission probability (or both) have an *exogenous* functional form. In contrast, in our model, the reproductive rate *endogenously* slows down as the virus spreads through the network, with a contact rate that is determined by the behaviour of our agents while the transmission probability is fixed.

To better understand how social learning improves the model fit over and above a calibration based on transmission

³ This is similar to other agent-based Covid-19 models such as [Rockett et al. \(2020\)](#), [Almagor and Picasia \(2020\)](#), and the well-known Imperial College Model of [Ferguson et al. \(2020\)](#).

⁴ The South African National Institute of Communicable Diseases (NICD) cites the lack of the inclusion of human behaviour within epidemiological models as a reason for over-projected deaths: "The known the unknown and the unknowable modelling Covid-19 between scarce data and the need to make decisions" –NICD, 24 July 2020.

⁵ This is based on [Hale et al. \(2020\)](#), who develop a lockdown stringency index which scores lockdowns worldwide on a 0–100 scale where 0 is the least and 100 is the most stringent lockdown. In response to the Covid-19 outbreak, South Africa went into a lockdown with a 87.96 stringency index at the start of our simulation period that was slowly reduced to 80.56 by the end of our period of interest.

⁶ See, "[Model prediction overview from the South African Government](#)"

⁷ Excess fatalities are generally considered to be more reliable, given widespread concerns about the under-counting of both infections ([Adepoju, 2020](#)) and fatalities ([Pasquariello and Stranges, 2020](#); [Leon et al., 2020](#)).

⁸ Testing in South Africa was severely limited at the beginning of the pandemic, hence the number of initial cases cannot be reliably estimated and we treat it as a free parameter.

probability—the standard free parameter in these types of models—we perform two sensitivity analyses, independently varying the transmission probability and the strength of the social learning impact on behaviour.

We find that a greater impact of social learning on isolation behaviour simultaneously decreases the height and duration of the curves for infections, hospitalizations, fatalities, and excess fatalities. For example, in our calibrated model, increasing the weight of social learning from 0.85 to 0.95 leads to 7% fewer infections and 14% fewer fatalities while keeping the peak infection day the same. In contrast, a decreased transmission probability of 2% reduces infections by 6% and fatalities by 14%. However, it simultaneously pushes the peak infection day out 3 days.

Furthermore, we find that in contrast to changing the transmission probability, the effect of changing the impact of social learning is non-linear. If we increase the strength of the social signal relative to the private signal, there is very little effect on infections for relative strengths between 0.0 and 0.8. However, between a relative strength of 0.8 and 1.0, the fatalities and infections are strongly reduced until plateauing at 76% of what infections would be without social learning. In our calibration exercise, we find a best-fit relative strength of the social signal of 0.85.

To further understand social learning, we explore how social learning affects the effectiveness of lockdown regulations by comparing the calibrated model with lockdown regulations to a hypothetical scenario in which there are no government interventions. In particular, we are interested if adding behaviour to the model would alter the conclusion that a strict lockdown was necessary, which is one of the main criticisms raised against the standard epidemiological Covid-19 models (see, for example, [Shen et al. \(2020\)](#); [Squazzoni et al. \(2020\)](#)). Comparing our baseline lockdown scenario to a no-intervention alternative in which agents voluntarily reduce their interactions, we find that lockdown regulations remain highly effective even when social learning is introduced.

With best-fit relative strength of social learning, lockdown regulations reduce end-of-simulation infections from 81% to 49% of the population. Furthermore, the intervention flattens the curve and pushes out the peak of infections by five weeks. Also, under the intervention scenario, hospitals are overburdened for 12 days, rather than for 51 days as in the no-intervention scenario. As a consequence of both reduced infections and a less overburdened hospital system, total deaths are reduced by 62% in the intervention scenario, compared to the no-intervention scenario.⁹ Under the intervention scenario agents reduce their contacts on average by 50% (from 18 daily contacts to 9), while under the no-intervention scenario they reduce contacts voluntarily by 19% (from 18 daily contacts to 16).

Social learning reduces infections in the intervention scenario, and even more so in the no-intervention scenario. Introducing social learning together with a lockdown reduces total infections by 4.2%. Compared to this, adding social learning to models of Covid-19 disease transmission has a stronger effect in the no-intervention scenario, reducing infections by 6%, peak critical cases by 12.4% and deaths by 5.1%. Despite this, 80% of the population becomes infected, which can be explained by the fact that even without social learning sheltering in the lockdown model is already high, averaging 82%, limiting the impact of social learning.

In [Section 5](#), we explore how three age-based vaccination strategies affect fatalities, critical cases, and infections, had the vaccine been available at the start of the virus outbreak in Cape Town. The first strategy is the control group in which age classes are randomly given priority. The second strategy is a risk-based strategy in which older agents get vaccinated first. In the third scenario, agents in age groups that have more connections are vaccinated first. Vaccinating the elderly first leads to a 72% reduction in fatalities compared to vaccinating in a random order (which is already effective, reducing fatalities relative to the no-intervention scenario by 65%). However, this comes at the cost of a 5% increase in total infections. On the other hand, the contact-based vaccination strategy reduces infections by only 0.9%, compared to the random benchmark. However, it leads to 42% more fatalities than the benchmark random vaccination scenario.

Finally, in [Section 6](#), we discuss related literature and conclude.

2. The model

We simulate the spread of Covid-19 in a network of physical interactions modelled as an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where the set of nodes \mathcal{N} represent agents and the set of edges $\mathcal{E} \subset (\mathcal{N} \times \mathcal{N})$ represents physical interactions between them.¹⁰

SEIR models employ the sequence of time-specific edges \mathcal{E}_t that represent meetings between agents to study the evolution of the epidemiological state of the system, which is the disease conditions $\{P_j(t)\}$ at all nodes. The evolution of the state depends on a number of features: disease and age-specific transmission and morbidity parameters conditional on a physical interaction, as well as behaviour that determines the likelihood of a physical interaction.

Our main contribution is a novel approach to modelling the impact of lockdowns via behavioural responses. In most models, lockdowns are treated as exogenous restrictions to the set of edges \mathcal{E}_t . In our model, they affect the probability that a meeting occurs along an active edge. Readers familiar with network implementations of SEIR models may turn to [Section 2.4](#) which discusses the behavioural aspects of the model.

In this section, we discuss: (i) the agents and their characteristics, (ii) the network that governs interactions among agents, (iii) the epidemiological status updates and disease transmission, and (iv) the technical details of how we implement the model numerically.¹¹

⁹ We explicitly focus on the cost of life and do not consider economic costs. For an analysis that incorporates economic cost, see for example [Krueger et al. \(2020\)](#), [Acemoglu et al. \(2020\)](#), [Birge et al. \(2020\)](#) and [Eichenbaum et al. \(2020\)](#).

¹⁰ We sometimes refer to physical interactions as social connections.

¹¹ In [Appendix Appendix C](#), we provide a detailed pseudo-code description of the model.

2.1. Agents

There is a set of agents \mathcal{N} , living in households \mathcal{H} in a city with districts \mathcal{W} .¹² Agents $j \in \mathcal{N}$ are characterized by their epidemiological state P_j and four agent-specific parameters, age $a_j \in \mathcal{A}$, home district $w_j \in \mathcal{W}$, district to travel to $w'_j \in \mathcal{W}$ and household $h_j \in \mathcal{H}$ that remain constant throughout the simulation. We model nine age groups, $\mathcal{A} = \{0 - 10, 10 - 20, \dots, 70 - 80, 80 +\}$, and, in our specific application to the cities of Cape Town, 116 districts \mathcal{W} .

An agent's epidemiological status P_j can take seven values: Susceptible to infection (S); Exposed but not infectious (E); Infectious but asymptomatic (I^{as}); Infectious and symptomatic (I^s); Critically ill (C); Recovered (R); or Deceased (D).¹³ The two sub-categories of the infectious status are important as there are clear indications that some individuals with Covid-19 never show symptoms but still infect others. The critically ill category is used to compare the number who require hospitalization to the critical care capacity of the health system—in our model, the probability of death increases if there are more critically ill patients than hospital capacity. While such agents are medically still infectious, we assume they are isolated in hospital where they cannot infect other agents.¹⁴

When initialising the agents $j \in \mathcal{N}$, we assign them a district to live in w_j proportional to the share of people living in this district relative to the full city population. For each agent $j \in \mathcal{N}_w \subset \mathcal{N}$, we randomly assign an age group a_j using the observed district level age distribution data F^{age} .

2.2. Agent interactions

Agents interact in the network \mathcal{E} that represents the possible interactions of individuals within a city. We model agent interaction with other agents separately for household and non-household interactions, with interactions represented by edges in the network calibrated to probability distributions implied by observed contact matrices from Prem et al. (2017), in which we denote F^{hc} for household interactions and F^{oc} for non-household interactions. The probability distributions are constructed from the contact matrices as follows. The raw data of contact matrix \tilde{F} records the best estimates of data on the daily number of contacts between individuals in different age categories. Row a of the square matrix F^{hc} contains the list of the average daily number of contacts that an individual in age category a is expected to have with an individual in each of the age categories represented by the columns. This is converted into a frequency distribution F of the percentage of contact with individuals of different ages by dividing each row by the sum of the row values (which is the total number of daily contacts expected of someone of the age category represented by that row). These frequency distributions are used in the calibration as the probability that an edge is created between two agents. Formally, let \tilde{f}_{ab} be the expected number of contacts between a person of age group a with a person of age group b . Then $f_{ab} = \frac{\tilde{f}_{ab}}{\sum_a \tilde{f}_{ab}}$ is the expected frequency of such a contact from the perspective of a person in age group a , which is the typical entry in the probability distributions F^{hc} and F^{oc} . See Table 2 for further detail.

2.2.1. Household and non-household interaction

We assign agents to households to match the observed distribution of (i) household sizes in each district and (ii) the age-specific within-household distribution. Agents in each household are all connected with one another. We then calibrate a network of non-household interactions to accurately reflect the travel patterns across districts (informed by travel survey data) and the observed age-specific probability distribution of non-household contacts (informed by contact matrices). We elaborate further on this process in Section 3.1.

2.3. Epidemiological status updates and disease transmission

In each period of the simulation $t \in [0, T]$, the epidemiological status of each agent ($P_j(t)$) is updated. Disease transmission takes place as agents interact on the network. Interactions are contacts between two agents who share an edge in \mathcal{E} . When a susceptible agent i ($P_i(t) = S$) interacts with an infectious agent j ($P_j(t) \in \{I^{as}, I^s\}$), transmission of the virus occurs with probability π^E .¹⁵ Importantly, π^E is only positive for infectious agents ($P_j(t) \in \{I^{as}, I^s\}$) and is 0 otherwise. Put differently susceptible, exposed, recovered, critical and dead agents cannot infect susceptible agents. While critical agents are still infectious, we assume they are isolated in hospital where they are unable to infect other agents.

¹² We use the convention that sets are denoted by formal script, e.g. \mathcal{N} ; the value of a variable (such as the cardinality of a set) are denoted by upper case letters (e.g. $N \equiv |\mathcal{N}|$), and generic elements are denoted by lower case letters (e.g. an agent is denoted as $j \in \mathcal{N}$). Indices denoting generic elements of a set are subscripts, while additional identifiers are superscripts.

¹³ The epidemiological literature denotes these as *compartments* and we use this nomenclature as well.

¹⁴ We use this simplifying assumption to model the situation where patients isolated in hospital cannot spread the disease in their regular social networks.

¹⁵ The basic infectiousness of the disease π^E is an uncertain parameter that we treat as a constant over time and across individuals in this model as a simplification (see Section 3.2 for our calibration.). As more certain results become available in the medical literature, simple extensions of the model can consider situations where e.g. children are less likely to become infected and/or transmit the disease than older individuals.

2.3.1. Initial infections

The first agents to update their epidemiological status from susceptible S to one of the infected compartments (E , I^{as} or I^s) are agents that are created at the start of the simulation ($t = 0$). We assign initial infections based on the observed number of cases per district at the beginning of the Covid-19 outbreak, F^{ca} . Within each district, we pick a random agent and update their initial status to $P_j(0) \in \{E, I^{as}, I^s\}$ (with equal probability), and assume that the agent has been in this state for a random number of days $T_j^P(0) \in [0, \tau^P]$, where τ^P is the tenure in compartment P .

2.3.2. Evolution of previously infected agents

Agents transition through the various phases or compartments of the disease via a system of between-state probabilities, which denote the probability that an agent moves from any given state to the following state. We present the disease progression schematically in Fig. 1. We calibrate these probabilities using estimates from the literature which we discuss in a later section.¹⁶

While the probability of infection π^F and the probability of developing a symptomatic version of the disease π^s are age-invariant probabilities, we model the probability of becoming critical π^{C,a_j} and the probability of death π^{D,a_j} as age-varying probabilities consistent with existing evidence of varying mortality rates by age-group (Verity et al., 2020). Lastly, we augment the probability of death π^{D,a_j} to include an important feature of the Covid-19 pandemic, increased mortality rates when the health system is overburdened. We model this via a multiplier δ^L :

$$\delta^L = \begin{cases} Q > 1 & \text{if } N^C(t-1) > L \\ 1 & \text{otherwise} \end{cases}, \quad (1)$$

where L is the capacity of the health system, $N^C(t-1)$ is the total number of agents in the critical state in period $t - 1$, and Q is the empirical multiplier that increases the probability of dying when hospital capacity is overwhelmed. This generates a probability of death of $\delta^L \pi^{D,a_j}$.

2.4. Agent behaviour, lockdown regulations and new infections

We study the impact that behavioural responses to lockdown regulations have on the transmission of the disease. These interventions can take many forms, such as stay-at-home orders, limits on indoor activities, and policies that encourage proper hygiene, mask-wearing and social distancing. For simplicity, we consider only one aspect of these policies in our model: how much they reduce physical interaction among agents.

The primary methodological contribution of this paper is to consider how social learning affects the desire of individuals to reduce contacts on their own and how that interacts with lockdown regulations and in turn study how this affects disease transmission¹⁷. Our implementation of social learning closely follows that of Dasaratha et al. (2020).

In each period t , every agent j observes a private signal $\zeta_j(t)$ which is composed of a public signal $\xi(t)$ and an individual opinion term $e_j(t)$.

$$\zeta_j(t) = \xi(t) + e_j(t).$$

The public signal $\xi(t)$ represents the stringency of lockdown measures put in place, and thus a signal of the proportion of connections the government recommends an individual should cut. We model the public signal using a time-varying lockdown stringency index $\xi(t) = \frac{F^{strin}(t)}{100}$ from Hale et al. (2020), measuring the stringency of the lockdown in South Africa relative to all other countries, where $\xi(t) = 1$ represents the most stringent lockdown and $\xi(t) = 0$ the least stringent. During the most stringent lockdown, the government signals that agents should only interact within their household plus contacts due to 'essential' activities, while the least stringent lockdown places no restrictions on physical interactions.

The opinion term $e_j(t)$ represents the individual's idiosyncratic attitude towards the lockdown policies. It is drawn independently for each j from a truncated normal distribution with support $[-\xi(t), 1 - \xi(t)]$, mean 0 and standard deviation σ , such that the private signal $\zeta_j(t) \in [0, 1]$. This means an agent with $e_j(t) = 0$ privately agrees with the government's suggestions (i.e. $\zeta_j(t) = \xi(t)$), while an agent with $e_j(t) = -\xi(t)$ (i.e. $\zeta_j(t) = 0$) privately thinks it is not necessary to cut any interactions; and an agent with $e_j(t) = 1 - \xi(t)$ (i.e. $\zeta_j(t) = 1$) privately thinks maximal isolation is appropriate, regardless of the government's advice.

After observing their private signal, agents choose their *desired proportional reduction in contacts* $\phi_j(t) \in [0, 1]$ (which we call compliance for short). We model the level of compliance $\phi_j(t)$ of each agent as a weighted average of the private signal $\zeta_j(t)$ (with a weight of ρ) and a social signal (with a weight of $(1 - \rho)$).

The social signal is the simple average of observed $t - 1$ desired contact reduction of all neighbours, $\frac{1}{N_j} \sum_{k \in \mathcal{N}_j} \phi_k(t - 1)$. This yields:

¹⁶ See footnote 15 for a discussion on the basic infectiousness of the disease π^F , which is an uncertain parameter.

¹⁷ An alternative configuration might be that regulations affect the transmission probability. In this paper, we have used the contact rate because this produced a better fit for Cape Town and the main tool that was varied by the government took the form of mobility restrictions.

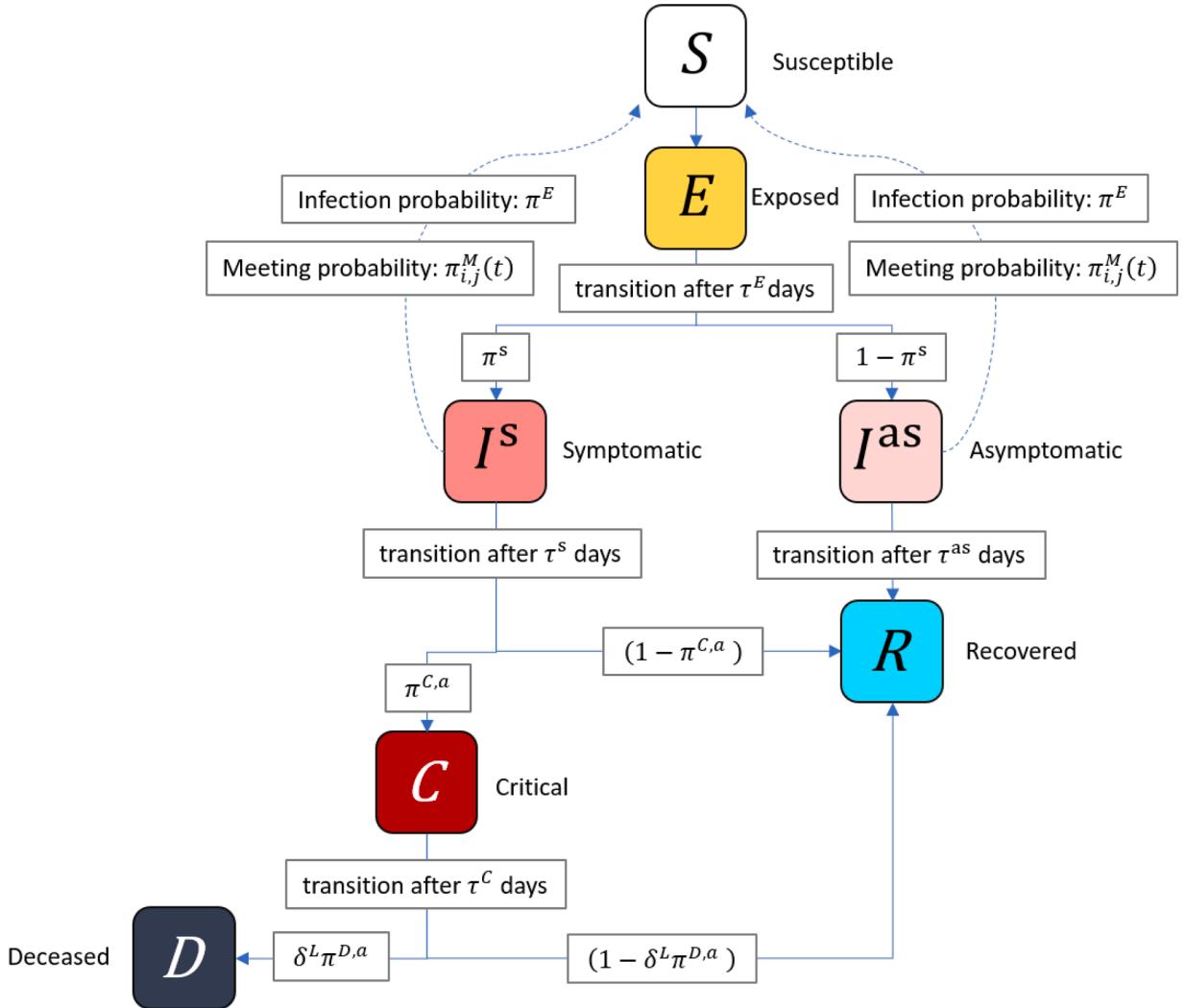


Fig. 1. Schematic of the disease progression. This schematic shows the disease progression modelled in our model. Solid arrows between compartments represent transitions of an agent to different disease statuses. Dashed arrows represent social connections along which the virus may be transmitted from infectious to susceptible individuals.

$$\phi_j(t) = \rho \zeta_j(t) + (1 - \rho) \frac{1}{N_j} \sum_{k \in \mathcal{N}_j} \phi_k(t-1)$$

In this way, desired contact reduction depends both on an agent's opinion of the aggregate stringency of lockdown measures as well as the observed desired contact reduction levels of their neighbours in the previous period.¹⁸ This behaviour (agents' contact reduction decisions) can be obtained as the optimal result of utility maximization by agents trying to match their desired contact rate to the state of the world, as in Dasaratha et al. (2020). This can be obtained in a model where agents dislike physical distancing (when there is no risk), but have a preference to reduce contacts when there is a risk of Covid-19 transmission.

Finally, to account for agent reactions to actual observable (symptomatic) infections, hospitalizations, or fatalities, if an agent's status changes to any of these categories, $P_j \in \{I^S, C, D\}$, it will set its desired contact reduction rate to 1, $\phi_j(t) = 1$.

After choosing a desired reduction in contact rate, a physical meeting between two connected agents i and j occurs with a time-varying, pair-specific probability $\pi_{i,j}^M(t)$ modelled as:

¹⁸ Idiosyncratic noise ($\varepsilon_j(t) \neq 0$) is necessary for there to be any interesting learning effects. If $\varepsilon_j(t) = 0 \forall j$ then $\zeta_j(t) = \xi(t) \forall j, t$ and $\phi_j(t)$ will differ from $\xi(t)$ only due to exogenous observation lag.

$$\pi_{ij}^M(t) = \begin{cases} 1 & \text{if } i \in \mathcal{N}_j^{\text{hc}} \\ (\omega + (1 - \omega)(1 - \phi_i(t))(\omega + (1 - \omega)(1 - \phi_j(t))) & \text{otherwise} \end{cases}, \quad (2)$$

The probability $\pi_{ij}^M(t)$ firstly depends on the nature of contacts - whether these contacts occur within the household or outside the household. Agent j has daily physical meetings with all members of their household, so that if $i \in \mathcal{N}_j^{\text{hc}} \Rightarrow \pi_{ij}^M(t) = 1$. Second, for members in the set of non-household contacts of agent j , the probability of a meeting between j and $i \in \mathcal{N}_j^{\text{oc}}$ depends on three parameters, $\omega \in [0, 1]$ and $\phi_i(t)$ and $\phi_j(t)$.

We model $\omega \in [0, 1]$ to represent essential contacts that each agent will visit even under the most stringent lockdown regulations, where $\omega = 0$ means that no contacts are essential, whereas $\omega = 1$ means that all contacts are essential. Therefore, in between these two extremes, this parameter reflects the fact that agents will still have some contacts even if they desire to reduce all of their contacts. After all, while lockdowns restrict movement, many agents will still need to leave home, most notably essential workers.

The probability that agent i and agent j meet is therefore a function of their own desired contacts and the fraction of non-household contacts that are essential. Put differently, even when governments enact a lockdown to which all agents comply completely, some mobility always occurs (ω). Over and above this level of minimal mobility agents can make additional contacts, and the extent to which they make additional contacts depends on their desired rate of (non-essential) contacts $(1 - \phi)$. To build intuition consider the extreme case where both agents fully reduce desired contacts, $\phi_i(t) = \phi_j(t) = 1$. The probability of meeting reduces to ω^2 , the minimum amount of mobility. This functional form therefore encodes two features: (i) even with a full desire to limit contacts, some physical meetings may happen during the course of everyday life, and (ii) individuals can always coordinate to ensure a meeting should they wish to do so strongly enough. An alternative way of thinking about this, within our model framework, is that without any intervention the base assumption is that individuals meet with their non-household contacts with certainty every day. Without the inclusion of ω , the probability of two agents meeting would reduce to a function of their desired contact rates, and two agents that want to reduce all contacts would never meet. Given what we know about essential workers and purchasing basic necessities, contacts do still occur, even in a full lockdown. The inclusion of ω therefore prevents agents from being able to fully cut themselves off from making connections in their network of non-household contacts.¹⁹

In summary, in any given period agent i and j , who are connected in the network, physically meets with probability $\pi_{ij}^M(t)$. Conditional on a meeting, if one agent is susceptible and the other is infectious, the virus is transmitted with probability π^E .

2.5. Implementation

We implement the model using Python and publicly available data, ensuring full reproducibility.²⁰ In addition to the variables defined above, we make use of auxiliary (internal) variables during the simulation to keep track of how long each agent has been in each epidemiological state. Furthermore, we keep track of the number of other agents who have been infected by each agent in the current period as well as in total. Finally, we store each agent's neighbours, their number of contacts, as well as the household number of their neighbours.

For our first simulation day, we record the sets of nodes \mathcal{N} and edges \mathcal{E} . On every simulated day t , we keep track of the epidemiological status P , district number w_j , and age group a_j of every agent along with the number of other agents the agent has infected during that day. This data is recorded and stored.

We use Monte Carlo methods, simulating the model V times. For each simulation, we seed the pseudo-random number generator with a different number $v \in \mathcal{V}$, ensuring both that its results can be replicated and that changes in output for the same seed can be attributed to changes in policy rather than changes in stochastic factors.

3. Calibration

Our model calibration can be divided into four parts, each of which we briefly discuss in this section.²¹ First, we set a baseline for the simulation-specific parameters and discuss how we initialize the model. Second, we assign values to parameters that are associated with the clinical features of Covid-19 progression. Third, we calibrate the geo-spatial and demographic features to data for Cape Town. Finally, we estimate the remaining parameters for which we could not find a value in the data or literature.

3.1. Model initialization

We run our simulations with $N = 100,000$ agents. Every simulation runs for a minimum of $T = 1,171$ days and is typically repeated

¹⁹ Economic epidemiological models encode a similar logic whereby agents trade off wanting to leave home to earn an income to satisfy their economic needs versus wanting to stay at home to minimize the risk of infection. In these models, no matter how high the infection cost is, there will always be agents for whom the economic benefit from making a contact with others exceeds the cost of infection, given a distribution of economic needs. The forces that drive these kinds of agents to always make trips, irrespective of infection risk, are captured in $\omega > 0$.

²⁰ Find the code and replication files on our GitHub repository, [here](#).

²¹ We provide a more extensive discussion of the calibration process in [Appendix Appendix B](#).

$V = 50$ times using Monte Carlo simulation methods to average out stochastic effects.²² We then initialize our model in order to match the observed features of household and non-household interactions.

3.1.1. Household interactions

As mentioned in Section 2.2 we implement an algorithm to calibrate household interactions in our model such that the sets of household edges in each district match (i) the size and age distributions in districts are representative in aggregate, and also (ii) the household size and age distributions are representative within each district.

In constructing the network of household interactions, we first create households of various sizes in order to match the observed empirical distribution of household sizes for each district using national census data. In the second step, we proceed to assign agents to households in order to match the observed within-household age probability distribution, which we derive from contact matrices that record household-level contacts. Agents in each household are all connected with one another.

3.1.2. Non-household interactions

Our model also features two key properties of non-household interactions: heterogeneous contact rates by age and realistic travel patterns. We initialize non-household interactions in two steps.

In the first step, we calibrate the network of each agent's non-household interactions to match the observed distribution of (work and school) travel patterns across districts, using travel survey data. This amounts to assigning a *destination* district to each agent, representing the district they travel to for work or school. In the second step, we calibrate the network of non-household interactions to match the observed age-specific distribution of non-household interactions from contact matrices. This involves constructing edges between each agent and other agents in the same *destination* district where these edges are informed by non-household contact matrices.

Iterating over all agents, the initialization of household and non-household contacts generates a network which contains, for each agent j , a network of household and non-household agents to or from whom the virus can be transmitted.²³ We present a stylized network for a model with 40 agents in two districts in Fig. 7 where household edges are dotted and other edges are continuous. The nodes are agents and the colours mark differences in their epidemiological status p_j . The main clusters that can be observed represent agents that travel to the same location every day. The virus can then spread to other districts via household links when agents travel back to their home district.

3.2. Covid-19 parameters

With respect to parameters which pertain to the Covid-19 pathogen, we choose the best available estimates from peer-reviewed medical journals. Despite there still being considerable uncertainty around the clinical course and transmission of the disease, we were able to find literature estimates for most parameters.

The Covid-19 parameters are chosen from recent studies (see Chen et al., 2020; Verity et al., 2020; Huang et al., 2020, among others). Table 1 provides an overview of all parameters and their sources. Because we could not find an estimate in the literature for π^E , the probability of transmission when an infectious agent comes into contact with a susceptible agent, we will estimate its value along with our policy parameters.²⁴ This procedure will be described in Section 3.4.

3.3. Applying the model to Cape Town

We calibrate our model to Cape Town, which covers over 2400 square kilometres and has a population of 3,740,026.²⁵ The city is sub-divided into $W = 116$ administrative districts, known as wards.²⁶ Calibrating our model to Cape Town means that we populate our input files with data that is representative of the city and set our health system parameter based on actual hospital capacities. In this way, our model can easily be applied to any other city where similar data is available.

Our first input data file F^{pop} contains the population and age distribution per ward, The second file F^{hs} is the empirically observed household size distribution for all wards in the city. For each of these files, we use ward-level data from the South African National Census for 2011.

Then, we make use of social contact matrices for South Africa as a proxy for social contact matrices in Cape Town, obtained from Prem et al. (2017). The contact matrices are the 9×9 age-group household contact matrix F^{hc} , and non-household contact matrix F^{oc} , respectively. These matrices specify how many average daily contacts people in a particular age group have with others, by age group. The matrices include total contacts, household contacts, as well as work and school contacts. We illustrate the total household and non-household social contact matrix in Table 2. Between the age groups 0–10 and 40–50 we see a significant share of interactions occur

²² $T = 1,171$ is the length of our empirical reference data.

²³ These networks can be impacted differently by lockdown regulations, which can, for example, prevent non-household contacts but has no impact on household contacts.

²⁴ There likely is no general transmission probability when two people meet since local factors such as climate, ventilation, and cultural proximity affect this parameter.

²⁵ The latest date for which we have survey data is 2011.

²⁶ We will refer to them as wards from here on out.

Table 1

Covid-19 pathogen-related parameters This table reports parameter choices for the Covid-19 pathogen used in the model. For each parameter, we report the chosen value and the applicable medical study.

Name	Value	Source(s)							
Latency period (τ^E)	4	Silal et al. (2020)							
Recovery period asymptomatic infection (τ^{as})	7	Silal et al. (2020)							
Transition period symptomatic infection (τ^s)	11	Huang et al. (2020); Silal et al. (2020)							
Days in critical condition (τ^C)	11	Chen et al. (2020); Silal et al. (2020)							
Probability of becoming asymptomatic (π^{as})	0.6165	Ing et al. (2020); Yang et al. (2020)							
Probability of becoming critically ill ($\pi^{C,a}$)	Panel (b)	Verity et al. (2020)							
Probability of death ($\pi^{D,a}$)	Panel (b)	NICD							
Health system overburdened multiplier (δ_L)	1.79	Chen et al. (2020)							
(a) Covid-19 pathogen related parameters									
	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80+
$\pi^{C,a}$	0.1%	0.3%	1.2%	3.2%	4.9%	10.2%	16.6%	24.4%	27.3%
$\pi^{D,a}$	2.1%	3.3%	3.4%	5.3%	9.7%	15.5%	24.9%	30.1%	37.1%
(b) Age-specific probabilities of entering critical or deceased states									

Table 2

Age-based social contact matrix This table shows the age-based social contact matrix of total contacts for South Africa obtained from Prem et al. (2017). While the original contact matrix contains the absolute number of contacts, in this representation, we normalize contacts row-wise within age group so each value can be interpreted as a percentage of all contacts for that specific age group. Values should be read across columns with each column representing the person from which the contact originates and each row representing the person receiving the contact. We add one additional column and row to this table corresponding to individuals older than 80 years old. While our model contains individuals in this age group, the data we use does not include this age group. We set the contact matrix values of individuals above the age of 80 equal to the contact matrix values of individuals between 70 and 80 and therefore explicitly assume an identical contact structure.

	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 +
0 - 10	0.51	0.11	0.08	0.17	0.13	0.16	0.19	0.17	0.17
10 - 20	0.12	0.61	0.16	0.12	0.20	0.19	0.14	0.25	0.25
20 - 30	0.11	0.10	0.43	0.19	0.15	0.18	0.14	0.08	0.08
30 - 40	0.14	0.07	0.16	0.27	0.20	0.16	0.19	0.11	0.11
40 - 50	0.06	0.07	0.10	0.15	0.21	0.16	0.14	0.13	0.13
50 - 60	0.03	0.02	0.05	0.06	0.07	0.11	0.10	0.09	0.09
60 - 70	0.02	0.01	0.01	0.02	0.02	0.03	0.07	0.07	0.07
70 - 80	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.05	0.05
80 +	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.05	0.05

on the diagonal, indicating predominantly within age-group contacts. However, for age groups 50–60 and above, the diagonals become less significant, indicating substantial cross-age group interaction. This is a particularly important feature given mortality rates are highest among the elderly, who also have the highest cross-age group interactions, especially with the young.

Next, we construct the travel matrix F^v using travel patterns from the 2013 National Household Travel Survey, a nationally representative travel survey undertaken by Statistics South Africa, to calibrate realistic travel patterns, as discussed in 3.1.²⁷ We describe the process by which we map the travel survey data to our model in more detail in Appendix B.

For the distribution of initial infections, we make use of F^{ca} , a data set that contains the total number of detected cases per ward. For each Ward, we divide the number of observed infections over the total number of infections in the metropolitan area and use this to distribute initial infections over the city. However, because the detection of cases in South Africa was initially severely limited by testing capacity, we do not use detected cases to initialise infections in the simulation. Instead, we treat this parameter as uncertain and estimate its value.

We calibrate the health system capacity as follows. According to official sources, there were 2162 acute beds in the Western Cape

²⁷ The travel survey only asks respondents where they live, work and where they attend school. As a result, travel patterns reflect patterns related specifically to work and school and not for other reasons such as leisure or shopping, for example.

province on the 22nd of May 2020.²⁸ An additional 1428 care beds were scheduled to be provided by temporary hospitals. Of these, 89% were to be in Cape Town. Assuming that this ratio holds for all beds, we set our health systems capacity to be $L_{ct} = 0.000917$, the fraction of acute beds available in Cape Town divided by the total population.

Finally, we calibrate $\omega \in [0, 1]$, the minimum share of contacts that agents need to have even under a full lockdown, using Google observed mobility data.²⁹ We take the value of the mobility index on 26 March 2020, the day upon which South Africans entered into a very strict lockdown in which the streets were almost empty. We therefore assume that any travel recorded on this day must have represented travel by essential workers. This results in $\omega = 0.46$.

For the government signal, we use the Stringency Index F^{strin} (Hale et al., 2020), published by the Oxford Blavatnik School of Government (Hale et al., 2020). The index scores lockdowns worldwide on a 0–100 scale where 0 is the least and 100 is the most stringent lockdown. In response to the Covid-19 outbreak, South Africa went into a 87.96 stringency index lockdown at midnight on 26 March 2020. The stringency of this lockdown was eased to 84.26 on the first of May, to 80.56 on the first of June, and then to 76.85 on June 8th.

3.4. Estimating uncertain parameters

There are four parameters for which we could not find reliable values in the literature or data: the standard deviation of the shock to the private signal $\sigma \in (0.0001, 0.1)$, the weight of the private signal $\rho \in (0.05, 0.7)$, the probability of transmission $\pi^E \in (0.025, 0.35)$, and initial infections $E + I^{as} + I^r = (1000, 3500)$. We proceed to estimate these parameters.

Since the possible number of combinations of four uncertain parameters is large, we first use Latin Hypercube Sampling (Stein, 1987) to efficiently select 5 initial four-parameter combinations. These values will serve as the starting point for our estimation procedure.

Next, we take a two-step approach in which we estimate the values of all parameters. For each initial set of parameters, we use the Simulated Method of Moments (SMM) methodology (Franke and Westerhoff, 2012).³⁰ The method requires that one chooses which moments of the empirical data one wants to replicate. In our case, we decide to focus on the observed excess fatality data for the Cape Town metropolitan area over a period of 117 days, the length of F^{strin} . Following this approach, we minimise a quadratic loss function B using a constrained Nelder-Mead simplex algorithm (Nelder and Mead, 1965) which terminates after 10 iterations, where B is defined as:³¹

$$B(\beta^{sim}, \beta^{emp}) = (\beta^{sim} - \beta^{emp})' (\beta^{sim} - \beta^{emp}). \quad (3)$$

Here, β^{emp} and β^{sim} are sets containing the empirical and model-generated moments respectively. For each iteration in the optimiser, we simulate $V = 15$ Monte Carlo Simulations and score each simulation independently. We then average these scores for a final score.

Using this procedure, we jointly estimate all four uncertain parameters for all 5 initial parameter sets, pick the best fit, and then use some manual calibration for the last mile. This yields the following estimates: $\sigma = 0.05$, $\rho = 0.15$, $\pi^E = 0.02989$ and $E + I^{as} + I^r = 928$ initially infected agents, where initial infections are randomly distributed across these three compartments. This last estimate translates to 34,707 initial infections, given that our 100,000 agents represent 3,740,000 people.

3.5. Model validation

To validate our model against the data, we choose to compare simulated fatalities with realized excess fatalities in Cape Town. Excess deaths represent the number of weekly deaths recorded during 2020 relative to the average number of deaths recorded during the same weeks in 2018 and 2019.³² Excess deaths are regarded as the most accurate measure of COVID deaths, given concerns regarding the under-reporting of true COVID deaths in national statistics. We report the number of estimated deaths produced by our model compared to the number of excess deaths reported for Cape Town in Panel (a) of Fig. 2. The root mean squared error (RMSE) between simulated and excess deaths is 11.4. Our model is able to replicate the pattern of excess deaths for Cape Town reasonably well, which provides reassurance that our model is able to capture the basic dynamics of COVID-19.

4. Results

In this section we present our simulation results with the goal of illustrating the role of social learning in a Covid-19 epidemiological network model. Throughout this section, we present results for simulations with our baseline model over $M = 50$ random seeds, reporting the average effects along with 95% confidence intervals.³³

²⁸ See for example: "Winde confirms pressure on hospital system increased, despite not yet hitting peak capacity" - News24, 22 May 2020

²⁹ Available to download, [here](#). Accessed on 04/17/2021.

³⁰ The seminal papers developing SMM are McFadden (1989); Duffie and Singleton (1990); Lee and Ingram (1991).

³¹ The constraint restricts input parameters to positive values only and was developed by Alex Blaessle. The source code is available [here](#).

³² See the South African Medical Research Council's [website](#) for more information regarding data cleaning and statistical methods used.

³³ Calculations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team ([hpc.uct.ac.za](#)) and the University of Stellenbosch's HPC1, Rhasatsha ([www.sun.ac.za/hpc](#)).

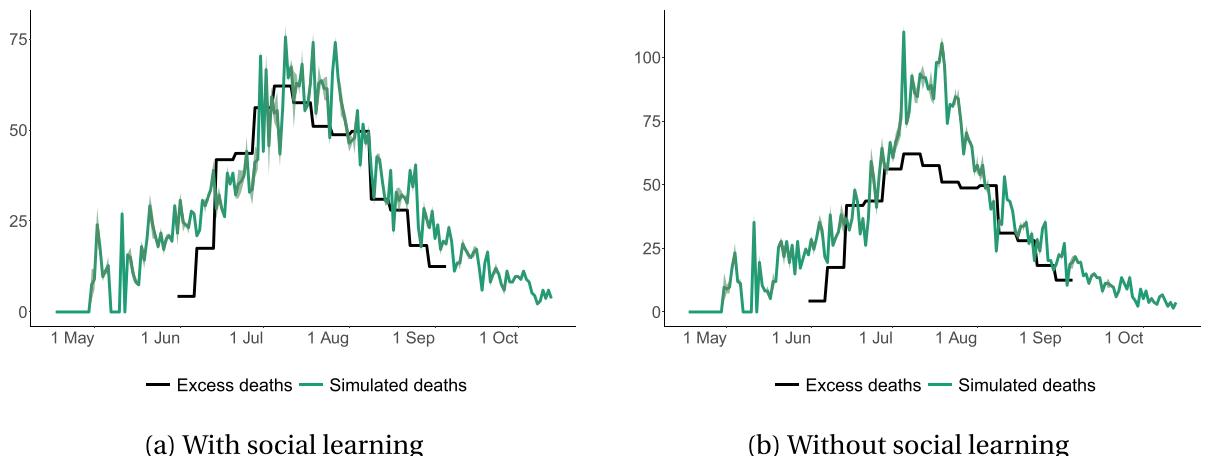


Fig. 2. Simulated deaths and observed excess deaths with and without social learning This figure shows how simulated deaths compare with excess deaths recorded in Cape Town in a model with social learning and a model without social learning. Both scenarios represent the best fit obtained using the calibration procedure outlined in Section 3.4. All figures show the mean outcome taken across 50 simulations. The shaded area indicates the standard deviation. In the model including social learning, the root mean squared error (RMSE) between simulated and excess deaths is 11.40. The RMSE in the model without social learning is 19.34.

This section is structured as follows. First, we show the best fit we could find for the model without learning and compare it to the fit with learning. Then, we show how differences in the weight of social learning affects the spread of the virus and contrast this to the impact of differences in transmission probability. Finally, we show how lockdown policies interact with social learning.

4.1. Fitting excess death curves with- and without social learning

First, we fit the empirically observed excess fatality curve in Cape Town without social learning. Fig. 2 compares the fit of a model with social learning to one without social learning. We show results for best fit obtained using the calibration procedure outlined in Section 3.4.

4.2. Social learning, transmission probability, and contact rates

Our fitting exercise in Section 4.1 shows that social learning can improve the model fit significantly. In this section we vary the weight placed on the social signal (relative to the private signal) in the compliance decision of each agent in order to better understand how the extent of social learning affects our findings. In addition to this, we also perform a similar robustness test where we vary the values of the other free model parameters: the transmission probability and the contact rate.

Specifically, we first check how sensitive our model is to changes in the weight of the private signal, ρ (conversely, changing the weight of social learning, $(1 - \rho)$). To do this, we vary ρ around the calibrated value of $\rho = 0.15$ in the range between $\rho = [0.0, 0.3]$ in steps of 0.05. In this scenario, $\rho = 0.0$ represents a model with only social learning. Second, we vary the transmission probability π^E around its calibrated value $\pi^E = 0.02989$ in the range $\pi^E = [0.02689, 0.03289]$ in steps of 0.01. Finally, we vary the minimum share of contacts that each agent will still visit under the most stringent lockdown regulations ω around its calibrated value $\omega = 0.407$ in the range $\omega = [0.257, 0.557]$ in steps of 0.05. Here, we note that while the contact rate is endogenous in our model, ω affects the contact rate at every point in the simulation independently from lockdown policies and learning. Therefore, it allows us to test the broader sensitivity of the contact rate.

Fig. 8 shows how these three variables impact the three most important epidemiological curves: the number of expected infections, hospitalised agents, and fatalities. It shows that while both a reduction in the weight of the private signal and a reduction in the transmission probability reduce the total number of infections, fatalities, and critical cases, social learning brings the peak of the curve forward while not making it longer whereas both a reduction in transmission probability and general contact rate flattens the curve. On the other hand, decreasing the weight of the private signal (and increasing the weight of the social signal) reduces the number of infections, but does not change their duration. When we compare two extreme cases we find that restricting the model to only social learning ($\rho = 0$) merely moves the date of peak infections forward by 2 days relative to a model without social learning ($\rho = 1$). While reducing the weight on the private signal reduces infections, there is a strong non-linearity in this relationship when the private signal is eliminated altogether ($\rho = 0$). In fact, going from $\rho = 1$ to $\rho = 0.05$ reduces total infections by 10%, while going from $\rho = 0.05$ to $\rho = 0$ reduces total infections by 16%.

On the other hand, reducing the transmission probability reduces the number of infections as well as the number of peak infections and also moves back the date of peak infections. When compared to a simulation with $\pi^E = 0.33$, a simulation with $\pi^E = 0.27$ produces 17% fewer total infections, 43% fewer peak infections and moves the date of peak infections back by 9 days. In addition, the virus also

takes longer before it dies out completely. As a result, changes in the two uncertain parameters ρ and π^E produce quantitatively different effects, while both an increase in ρ and a reduction in π^E reduce infections, the former brings forward the date of the peak while the latter pushes it back in time.

The fact that social learning reduces the height of the curve without stretching it out can be explained by the fact that agents will only increase their desired contact reductions, and hence reduce their contacts, once the virus has affected their neighbours in their social network. That being said, lockdown policies by the government are also a major factor in determining desired contact reduction rates and contacts. As Fig. 3 shows, increasing the weight of social learning non-linearly increases desired contact reduction rates and reduces contacts, especially once the virus became widespread in July 2020. Finally, as can be seen in Fig. 3, with the full weight on social learning, the increased rate of fatalities results in almost 100% reduction in desired contacts by agents.

4.3. Social learning and government intervention

Finally, since our calibration period coincided with government lockdown policies, we examine how social learning affects the spread of the virus in the hypothetical absence of such measures.

We consider two scenarios. The first scenario is our baseline, the scenario discussed in Section 3. We contrast this with a second ‘no-intervention’ scenario in which we assume no interventions by the government, meaning that the public signal $\xi(t) = 0$ for $t \in T$. This allows us to analyse the effectiveness of the government lockdown, which is important to quantify, given the high economic cost of the lockdown in South Africa (Arndt et al., 2020), as well as the effect of social learning when the government does not intervene.

First, we consider the effect of the lockdown in the model with social learning. In the no-intervention scenario, infections among the population exceed 3 million, with 81% of the population becoming infected. Peak infections are reached on day 43 of the simulation and the healthcare system becomes overwhelmed, with the number of critical cases exceeding the hospital capacity for 51 days, resulting in additional deaths as a result of insufficient healthcare. When no more agents are infected or in the hospital, there are 12,973 deaths resulting in an infection fatality ratio (IFR) of 0.43%. Panel (e) in Fig. 4 shows the evolution of desired contact reduction rates. Desired contact reduction rates have a similar shape to infections and critical cases and desired contact reduction rates approach their peak roughly at the same time as infections and agents begin to reduce their number of contacts, as illustrated in Panel (f). This endogenous increase in desired contact reduction rates and a reduction in contacts in response to rising infections illustrates the effect of social learning in our model. Despite this, social learning happens too slowly and always lags infections—once desired contact reduction rates reach their peak, infections have already reached their peak. This highlights an important role for policy in the form of a lockdown: increasing desired contact reduction rates when infections are low, in anticipation of a rise in infections in the near future.

In the lockdown scenario, infections are significantly lower, with infections reaching 1.8 million, resulting in 49% of the population becoming infected. The trajectory of infections is also quantitatively different from the trajectory in a no-intervention scenario: the curve is flatter (i.e. with a smaller peak and larger standard deviation) and the virus takes longer to die out completely. Peak infections are reached on day 79, more than five weeks later than in a no-intervention scenario. Furthermore, peak infections and peak critical cases are significantly lower, ensuring that the healthcare system is only overburdened for 12 days. Altogether, this results in 4934 deaths and an IFR of 0.27%, a little more than half the IFR in a no-intervention scenario. Desired contact reduction rates are also distinctly different under a lockdown, as illustrated in Panel (e). At the beginning of the simulation, desired contact reduction rates are already high, taking a value of around 0.8, compared to a mechanical value of 0 in the no-intervention scenario. This is because under a lockdown, desired contact reduction rates become a function of a private signal, $\zeta_j(t)$ which reflects lockdown stringency and social learning $\frac{1}{N_j} \sum_{k \in N_j} \phi_k(t-1)$, whereas in the no-intervention scenario, the private signal equals zero by definition, $\zeta_j(t) = 0$. As a result, the introduction of a lockdown leads to an upward shift in desired contact reduction rates early in the simulation, at precisely the time when infections and desired contact reduction rates due to social learning are low. Desired contact reduction rates remain high throughout the simulation as the lockdown (with varying intensity at different times) remains intact.

These results highlight that four important features of a lockdown are still present in a network model with social learning. First, lockdown regulations reduce mobility and thus infections and consequently the number of deaths. Second, they have an additional dampening effect on deaths by limiting the extent to which the hospital system becomes overburdened, reducing the number of preventable deaths which arise solely due to inadequate access to health services. Third, lockdown regulations delay the peak of infections, providing additional time for policymakers to deal with the virus. One benefit of this additional time is that it gives policymakers time to expand the capacity of the healthcare system. Finally, lockdowns increase desired contact reduction rates early in the simulation when endogenous desired contact reduction rates due to social learning is low. While social learning is effective in inducing endogenous increases in desired contact reduction rates, independent of policy action, it always lags infections, reducing its effectiveness in suppressing infections. Lockdowns create the initial increase in desired contact reduction rates and a reduction in contacts necessary to suppress the extent to which the virus seeds. Highlighting this effect is a key contribution of our paper.

Next, we test how varying the weight of social learning affects the virus curves in the no-intervention scenario and compare it to our baseline. Comparing Fig. 8 to Fig. 9, the same basic patterns emerge when we vary the weight of the private signal ρ . Introducing and increasing the extent of social learning lowers the peak of the infections, but does not change the duration of the infections. For example, comparing the two extreme cases, a model with only social learning ($\rho = 0$) merely moves the date of peak infections forward by 2 days relative to a model without social learning ($\rho = 1$). The strong non-linearity in the relationship between infections and the private signal from the baseline scenario also holds in the no intervention scenario.

The biggest difference is that now the private signal $\rho = 0$ acts as an anchor for desired contact reduction rates. Again, desired

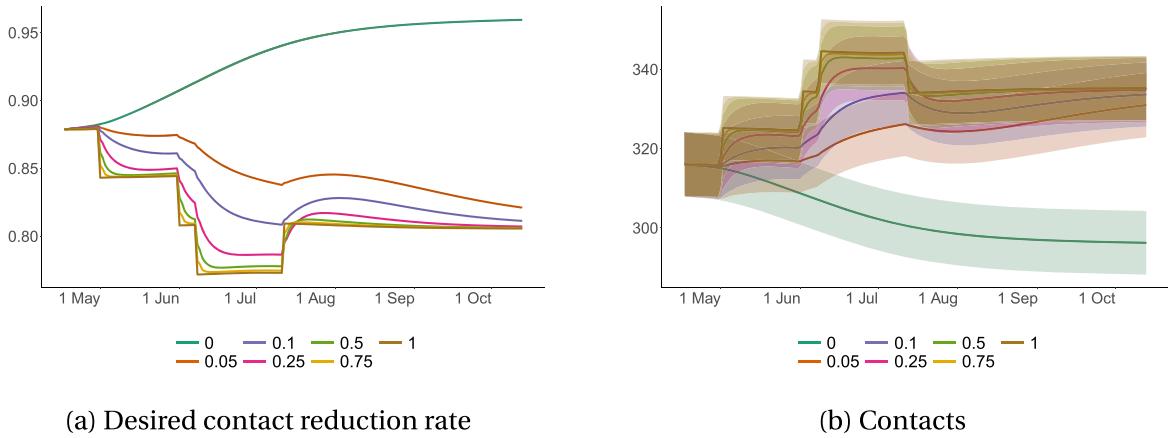


Fig. 3. Desired contact reduction rate and the number of contacts with variations in the strength of the private signal. This figure shows desired contact reduction rates and realised contacts for Cape Town in a Lockdown scenario, for different values of the weight, ρ , of the private signal, $\zeta_j(t)$, in the agent's desired contact reduction decision. All figures show the mean outcome taken across 50 simulations. For reference, in our baseline specifications, our calibrated value of ρ is 0.15. The shaded area indicates the standard deviation.

contact reduction rates will remain higher only in the extreme case of a full weight on social learning. In all other cases desired contact reduction rates tend to go back to zero once the virus subsides. As a result, increasing the weight of the social signal mechanically increases desired contact reduction rates as indicated in Panel (c) of Fig. 9. However, when comparing $\rho = 0.05$ to $\rho = 0$, the difference in desired contact reduction rates is small as the virus takes off all the way through to the peak of infections, with the major differences occurring after the peak of the virus. It is also worth discussing the two polar cases, namely $\rho = 0$ and $\rho = 1$. When $\rho = 0$, agents base their desired contact reduction rates solely on social learning. Desired contact reduction rates reach a peak after infections peak and remains at this peak. Given the fact that in a no-intervention scenario, deaths are significantly higher, agents are much more likely to be exposed to a death in their network of connections and since we encode the observed desired contact reduction rates of dead agents to be 1 at all times after their death, this causes aggregate desired contact reduction rates to converge to a high level even as infections fade. Contrast this with the case where $\rho = 0$. In this scenario, by default, all agents do not listen to lockdown policies. Agents only increase their desired contact reduction rates, when they become infected. If agents die, their desired contact reduction rates remain 1 throughout, while recovered agents revert to the desired contact rate norm. As a result, desired contact reduction rates are always non-zero due to initial infections and deaths, and there is a slight increase in the desired contact reduction rate driven by an increased desire to reduce contacts among the infected population around the time infections peak.

Now, contrast the behaviour in the lockdown scenario. In all cases, desired contact reduction rates already begin at a high level, given the stringency of the lockdown which affects the private signal. In the special case where $\rho = 1$, desired contact reduction rates simply become the private signal, which is calibrated to the lockdown stringency index. This explains the step changes in desired contact reduction rates. As soon as social learning is introduced ($\rho < 1$), desired contact reduction rates are smoothed. In the case of social learning only, $\rho = 0$, desired contact reduction rates once again plateau at a high level, with the transition to this plateau being more gradual than in the no intervention case, driven by the slower progression of infections. However, unlike the no-intervention scenario, here there are clear differences in desired contact reduction rates early in the simulation when comparing $\rho = 0.05$ to $\rho = 0$. Put differently, under a lockdown, when agents only make use of social learning this leads to a significant increase in desired contact reduction rates, even when compared to a calibration where a small weight is placed on the private signal. This change in desired contact reduction rates happens early enough to meaningfully impact infections, explaining the larger decrease in total infections (16%) than in the no-intervention scenario (7.3%).

The main insight from these results is that the inclusion of social learning has significant effects on the projected progression of the virus. Comparing results from simulations using our calibrated specification, $\rho = 0.15$ with an identical model without social learning $\rho = 1$, shows a 4% decrease in predicted infections and a 17.4% decrease in predicted deaths. The more pronounced reduction in deaths stems from the fact that in the model with social learning, critical cases never exceed healthcare capacity, whereas in the model without social learning, the healthcare system is overburdened for 28 days, resulting in many preventable deaths.

4.4. An alternative model of endogenous behaviour

A primary contribution of this paper is to show that the inclusion of endogenous behaviour in a standard agent-based SEIR model improves the ability of this class of models to match observed patterns of Covid-19 fatalities. Our primary focus is on social learning, but alternative models of endogenous behaviour may also provide an improvement on the standard agent-based SEIR models without learning.

In this section, we consider one alternative type of endogenous behaviour which we term ‘lexicographic’ learning, where information on a salient event—death or critical illness in an agent’s network—determines behaviour over and above all other information.

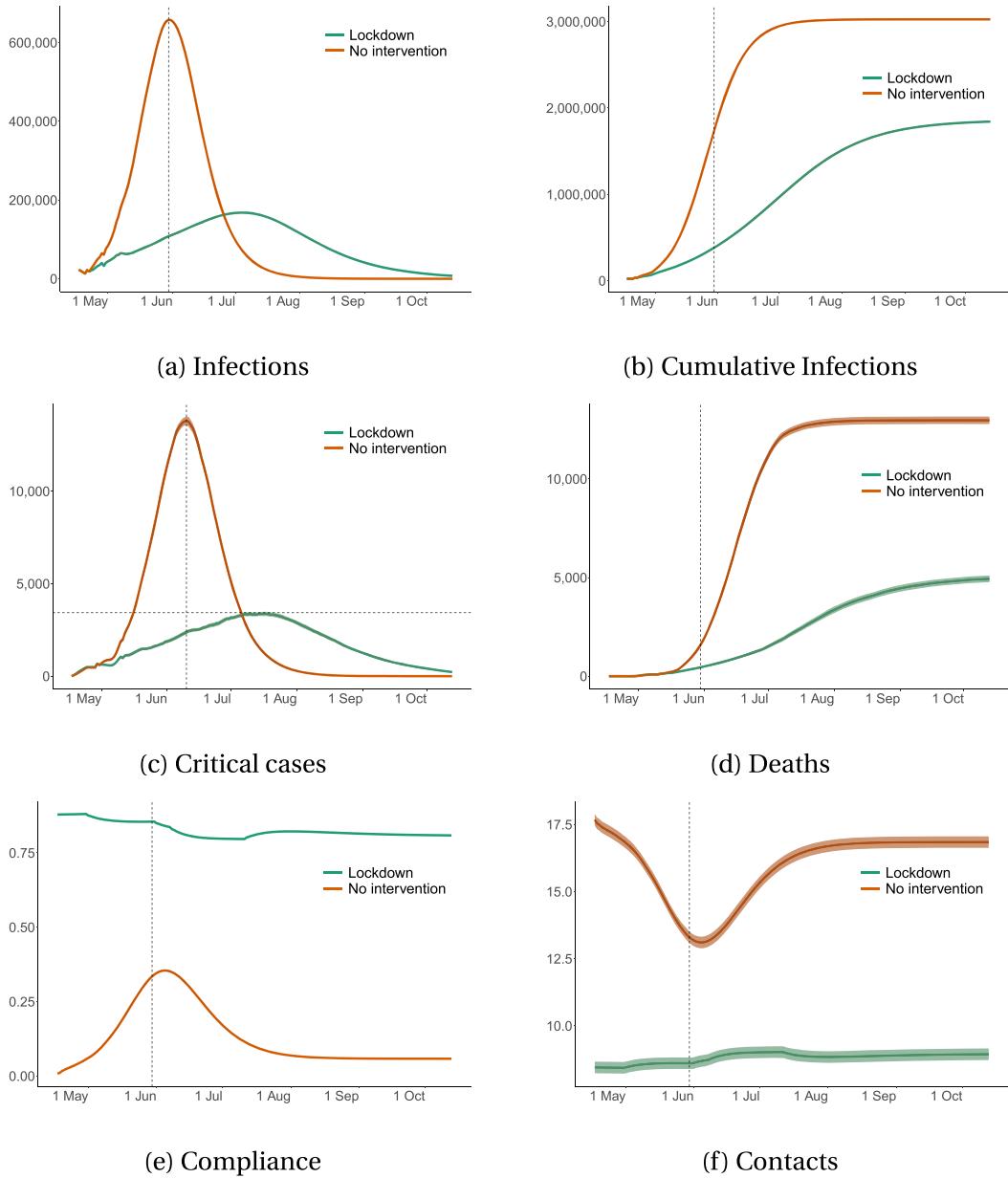


Fig. 4. The impact of a lockdown on key model observables. This figure shows simulation outcomes for Cape Town across the lockdown and no-intervention scenarios. Panels (a)–(d) show the number of infections, cumulative infections, critical cases, and deaths, respectively. Numbers are scaled to the size of Cape Town. Panel (e) shows the level of desired contact reduction rates and panel (f) shows the average number of contacts per household. Panels (a), (b), (c), (e), and (f) include vertical lines indicating the date of peak infections. In Panel (c) we also include a dotted horizontal line indicating hospital capacity in Cape Town. All figures show the mean outcomes taken across 50 simulations.

Under this model of endogenous learning, agents desire the maximum level of isolation whenever there is a Covid-19 critical case or death in their network: When agent j has a neighbour who is critically ill or deceased, they will set their desired degree of contact reduction to the maximum $\tilde{\phi}_j(t) = 1$. Otherwise, they ignore all restrictions and do not choose to reduce their contacts at all, $\tilde{\phi}_j(t) = 0$. The remainder of our model remains unchanged with $\tilde{\phi}_j(t)$ replacing $\phi_j(t)$ in (2) being the only change, which determines the probability of a meeting between two agents.

Under this model of learning, isolation following exposure to the virus via the network of an agent increases much faster than under

a model of social learning, where isolation takes time to occur as agents slowly internalize the infection state of their neighbours. This is what we observe through the simulation results.³⁴ In a model with lexicographic learning, we observe lower infections, deaths and critical cases relative to a model with social learning driven by agents making fewer contacts and having higher levels of compliance. With fewer critical cases, the number of days that critical cases exceed hospital capacity drops sharply from 12 to 3 days, which further reduces the number of deaths. Lexicographic learning also changes the dynamics of the virus, bringing forward the date of peak infections by 23 days and reducing the number of peak infections by 14%, thereby flattening the peak more than in the model with social learning.

Crucially, however, when we compare the ability of a model with lexicographic learning to match observed fatalities in Fig. 5, we see stark differences in both the baseline model without learning and the model with social learning. Even under the best fit, while lexicographic learning produces peak infections which match peak observed excess deaths (something the model with no learning is unable to replicate) it brings the peak of infections too far forward in time. On the other hand, the model with social learning is able to accurately match both the date and number of peak infections. This highlights an important point: while the inclusion of endogenous behaviour can improve the ability of standard agent-based SEIR models to match observed patterns of Covid-19 fatalities, the type of endogenous behaviour included also matters.

5. The effectiveness of different vaccination strategies

Covid-19 network models are particularly useful to study the efficacy of different vaccination strategies because there is no heterogeneous mixing and therefore local herd immunity can be achieved. Globally, the most common vaccination strategy (World Health Organization and others, 2020) is risk-based. Under such a strategy, particularly vulnerable agents—for example the elderly or those with comorbidities—are first in line to be vaccinated after healthcare and front-line workers. The rationale behind this strategy is that a small percentage of the population is particularly susceptible to being severely affected by Covid-19. By vaccinating those first, governments can quickly prevent hospitals from being overwhelmed and effectively reduce the death toll. One possible alternative, put forward by Glover et al. (2022) and Boppart et al. (2022), is to protect those most at risk indirectly by vaccinating those who transmit the virus most widely. The idea is that the spread of the disease can be effectively curbed if those who are most likely to transmit it are vaccinated.³⁵ Using our calibrated model, we compare the efficacy of these two vaccination strategies with a third vaccination strategy where the order in which age categories are vaccinated is random. Specifically, we study how these three vaccination strategies affect fatalities, critical cases, and infections.

We explore a hypothetical scenario in which one-shot perfectly effective vaccines are available at the start of our simulation. This means that each agent who is vaccinated will immediately change her status to recovered. For simplicity, we assume that only susceptible agents will be vaccinated. Furthermore, we assume that, while the vaccine technology is available at the start of the simulation, only 500 vaccines can be produced and distributed each period (i.e., per day). This means that the 100,000 agents in our simulation are vaccinated within 200 days.

The shortage of vaccines implies that a choice needs to be made about which agents get vaccinated first. For our experiments, we first divide agents into their respective age classes: 0 – 10, 10 – 20, ..., 80 + . Regardless of strategy, agents within a selected age class are vaccinated in random order. For the risk-based strategy, age groups that have a higher risk of dying are selected first. Once all agents in the most at-risk age group have been vaccinated, the algorithm proceeds to vaccinate agents in the age group with the next highest risk. For the connection-based strategy, we use the contact matrix from Table 2, which captures the number of connections that each age group has. We use this information as a proxy for an agent's connectedness. Following this strategy, the order of vaccination is from most to least connected age group. Lastly, for the random vaccination strategy, each day 500 agents are selected at random for vaccination, irrespective of age group.

For each of these vaccination scenarios, we simulate our calibrated model with $V = 50$ Monte Carlo simulations, shown in Fig. 6. We find that, while the vaccination speed is too slow to stop the spread of infections by much, critical cases are reduced and fatality curves are flattened considerably by all three. When comparing the three strategies, the connection-based strategy is the most effective at controlling infections. As shown in Table 8, the connection-based strategy leads to 0.9% fewer infections than the random strategy, while the risk-based strategy, on average, leads to a 5.3% increase in cumulative infections. Even when compared to the risk-based strategy, the connection-based strategy only leads to a 5.9% decrease in infections.

In comparison, fatality differences are much bigger per scenario. Compared to the random benchmark, the connection-based vaccination strategy, in which highly-connected low-risk individuals get vaccinated first, results in 42% *more* fatalities than the random vaccination strategy, while the risk-based scenario leads to a reduction of roughly 72%. Comparing the two scenarios leads to an even more stark difference. By changing from a connection-based to a risk-based vaccination strategy, fatalities can be reduced by a little over 80%.

Since both total infections and government stringency hardly change in these scenarios, the differences in desired contact reduction rates and contacts are very small.

³⁴ We report the full set of results for these simulations in Table 9 and Fig. 10 in the Appendix.

³⁵ Thompson (2021); Voysey et al. (2021) present evidence that the vast majority of vaccinated individuals will both become immune to the effects of the virus and will no longer transmit it.

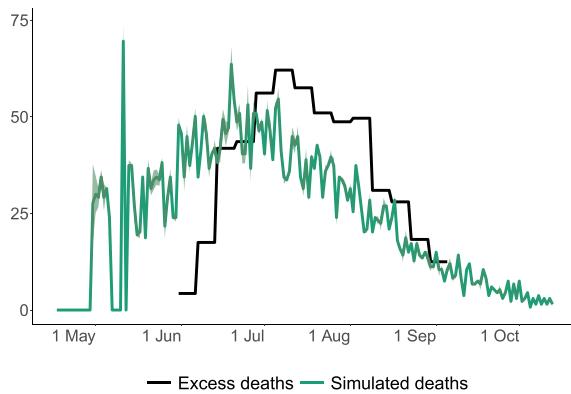


Fig. 5. Simulated deaths and observed excess deaths across different models of learning. This figure shows how simulated deaths compare with excess deaths recorded in Cape Town in a model with lexicographic learning. As in Fig. 2, we use a scenario that represents the best fit obtained using the calibration procedure outlined in Section 3.4. The figure shows the mean outcome taken across 50 simulations.

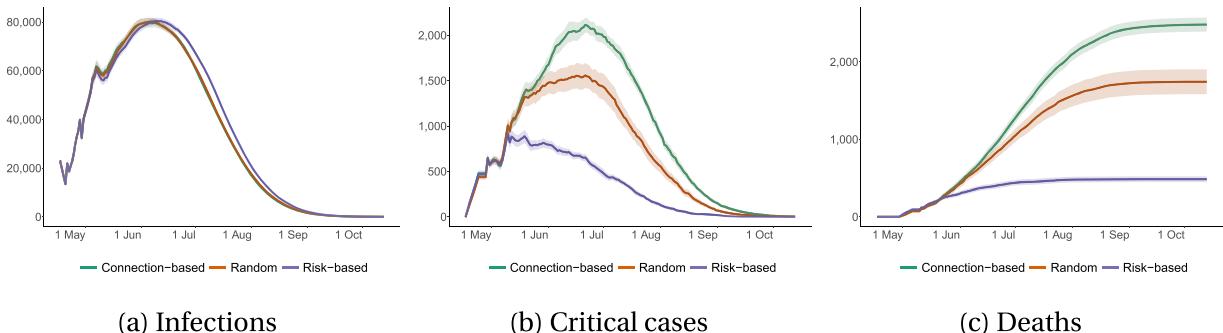


Fig. 6. The efficacy of different vaccination strategies. This figure shows infections, critical cases and deaths of agents over time, under the three vaccination scenarios: connection based; random and risk-based. All figures show mean results across 50 simulations. The shaded regions represent 95% confidence bands.

6. Related literature and conclusion

6.1. Related literature

By adding social learning to an agent-based Covid-19 model, our paper relates to several separate strands of literature.³⁶ When it comes to Covid-19 models, our model is most closely related to other detailed agent-based³⁷ Covid-19 network models, such as those of (Akbarpour et al., 2020) and the Imperial College Model (Ferguson et al., 2020). The latter has been used to inform the Covid-19 strategy of the UK government. Furthermore, our model is related to the macroeconomic ABM literature that has started to include an epidemiological component, such as (Basurto et al., 2020; Delli Gatti and Reissl, 2020). We also use the classic Susceptible-Infected-Recovered (SIR) Kermack and McKendrick (1927) structure that is the backbone of standard differential equation models and implement this within a Covid-19 network model.³⁸ Our major contribution to the literature relative to these models is that we incorporate social learning, in which agents learn from their neighbors' state rather than the global state.³⁹

Consequently, our model is also related to the extensive epidemiological literature that incorporates behaviour and learning into pandemic models in addition to a growing literature on economic epidemiological models (e.g. Toxvaerd (2020), Eichenbaum et al.

³⁶ For a review of the literature, see Mwalili et al. (2020).

³⁷ In this context also known as individual-based, or micro-simulation models

³⁸ Our model could be classified as a Susceptible Exposed Infected without symptoms, Infected with symptoms, in need of Critical care, Recovered, Dead model. However, like most Covid-19 models—see e.g. Mwalili et al. (2020); Calafiore et al. (2020)—we still classify this as a SIR model because each additional compartment can be seen as a sub-category of these three. We also calibrate our model using contact matrices and detailed population data. We choose the agent-based structure because it naturally incorporates a network structure and using heterogeneous agents means that *local* learning is possible.

³⁹ While Akbarpour et al. (2020) include behaviour they do so in a very stylized way by re-estimating the transmission probability and as a result do not model behaviour or learning explicitly.

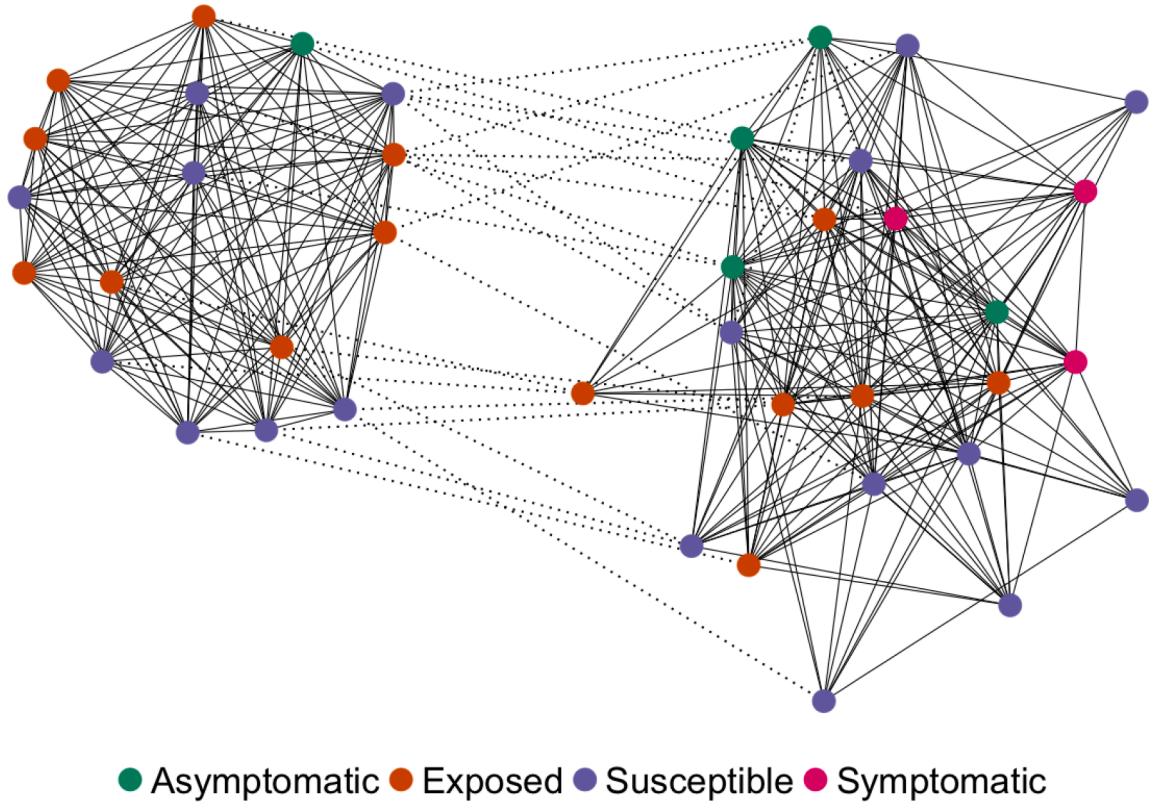


Fig. 7. City network example. This schematic shows a 40-agent, 2-district city network at $t = 14$ generated by our algorithm. pure Household edges are dotted and other edges are continuous (these may overlap). The nodes are agents and the colours mark differences in their epidemiological status p_j .

(2020), Krueger et al. (2020), Dasaratha (2020)) which incorporates optimising behaviour.⁴⁰ The optimising mechanisms in this class of models typically involve agents who face a trade-off between wanting to leave home to earn an income to satisfy their economic needs and wanting to stay at home to minimize the risk and cost of infection, which is typically a function of aggregate infections. Our model is similar to these efforts in introducing behaviour to an epidemiological model. However, rather than individuals optimising their behaviour based on an aggregate state of infections, learning is social, and therefore local, in our model.

Our paper is also related to the literature on social learning in networks (Golub and Sadler (2016); DeGroot (1974); Golub and Jackson (2010)) and the epidemiological literature using network models to model the spread of diseases (Danon et al. (2011); Keeling and Eames (2005)). Our findings that social learning limits connections and reduces infections is consistent with empirical evidence from the recent and growing literature on the role of social networks in individual decision making and behaviour (Bailey et al., 2018). In particular it aligns with research by Bailey et al. (2021) and Charoenwong et al. (2020) who find that increases in Covid-19 cases within an individual's social network leads to reduced mobility and an increased desire to reduce connections.⁴¹ In this literature, our paper is also closely related to Makridis and Wang (2020) who study the impact of social learning on consumption during a pandemic within a consumption and savings network-based model. They show that, compared to a model without social learning, including social learning leads to a greater reduction in consumption, as learning leads to an increase in perceived infections by agents as they internalise the infectious state of the nodes in their network. However, these authors focus explicitly on the role of social learning on consumption during a pandemic and do not allow for social learning to affect the progression of the disease. In this regard, our model contributes to this literature by revealing how social learning affects disease progression. Our work is also distinguished by how we allow social learning to influence behaviour. In Makridis and Wang (2020), agents learn about the true state of infections solely through the infection state of the nodes in their network, in the vein of Golub and Jackson (2010). Our model, rather, closely follows the implementation of social learning in Dasaratha et al. (2020), whereby agents receive a noisy private signal of the state of the world and use a weighted combination of this noisy private signal and a social signal, derived from social learning, to form a decision to reduce connections.

Finally, our paper extends the literature on optimal vaccination strategies which largely explores two broad vaccination strategies:

⁴⁰ See Verelst et al. (2016) for a survey of this literature

⁴¹ Glaeser et al. (2020) document a causal link between mobility and the spread of Covid-19.

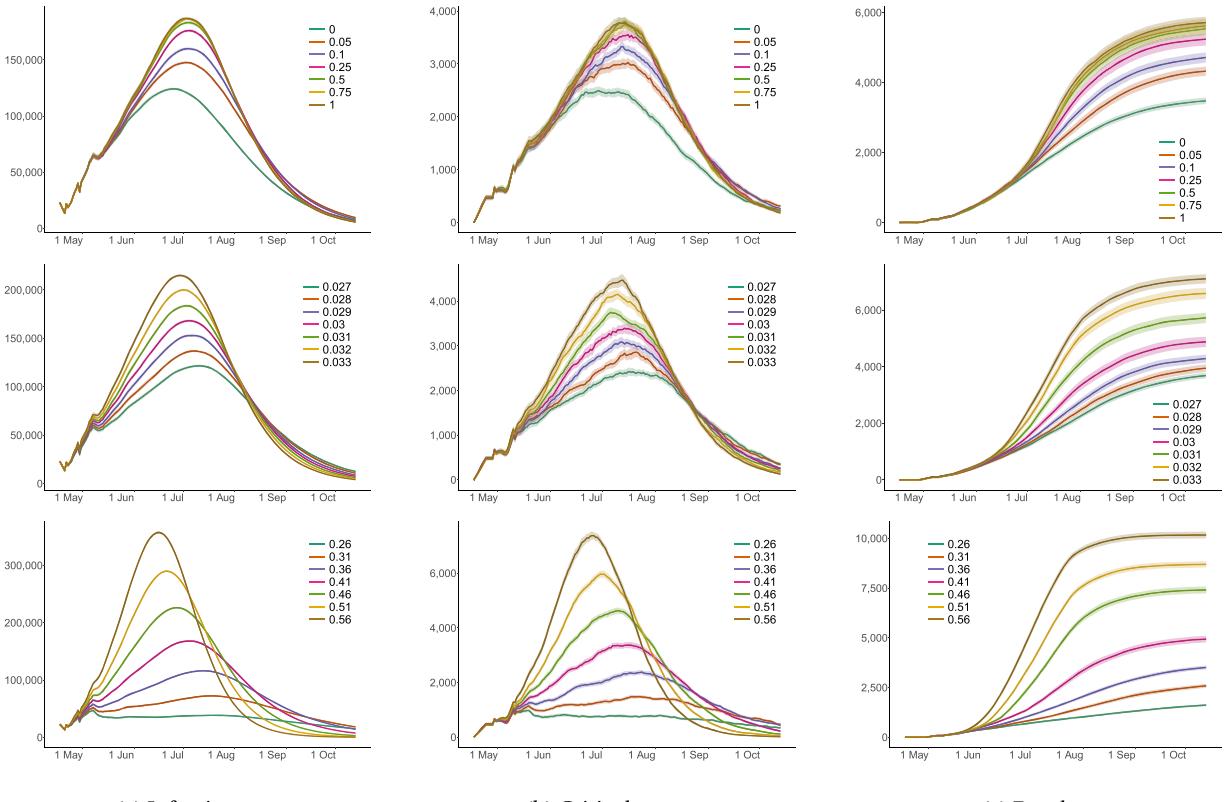


Fig. 8. Key observables as a function of varying key parameters. This figure shows infections, critical cases and deaths for Cape Town in a Lockdown scenario, for different values of the weight, ρ , of the private signal, $\zeta_j(t)$, in the agent's desired contact rate (top), the probability of transmission, π^E (middle) and the share of contacts deemed essential, ω . All figures show the mean outcome taken across 50 simulations. For reference, in our baseline specifications, $\rho = 0.15$, $\pi^E = 0.030$ and $\omega = 0.46$. The shaded area indicates the standard deviation.

Vaccinating those at higher risk of hospitalisation and death, and vaccinating those with the highest likelihood of spreading the virus. Both of these strategies amount to age-based vaccination strategies, given age is positively correlated to mortality risk and negatively correlated with mobility and interactions. Economic epidemiological models by [Glover et al. \(2022\)](#) and [Boppert et al. \(2022\)](#) find that vaccinating the elderly (and thus, those with greater mortality risk) results in the greatest reduction in mortality, a finding echoed in [Bubar et al. \(2021\)](#). While we evaluate two similar vaccination strategies in this paper, we innovate over the existing literature in two ways. Firstly, in order to remain tractable, the economic epidemiological models consider simplified scenarios where the population group is divided into two age groups, young and old. In our model, the population is divided into nine age-groups allowing for a richer and more realistic treatment of vaccination strategies. Secondly, our model features heterogeneous contact rates calibrated using an age-based social contract matrix, and as a result, we can explore vaccination experiments where we target the age groups who make the most contacts, as opposed to using age to proxy for the number of contacts made, which would amount to targeting the youngest age groups first. This makes our model flexible to being applied in locations where the relationship between age and contacts is not as strongly positively correlated, or, is highly non-linear.

6.2. Conclusion

Our simulations show that social learning helps explain the unusually flat and short Covid-19 curves in Cape Town. The social learning signal is strong when there are many infections in the neighbourhood of an agent. This decreases the spread of Covid-19 at the peak. Further, since agents desire to reduce contacts for an extended periods, social learning speeds up the decline of the virus. This is different from the effect of lockdown measures that reduce the transmission probability, which reduces total infections by spreading them out over time. Nevertheless, social learning alone was not enough to flatten the curve in Cape Town so that hospital capacity was not breached, increasing the excess fatality rate. A lockdown was still needed to flatten the curve sufficiently. However, a reduction in the number of contacts as a consequence of a lockdown is smaller when taking into account social learning because otherwise agents will also reduce some contacts voluntarily. Finally, we show that in our calibrated model the strategy of vaccinating the elderly first effectively reduces the number of fatalities, even though it leads to more infections compared to both a random and connection-based strategy. That being said, our model does not consider the possibility of re-infection which can come about through the emergence of

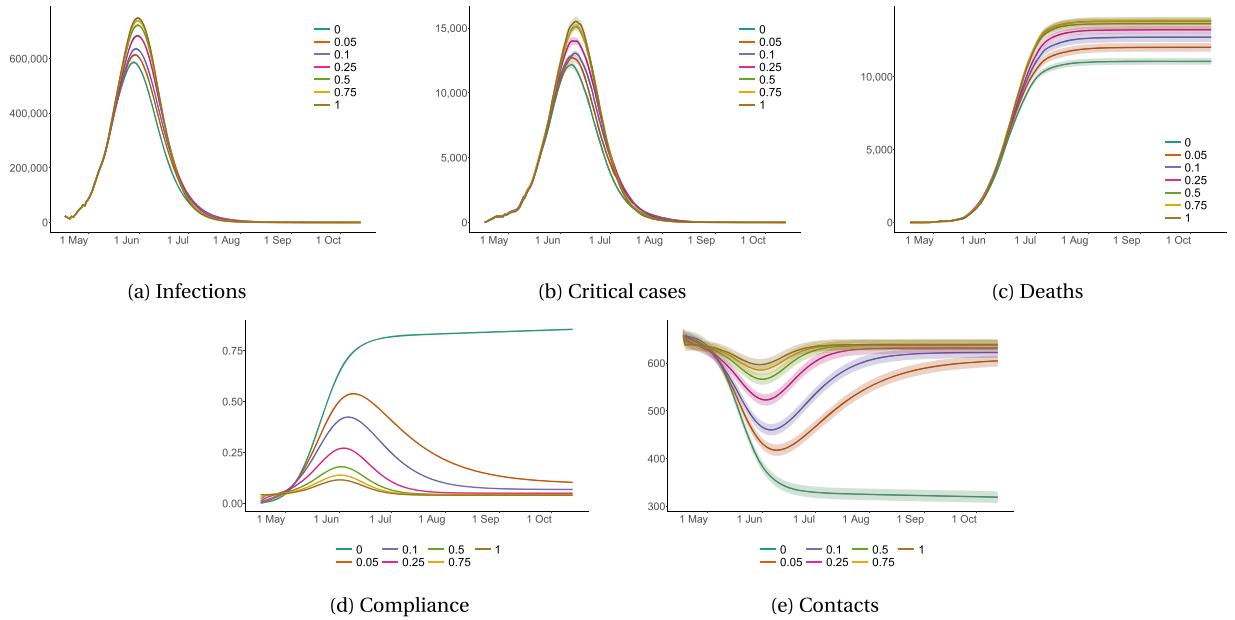


Fig. 9. Key observables as a function of a changing weight of the private signal. This figure shows infections, critical cases, deaths, desired contact reduction rates and contacts for Cape Town in a no-intervention scenario, for different values of the weight, ρ , of the private signal, $\zeta_j(t)$, in the agent's contact reduction decision. All figures show the mean outcome taken across 50 simulations. Shaded areas indicate the standard deviation. For reference, in our baseline specifications, our calibrated value of ρ is 0.15.

Covid-19 variants that escape existing immunity and/or the possibility that immunity may wane across time. Existing evidence suggests that Covid-19 immunity results in low risk of subsequent infection with antigenically similar variants for at least 6 months, irrespective of whether the immunity is the result of infection or vaccination (Centre for Disease Control and Prevention, 2021). As such, we see the time scale of the validity of our results to be between 6 and 8 months.

This paper contains three important insights for policymakers. First, standard epidemiological models without social learning are biased in overstating either the height or length of the infections. These models therefore overestimate deaths due to Covid-19 because they do not consider that in the no-intervention scenario people may voluntarily reduce their contacts as a result of social learning. Second, models without social learning overestimate the effect of lockdown measures in terms of reduced contacts. Finally, a risk-based vaccination strategy is highly effective in reducing fatalities, even in cities with relatively young populations such as Cape Town.

The model presented in this paper offers many possibilities for extensions which will make it applicable to particular policy concerns. For example, agent behaviour can easily be modified to include some sort of trade-off between contacts and age-specific risks. Furthermore, because of its heterogeneity, the model lends itself well to policy experiments that target specific locations or age profiles.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property. We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author.

Data availability

Data will be made available on request.

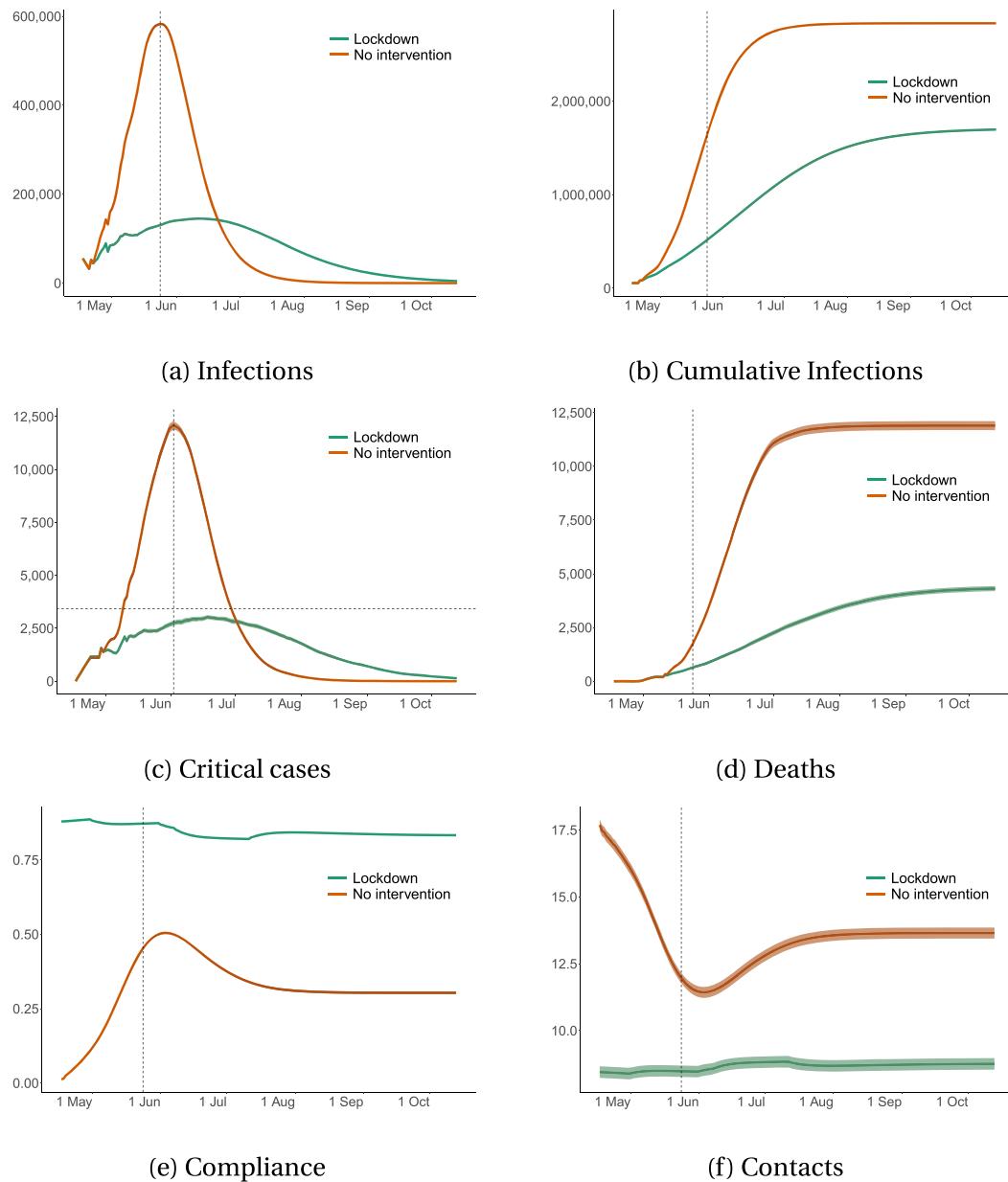


Fig. 10. The impact of a lockdown on key model observables under lexicographic learning. This figure shows simulation outcomes for Cape Town across the lockdown under lexicographic learning. Panels (a)–(d) show the number of infections, cumulative infections, critical cases, and deaths, respectively. Numbers are scaled to the size of Cape Town. Panel (e) shows the level of desired contact reduction rates and panel (f) shows the average number of contacts per household. Panels (a), (b), (c), (e), and (f) include vertical lines indicating the date of peak infections. In Panel (c) we also include a dotted horizontal line indicating hospital capacity in Cape Town. All figures show the mean outcomes taken across 50 simulations.

Appendix A

Table 3

Appendix B. Calibration Details

B1. Household interactions

As mentioned in [Section 2.2](#) we implement an algorithm to calibrate household interactions in our model such that the sets of household edges in each district match (i) the size and age distributions in districts are representative in aggregate, and also the (ii)

Table 3

Simulation results under both lockdown and no-intervention scenarios We include (i) Total Infections (ii) Total Deaths (iii) Total Recoveries (iv) Peak Infections (v) Peak Critical (vi) Peak Infections (Day) - the day on which peak infections are reached (vii) Peak Critical (Day) - the day on which peak critical is reached, (viii) End of simulation infections - the number of agents still infected at the end of the simulation, (ix) Critical > Capacity (Days) - the number of days in which the number of critical cases exceeds hospital capacity, (x) the average number of contacts and (xi) the average level of compliance. We report minimum, mean and maximum values, as well as the standard deviation across 50 simulations.

Outcome	Min	Mean	Max	SD
Total Infections	1,801,982	1,843,002	1,889,125	22,008
Total Deaths	3,740	4,934	6,582	591
Total Recoveries	1,788,293	1,830,547	1,879,363	22,474
Peak Infections	155,735	167,847	176,941	4,816
Peak Critical	2,655	3,368	4,413	409
Peak Infections (Day)	67	79	87	4
Peak Critical (Day)	54	86	105	10
End of simulation infections	5,124	7,520	10,098	1,243
Critical > Capacity (Days)	0	12	30	9
Contacts	7.3	8.87	10.61	0.76
Compliance	0.82	0.82	0.82	0
(a) Lockdown				
Min	Mean	Max	SD	
3,010,721	3,026,825	3,041,277	7,938	
11,818	12,973	14,736	665	
2,997,893	3,013,849	3,028,187	7,782	
639,881	657,681	675,149	6,661	
11,744	13,803	16,344	827	
42	43	46	1	
48	54	57	2	
0	4	37	11	
47	51	54	1	
14.76	16.31	18.06	0.76	
0.1	0.1	0.1	0	
(b) No-intervention				

Table 4

Simulation results when varying the extent of social learning in a lockdown scenario This table shows model outcomes for Cape Town for different values of the weight, ρ , of the private signal, $\zeta_j(t)$, in the agent's compliance decision in a lockdown scenario. We include (i) Total Infections (ii) Total Deaths (iii) Total Recoveries (iv) Peak Infections (v) Peak Critical (vi) Peak Infections (Day) - the day on which peak infections are reached (vii) Peak Critical (Day) - the day on which peak critical is reached, (viii) End of simulation infections - the number of agents still infected at the end of the simulation, (ix) Critical > Capacity (Days) - the number of days in which the number of critical cases exceeds hospital capacity, (x) the average number of contacts and (xi) the average level of compliance. We report minimum, mean and maximum values, as well as the standard deviation across 50 simulations.

Outcome	$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1.0$
Total Infections	1,450,666	1,726,480	1,803,981	1,843,002	1,875,291	1,903,463	1,915,957	1,919,512
Total Deaths	3,477	4,326	4,714	4,934	5,236	5,534	5,623	5,710
Total Recoveries	1,439,665	1,712,503	1,790,852	1,830,547	1,863,480	1,891,933	1,904,569	1,908,184
Peak Infections	124,312	147,684	159,934	167,847	176,092	183,248	186,619	186,885
Peak Critical	2,495	3,028	3,331	3,368	3,555	3,800	3,749	3,778
Peak Infections (Day)	69	76	77	77	78	77	77	77
Peak Critical (Day)	73	90	87	90	90	88	88	85
End of simulation infections	7,523	9,651	8,415	7,520	6,575	5,996	5,765	5,618
Critical > Capacity (Days)	0	0	0	0	16	22	25	28
Contacts	8.02	8.75	8.83	8.87	8.9	8.92	8.93	8.93
Compliance	0.94	0.84	0.82	0.82	0.81	0.81	0.81	0.81

household size and age distributions are representative within each district.

We implement this algorithm as follows. The set of agents in district w , denoted \mathcal{N}_w , is partitioned into a disjoint set of households \mathcal{H}_w of different sizes. Let $\mathcal{N}_{h,w}$ be the set of agents in household h in district w (i.e. $\bigcup_{h \in \mathcal{H}_w} \mathcal{N}_{h,w} = \mathcal{N}_w$) with household size $N_{h,w} = |\mathcal{N}_{h,w}|$.

Table 5

Simulation results when varying the extent of social learning in a no-intervention scenario This table shows model outcomes for Cape Town for different values of the weight, ρ , of the private signal, $\zeta_j(t)$, in the agent's compliance decision in a no-intervention scenario. We include (i) Total Infections (ii) Total Deaths (iii) Total Recoveries (iv) Peak Infections (v) Peak Critical (vi) Peak Infections (Day) - the day on which peak infections are reached (vii) Peak Critical (Day) - the day on which peak critical is reached, (viii) End of simulation infections - the number of agents still infected at the end of the simulation, (ix) Critical > Capacity (Days) - the number of days in which the number of critical cases exceeds hospital capacity, (x) the average number of contacts and (xi) the average level of compliance. We report minimum, mean and maximum values, as well as the standard deviation across 50 simulations.

Outcome	$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1.0$
Total Infections	2,650,236	2,858,741	2,964,293	3,026,770	3,094,962	3,166,939	3,194,733	3,209,114
Total Deaths	11,052	12,016	12,710	12,853	13,220	13,804	13,837	13,633
Total Recoveries	2,639,184	2,846,712	2,951,578	3,013,914	3,081,743	3,153,135	3,180,897	3,195,481
Peak Infections	586,456	613,925	635,849	657,669	684,372	723,142	739,804	749,504
Peak Critical	12,167	12,730	13,054	13,739	14,014	15,087	15,173	15,518
Peak Infections (Day)	42	42	43	44	44	44	44	44
Peak Critical (Day)	52	52	54	54	52	54	55	54
End of simulation infections	0	13	4	3	0	0	0	0
Critical > Capacity (Days)	47	49	51	50	50	49	50	49
Contacts	10.34	14.56	15.49	15.89	16.27	16.67	16.8	16.87
Compliance	0.67	0.24	0.16	0.13	0.1	0.07	0.06	0.05

Table 6

Simulation results when varying the transmission probability in a lockdown scenario This table shows model outcomes for Cape Town for different values of the disease probability of transmission, π^E in a lockdown scenario. We include (i) Total Infections (ii) Total Deaths (iii) Total Recoveries (iv) Peak Infections (v) Peak Critical (vi) Peak Infections (Day) - the day on which peak infections are reached (vii) Peak Critical (Day) - the day on which peak critical is reached, (viii) End of simulation infections - the number of agents still infected at the end of the simulation, (ix) Critical > Capacity (Days) - the number of days in which the number of critical cases exceeds hospital capacity, (x) the average number of contacts and (xi) the average level of compliance. We report minimum, mean and maximum values, as well as the standard deviation across 50 simulations.

Outcome	$\pi^E = 0.027$	$\pi^E = 0.028$	$\pi^E = 0.029$	$\pi^E = 0.030$	$\pi^E = 0.031$	$\pi^E = 0.032$	$\pi^E = 0.033$
Total Infections	1,528,530	1,642,133	1,745,812	1,843,391	1,929,874	2,018,460	2,095,866
Total Deaths	3,684	3,944	4,285	4,876	5,727	6,587	7,104
Total Recoveries	1,512,348	1,627,490	1,732,720	1,831,122	1,918,086	2,006,725	2,084,686
Peak Infections	121,602	136,931	152,983	168,051	183,468	200,145	214,966
Peak Critical	2,423	2,858	3,090	3,394	3,751	4,155	4,470
Peak Infections (Day)	85	81	81	78	77	75	73
Peak Critical (Day)	91	95	86	88	80	84	87
End of simulation infections	12,498	10,699	8,808	7,393	6,062	5,148	4,076
Critical > Capacity (Days)	0	0	0	0	23	36	44
Contacts	8.87	8.87	8.87	8.87	8.86	8.86	8.86
Compliance	0.82	0.82	0.82	0.82	0.82	0.82	0.82

Table 7

Simulation results when varying the transmission probability in a no-intervention scenario This table shows model outcomes for Cape Town for different values of the disease probability of transmission, π^E in a no-intervention scenario. We include (i) Total Infections (ii) Total Deaths (iii) Total Recoveries (iv) Peak Infections (v) Peak Critical (vi) Peak Infections (Day) - the day on which peak infections are reached (vii) Peak Critical (Day) - the day on which peak critical is reached, (viii) End of simulation infections - the number of agents still infected at the end of the simulation, (ix) Critical > Capacity (Days) - the number of days in which the number of critical cases exceeds hospital capacity, (x) the average number of contacts and (xi) the average level of compliance. We report minimum, mean and maximum values, as well as the standard deviation across 50 simulations.

Outcome	$\pi^E = 0.027$	$\pi^E = 0.028$	$\pi^E = 0.029$	$\pi^E = 0.030$	$\pi^E = 0.031$	$\pi^E = 0.032$	$\pi^E = 0.033$
Total Infections	2,894,028	2,941,954	2,985,157	3,028,015	3,064,321	3,102,929	3,133,917
Total Deaths	12,092	12,441	12,652	12,750	12,915	13,217	13,419
Total Recoveries	2,881,907	2,929,512	2,972,501	3,015,252	3,051,406	3,089,712	3,120,497
Peak Infections	574,730	603,624	630,594	657,482	682,598	709,875	734,193
Peak Critical	11,700	12,448	13,090	13,834	13,905	14,616	15,081
Peak Infections (Day)	46	45	44	44	43	42	42
Peak Critical (Day)	57	54	55	54	53	53	52
End of simulation infections	28	1	4	12	0	0	1
Critical > Capacity (Days)	52	51	51	50	50	49	50
Contacts	15.97	15.96	15.95	15.95	15.95	15.95	15.94
Compliance	0.12	0.13	0.13	0.13	0.13	0.13	0.13

Table 8

Vaccination experiments. This table shows model outcomes for Cape Town under different vaccination scenarios in a lockdown scenario. We include (i) Total Infections (ii) Total Critical Cases (iii) Total Deaths (iv) Total Recoveries (v) Contacts and (vi) Compliance. We report mean values across 50 simulations.

Outcome	Connection-based	Random	Risk-based
Total Infections	902,706	911,008	959,736
Total Critical cases	545	413	201
Total Deaths	2,481	1,743	485
Total Recoveries	3,470,921	3,482,767	3,504,965
Contacts	8.9	8.9	8.9
Compliance	0.81	0.81	0.81

Table 9

Simulation results under social learning and lexicographic learning. We include (i) Total Infections (ii) Total Deaths (iii) Total Recoveries (iv) Peak Infections (v) Peak Critical (vi) Peak Infections (Day) - the day on which peak infections are reached (vii) Peak Critical (Day) - the day on which peak critical is reached, (viii) End of simulation infections - the number of agents still infected at the end of the simulation, (ix) Critical > Capacity (Days) - the number of days in which the number of critical cases exceeds hospital capacity, (x) the average number of contacts and (xi) the average level of compliance. We report minimum, mean and maximum values, as well as the standard deviation across 50 simulations.

Outcome	Min	Mean	Max	SD
Total Infections	1,801,982	1,843,002	1,889,125	22,008
Total Deaths	3,740	4,934	6,582	591
Total Recoveries	1,788,293	1,830,547	1,879,363	22,474
Peak Infections	155,735	167,847	176,941	4,816
Peak Critical	2,655	3,368	4,413	409
Peak Infections (Day)	67	79	87	4
Peak Critical (Day)	54	86	105	10
End of simulation infections	5,124	7,520	10,098	1,243
Critical > Capacity (Days)	0	12	30	9
Contacts	7.3	8.87	10.61	0.76
Compliance	0.82	0.82	0.82	0
(a) Social learning				
Min	Mean	Max	SD	
1,640,001	1,696,861	1,753,175	22,758	
3,590	4,311	5,423	407	
1,630,165	1,688,098	1,744,386	23,067	
134,641	145,075	155,174	5,038	
2,319	3,028	3,703	339	
45	56	67	5	
45	63	95	11	
3,067	4,452	6,620	923	
0	3	15	4	
7.1	8.66	10.41	0.76	
0.84	0.85	0.85	0	
(b) Lexicographic learning				

Agents in each household are all connected with one another. The set of households in each district, \mathcal{H}_w , is constructed iteratively, by randomly drawing household sizes $N_{h,w}$ from the empirical observed distribution of household sizes specific to each district until $\sum_{h \in \mathcal{H}_w} N_{h,w} = N_w$.⁴² This yields, for each district w , a set of households \mathcal{H}_w of different sizes so that there are a total of $H_w = |\mathcal{H}_w|$ households in district w , representative of the observed household size distribution for each district.

After initializing a realistic district-level distribution of households of varying sizes, we then proceed to assign each agent $j \in \mathcal{N}$ to a household \mathcal{H}_w in order to match the observed within-household age probability distribution F^{hc} . We proceed in two steps.

First, for each household h in district w , we select a household head as follows: From the set of agents in the district, we randomly select an agent $j \in \mathcal{N}_w^-$ (sequentially over households, without replacement) to be head of household h , where \mathcal{N}_w^- denotes the subset of all agents in \mathcal{N}_w unassigned to a household. The selected household head has age-category a_j , which was assigned across agents, by district to match available data on age-distributions per district (as described in Section 2.1). This process continues until each of the H_w households in region w has a household head.

⁴² We observe the empirical distribution of households sizes for each district using national census data. Since the algorithm is stochastic and we use representative agent populations smaller than the actual populations in the city modelled, there are some additional algorithmic features that ensure that (i) the maximum household size randomly drawn remains smaller than the remaining number of agents to be assigned to a household at every point in the algorithm, and (ii) that the final number of agents in a district correspond to the proportional size of that district given the ratio of total modelled agents to the actual population of the city.

Second, for household h with a household head of age a_h and household size $N_{h,w}$, we select additional household members sequentially, randomly, and without replacement from \mathcal{N}_w^- according to the probability distribution constructed from the empirical household contact matrix F^{hc} until there are $N_{h,w}$ members, then move on to the next household.

B2. Non-household interactions

In addition to matching key properties of household interactions, our model also features key properties of non-household interactions, namely heterogeneous contact rates by age and realistic travel patterns.

We initialise non-household interactions as follows. The set of agents with whom agent j forms non-household edges is constructed in two steps. First, agent j selects a *destination* district w' according to the empirical probabilities in F^{tv} . Iterating over all agents, this yields a *population of potential contacts* in each district that consists of all agents from other districts who selected it as a destination. We denote the population of potential contacts in district w' as \mathcal{N}_w^{tv} . Second, for each agent j with destination district w' , we create edges to agents that are randomly selected from $\mathcal{N}_{w'}^{tv} \setminus \{j\}$, without replacement, according to the probability distribution implied by the non-household contact matrix F^{oc} , until N_j^{oc} new edges have been created.⁴³

To summarize, the algorithm generates the network \mathcal{G} which contains, for each agent j a neighbourhood of all contacts $\mathcal{N}_j \subset \mathcal{N}$ to or from whom the virus can be transmitted. Let agent j be in household h , then \mathcal{N}_j is composed of two components that can be impacted differently by lockdown regulations: a neighbourhood of household contacts \mathcal{N}_j^{hc} and a neighbourhood of non-household contacts \mathcal{N}_j^{oc} , i.e. $\mathcal{N}_j = \mathcal{N}_j^{hc} \cup \mathcal{N}_j^{oc}$.

B3. Travel pattern calibration

In order to calibrate our model to realistic travel patterns, we use the 2013 National Household Travel Survey, a nationally representative travel survey. Importantly, for our purposes, the survey records where respondents live, where they travel to for work and/or education, the frequency of travel, and the time spent travelling. The travel survey allocates respondents to Transport Analysis Zones of which there are 18 in Cape Town. Our challenge then is to relate the 116 wards we use in this paper, to the 18 travel regions. We illustrate a simplified schematic of how these two structures relate in Fig. 11.

Our goal is to create a ward-level travel probability matrix using regional travel data. To implement this, we allocate region-level flows to wards in proportion to the size of both the origin and destination ward population as a share of their respective total regional population. We illustrate this process in Fig. 12.

We encounter three scenarios. The first involves travel between two wards, which are both perfectly located within two different regions, as is the case with Ward 2, located in Region 1, and Ward 3, located in Region 2. We construct the flow of people between these wards as the product of:

- $\frac{\mathcal{N}_2}{\gamma \mathcal{N}_1 + \mathcal{N}_2}$: The size of Ward 2's population relative to the total population of Region 1;
- $\frac{\mathcal{N}_3}{\mathcal{N}_3 + \mathcal{N}_4}$: The size of Ward 3's population relative to the total population of Region 3;
- $f_{1,2}$: The flow of people between Region 1 and Region 2;

We introduce a parameter γ which scales the population of Ward 1, \mathcal{N}_1 . This accounts for a scenario where Ward 1 overlaps with two regions, as illustrated in Fig. 11. γ then reflects the percentage of geographic overlap with Ward 1 and Region 1. We use this geographic overlap to assign the ward population to the respective region.⁴⁴

Our second scenario involves travel between two wards, in which only one ward is perfectly located within a region, as is the case between Ward 1, located in Region 1 and Region 3, and Ward 3, located in Region 2. In this scenario, we map two flows⁴⁵

- $\frac{\gamma \mathcal{N}_1}{\gamma \mathcal{N}_1 + \mathcal{N}_2} \frac{\mathcal{N}_3}{\mathcal{N}_3 + \mathcal{N}_4} f_{1,2}$: The flow of people from the part of Ward 1 located in Region 1;
- $\frac{(1-\gamma) \mathcal{N}_1}{(1-\gamma) \mathcal{N}_1 + \mathcal{N}_5} \frac{\mathcal{N}_3}{\mathcal{N}_3 + \mathcal{N}_4} f_{3,2}$: The flow of people from the part of Ward 1 located in Region 3;

Our third scenario represents a case where a respondent reports living and working/attending school in the same region. Within this scenario, we need to decide how to assign respondents who are likely to live and travel within the same ward versus respondents who are likely to live in one ward but travel to another ward *within the same region*. In our illustrative example, such as case occurs between Ward 1 and Ward 2 and between Ward 1 and Ward 5. In order to allocate these flows, we introduce a new parameter $\delta_{i,i}$ which takes a value between 0 and 1, indicating the likelihood that an individual respondent lives and works or attends schools in different wards within the same region.

⁴³ The algorithm is robust to implementations with smaller populations where it is possible that $N_j^{oc} > |\mathcal{N}_w^{tv}|$. See Algorithm 4.

⁴⁴ The inherent assumption here is that the ward-level population is distributed evenly across the ward.

⁴⁵ In the event that the destination ward also overlaps multiple regions, we can modify these flow equations with an additional overlap parameter which scales the destination wards population.

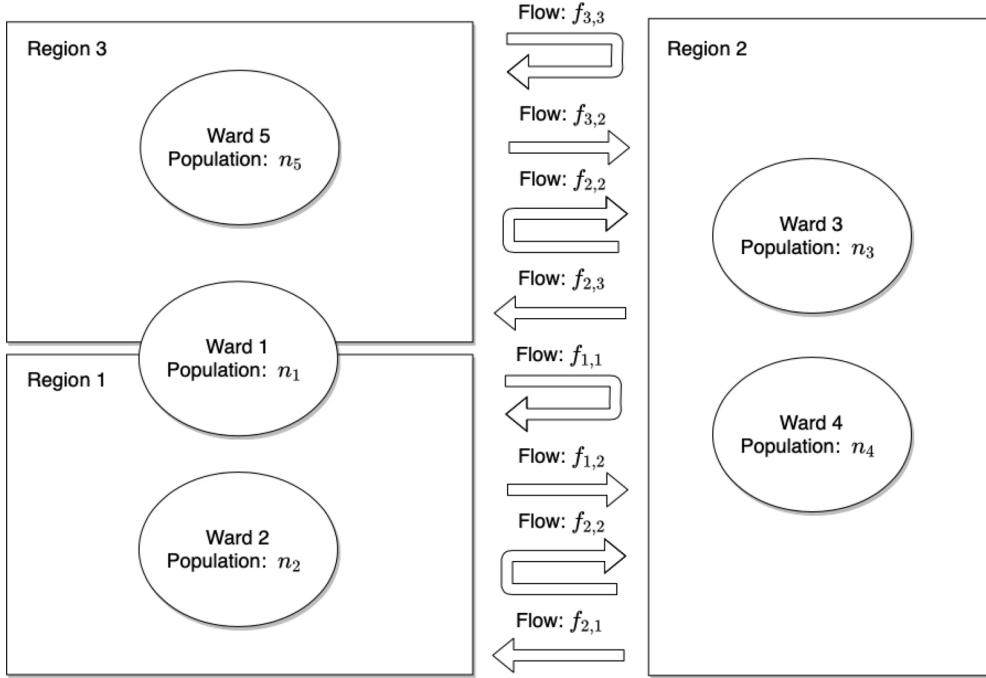


Fig. 11. Travel data illustrative schematic. This figure illustrates how the travel survey data we use corresponds to wards, our geospatial structure. The travel survey data records flows of people between regions. Respondents are asked in which region they reside and then in which region they commute to for work and/or education reasons. There are 18 regions in Cape Town, while there are 116 wards. In this illustrative schematic, we have 3 regions where travel only occurs between Region 1 and Region 2, and between Region 2 and Region 3. In some cases, wards will fit perfectly into a single region, as is the case with Wards 2 through 5. However, there may be a case where a ward overlaps with two regions, as is the case with Ward 1 which overlaps Regions 1 and 2.

We leverage a question in the travel survey, which asks how much time a daily commute takes. Using this question, we take the distribution of travel times for all respondents who live and work or attend school in the same region and assign any individual who has a travel time greater than the 25th percentile of this distribution as a *cross-ward traveller*, assigning the rest of respondents as *within-ward travellers*. $\delta_{i,i}$ then reflects the share of within-region travellers, who are likely to be *cross-ward travellers*. The assumption here is that the likelihood of cross-ward travel increases with travel time. Using this, we now assign 4 travel flows for Ward 1

- $\frac{\gamma \cdot r'_1}{\gamma \cdot r'_1 + \gamma' \cdot r'_2} f_{1,1} \delta_{1,1}$: The flow of people from the part of Ward 1 located in Region 1 to Ward 2;
- $\frac{\gamma \cdot r'_1}{\gamma \cdot r'_1 + \gamma' \cdot r'_2} f_{1,1} (1 - \delta_{1,1})$: The flow of people from the part of Ward 1 located in Region 1 who do not travel;
- $\frac{(1-\gamma) \cdot r'_1}{(1-\gamma) \cdot r'_1 + r'_3} f_{3,3} (1 - \delta_{3,3})$: The flow of people from the part of Ward 1 located in Region 3 to Ward 5;
- $\frac{(1-\gamma) \cdot r'_1}{(1-\gamma) \cdot r'_1 + r'_3} f_{3,3} (1 - \delta_{3,3})$: The flow of people from the part of Ward 1 located in Region 3 who do not travel;

Following this approach, we obtain a ward-level travel flow between each ward. To convert this to travel probability, we normalize each outgoing travel flow from a given ward by the total outgoing flows from that same ward. For Ward 1 then, the probability of travelling to Ward 2 can be calculated as follows:

- Flow from Ward 1 to Ward 2: $f_{w1,w2}$;
- Total flows from Ward 1: $\sum_{k=1}^l f_{k,l}$;
- Probability of travel between Ward 1 and Ward 2: $c_{1,2} = \frac{f_{w1,w2}}{\sum_{k=1}^l f_{k,l}}$;

Appendix C. Pseudo Code: Initialisation Algorithm

Our algorithms can be split up into two main algorithms. First, we use the initialisation [Algorithm 1](#) that consists of three sub-algorithms.

Next, we describe each of these sub-algorithms in more detail.

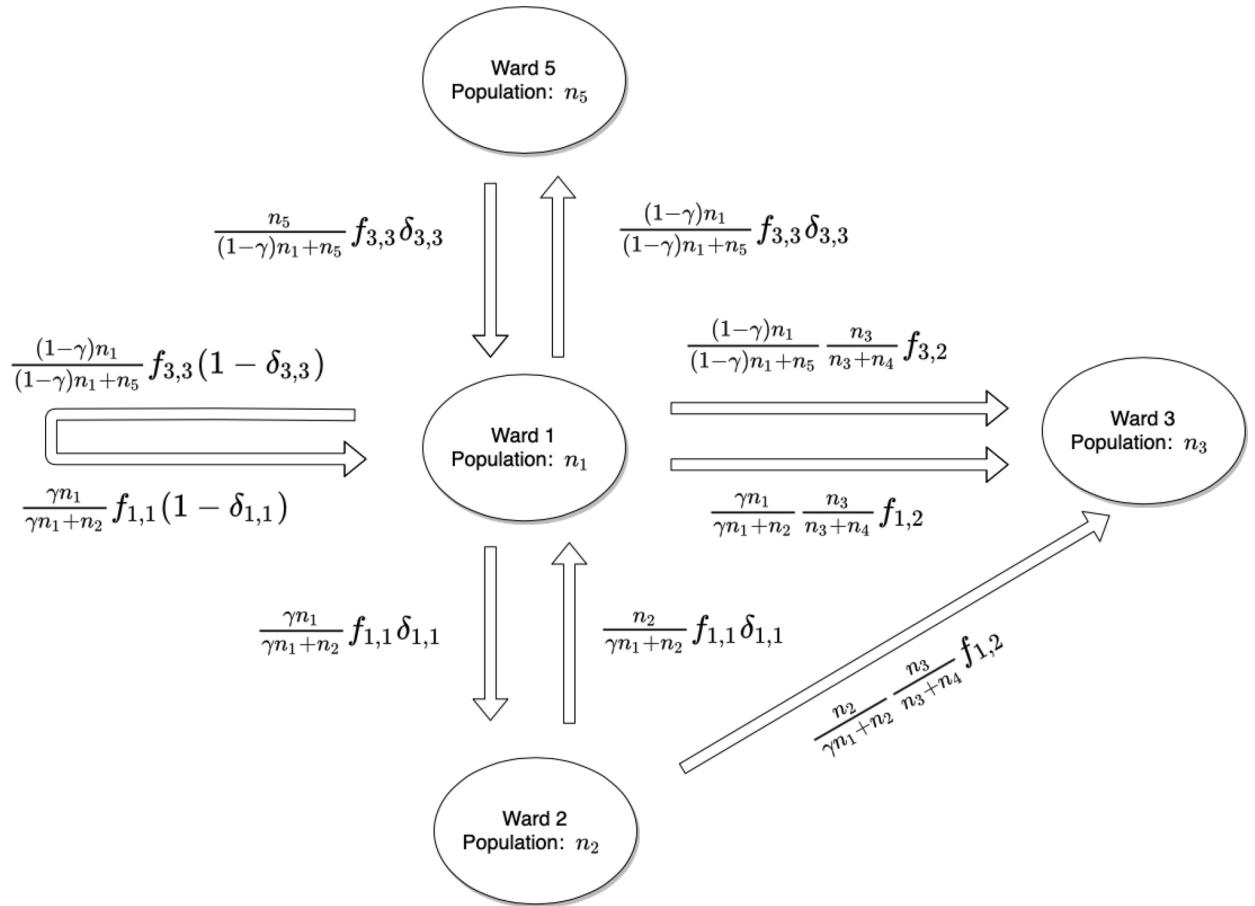


Fig. 12. Mapping region-based travel data to ward. This figure illustrates how we map the travel survey data we collect to wards, our suburb structure. We allocate region-level flows to wards in proportion to the size of both the origin and destination ward population as a share of their respective total regional population. γ reflects the percentage overlap between a ward and a region, in the case of an overlap. $f_{i,j}$ represents flows between regions. $\delta_{i,i}$ represents the share of survey respondents who both live and work or attend school in region i and who have an average travel time that is above the 25th percentile of the travel time distribution of all individuals who both live and work or attend school in region i .

Initialise city districts: Execute Algorithm 2

Create the household network structure: Execute Algorithm 3

Create the a city-wide network structure of recurring contacts: Execute Algorithm 4

Algorithm 1. Initialisation.

C1. Initialise city districts

We use [Algorithm 2](#) to calculate how many agents there should be in the simulation and what properties they should have to be proportional to the modelled city.

C2. Create the household network structure

Next, the creation of the household network structure is described in [Algorithm 3](#).

C3. Create the city-Wide network structure of other contacts

Finally, [algorithm 4](#) creates a city-wide network for all non-household contacts.

After these algorithms have been executed, we save the modelled city so it can be simulated.

```

Generate set of districts:  

 $\mathcal{W} \leftarrow$  list of city district identifiers  

for each  $w \in \mathcal{W}$  do  

    Assign population size to model district:  

     $N_w \leftarrow F_w^{\text{pop}} \frac{N}{N^{\text{pop}}}$   

    for each  $j \in \mathcal{N}_w$  do  

        Assign characteristics to agent:  

         $w_j \leftarrow w$   

         $a_j \leftarrow$  random draw from  $F_w^{\text{age}}$   

         $w'_j \leftarrow$  random draw from  $F_w^{\text{tv}}$   

         $\zeta_j(0) \leftarrow \xi(0) + \varepsilon_j(0)$   

    end for  

end for

```

Algorithm 2. Initialise city districts.

```

for each  $w \in \mathcal{W}$  do  

    Truncate household size distribution to district size:  

     $\tilde{F}_w^{\text{hs}}(N_{h,w}) = F_w^{\text{hs}}(N_{h,w}|N_{h,w} \leq N_w)$   

    Initialize set of households in district:  $\mathcal{H}_w = \emptyset$   

    Draw household sizes  $N_{h,w}$  until district population exhausted:  

    while  $\sum_{h \in \mathcal{H}_w} N_{h,w} < N_w$  do  

        draw new  $N_{h',w}$  from  $\tilde{F}_w^{\text{hs}}(N_{h,w})$  and add  $h'$  to  $\mathcal{H}_w$ .  

        if  $N_{h',w} > N_w - \sum_{h \in \mathcal{H}_w} N_{h,w}$  then set  $N_{h',w} = N_w - \sum_{h \in \mathcal{H}_w} N_{h,w}$   

        end if  

    end while  

    Denote the set of agents not assigned to a household  $\mathcal{N}_w^-$   

    Draw  $H_w = |\mathcal{H}_w|$  household heads randomly, without replacement from  $\mathcal{N}_w^-$ .  

    for each household head  $j \in \mathcal{H}_w$  (with age  $a_j$ ) do  

        Assign  $j$  to empty household  $h_j \in \mathcal{H}_w$  with size  $N_{h_j,w}$   

        Draw  $N_{h_j,w}$  household members without replacement from  $\mathcal{N}_w^-$  according to  $F_{a_j}^{\text{hc}}$   

        Create network edges between all members of the household  

    end for  

end for  

This yields the sets of household contacts for each agent  $\mathcal{N}_j^{\text{hc}} \forall j \in \mathcal{N}$ 

```

Algorithm 3. Create the household network structure.

Generate effective populations of districts:

```

for each  $w \in \mathcal{W}$  do  

    Construct  $\mathcal{N}_w^{\text{tv}} = \{j \in \mathcal{N} | w'_j = w\}$   

end for  

for each  $j \in \mathcal{N}$  do  

    Find total number of contacts  $N_j^{\text{oc}}$  from  $F_{a_j}^{\text{oc}}$   

    Limit the number of contacts to be feasible:  $N_j^{\text{oc}} \leftarrow \max\{N_j^{\text{oc}}, |\mathcal{N}_{w'_j}^{\text{tv}} \setminus \{j\}|\}$   

    Draw  $N_j^{\text{oc}}$  agents, without replacement, from  $\mathcal{N}_{w'_j}^{\text{tv}} \setminus \{j\}$  according to probabilities from  $F_{a_j}^{\text{oc}}$   

    Create edges between  $j$  and selected agents  

end for  

This yields the sets of non-household contacts for each agent  $\mathcal{N}_j^{\text{oc}} \forall j \in \mathcal{N}$   

The full set of contacts for each agent is  $\mathcal{N}_j = \mathcal{N}_j^{\text{hc}} \cup \mathcal{N}_j^{\text{oc}}$ 

```

Algorithm 4. Create the other contacts network structure.

```

Initialization of infections: Execute Algorithms 6
for each  $t \in [1, T]$  do
    Update health system load: Execute Algorithm 7
    Update compliance of each individual: Execute Algorithm 11
    Update status of infected individuals: Execute Algorithm 9
    Compute new infections: Execute Algorithm 10
end for
Store time paths of all variables

```

Algorithm 5. Main simulation.

Appendix D. Pseudo Code: Main Simulation Algorithm

An overview of the main simulation algorithm is described in [algorithm 5](#)

D1. Initial infections

[Algorithm 6](#) infects a number of initial agents. These agents are spread over the different districts in proportion to the initially detected cases per district from F^{ca} . All initially infected agents will have their initial status ($P_j(0)$) updated to exposed, infected without symptoms, or infected with symptoms, and the number of days for which they have been in the status at the start of the simulation $T_j^P(0)$ set as a random integer between zero and the maximum days that agents can be in that compartment.

After the initial infections have occurred, the day loop simulation starts and the next algorithms will be called every day.

D2. Check health system capacity

[Algorithm 7](#) checks if the health system is overburdened and activates the fatality multiplier δ^L if it is.

D3. Update compliance

Next, every agent j updates their degree of compliance with lockdown regulations, through a naive deGroot learning [Algorithm 8](#). Each agent has a private signal $\zeta_j(t)$ which consists of a public signal $\xi(t)$ and an individual noise term $\varepsilon_j(t)$, where the public signal is informed by the stringency index of the country being modelled $F^{strin}(t)$. $\xi(t) \in [0, 1]$ and $\varepsilon_j(t)$ is drawn independently for each j from a truncated normal distribution with support $[-\xi(t), 1 - \xi(t)]$ and mean 0 and variance σ , such that $\zeta_j(t) \in [0, 1]$.

The level of compliance $\phi_j(t)$ of the agents is a weighted average of the private signal $\zeta_j(t)$ (with a weight of ρ) and the social signal (with a weight of $(1 - \rho)$). The social signal is the simple average of observed previous $(t - 1)$ compliance of all neighbours $\frac{1}{N_j(t)} \sum_{k \in \mathcal{N}_j} \phi_k(t - 1)$.

D4. Update agents' infection status

In [Algorithm 9](#), each agent j who is in the exposed, asymptomatic, symptomatic, or critical compartments will have the number of days that they have been in this status updated and, and possible transition to a new disease status recorded. Let $T_j^P(t)$ denote the number of days that agent j has been in disease compartment P by period t (where τ^P is the parameter that calibrates the tenure in compartment P). Since some transitions to different compartments are stochastic, we use random draws from a uniform distribution

```

Initialize all agents as susceptible:  $P_j(0 : T) = S \forall j \in \mathcal{N}$ 
Find districts  $\mathcal{W}^{ca}$  and number of initial cases  $\tilde{N}_w^{ca}$  in  $F^{ca}$ 
Normalize initial cases to model population:  $N_w^{ca} = \tilde{N}_w^{ca} \frac{N}{N^{pop}}$ 
for each  $w \in \mathcal{W}^{ca}$  do
    Randomly draw  $N_w^{ca}$  agents to form  $\mathcal{N}_w^{ca} \subset \mathcal{N}_w$ 
    for each  $j \in \mathcal{N}_w^{ca}$  do
        Change  $P_j(0)$  to one of the disease statuses in  $\{E, I^{as}, I^s\}$  with equal probability
        Set initial days in status,  $T_j^P(0)$ , by uniform random draw from  $[0, \tau^P]$ 
    end for
end for

```

Algorithm 6. Initialize infections.

```

Compute number of critically ill patients:  $N^C(t-1) = |\{j \in \mathcal{N} | P_j(t-1) = C\}|$ 
if  $N^C(t-1) > L$  then
     $\delta^L \leftarrow Q$ 
else
     $\delta^L \leftarrow 1$ 
end if

```

Algorithm 7. Check health system capacity.

```

Generate public signal  $\xi(t)$  from  $F^{strin}(t)$ 
for each  $j \in \mathcal{N}$  do
    if  $P_j \in [I^s, C, D]$  then
         $\phi_j(t) \leftarrow 1$ 
    else
        Draw private signal:  $\zeta_j(t) \leftarrow \xi(t) + \varepsilon_j(t)$ 
        Compute compliance:  $\phi_j(t) \leftarrow \rho \zeta_j(t) + (1 - \rho) \frac{1}{N_j} \sum_{k \in \mathcal{N}_j} \phi_k(t-1)$ 
    end if
end for

```

Algorithm 8. Update agent compliance.

with support on $[0,1]$ to determine the outcome of a stochastic event Z for each agent in each period. We denote the random draw by $\tilde{\pi}^Z$. Thus, if the probability of event Z is π^Z , then the event occurs only when $\tilde{\pi}^Z < \pi^Z$.

D5. Compute new infections

Next, [Algorithm 10](#) will compute which agents will become infected next by looping over all infected agents and determining how many neighbours they visit and ultimately infect. Infection occurs as a result of two stochastic events in the model: a physical meeting between two agents and the stochastic transmission of the virus conditional on meeting. First, a physical meeting between two agents i and j occurs with a time-varying, pair-specific probability $\pi_{ij}^M(t)$ and second, conditional on a physical meeting between an infectious and a susceptible agent, transmission of the disease occurs with probability π^E (the fundamental transmissibility of the virus which we estimated to fit to incidence data). $\pi_{ij}^M(t)$ depends on a variety of features within the model. First, agent j inevitably has daily physical meetings with all members of their household, i.e. $i \in \mathcal{N}_j^{hc} \Rightarrow \pi_{ij}^M(t) = 1$. Second, for members in the set of non-household contacts of agent j , the probability of a meeting between j and $i \in \mathcal{N}_j^{oc}$ depends on two components: (i) a fundamental parameter that calibrates the likelihood of a meeting between the two agents if they are both fully compliant with lockdown regulations of authorities ($\omega \in [0, 1]$) and (ii) the degree compliance of *each* agents with lockdown regulations of authorities that aim to prevent transmission. This is modelled as follows: in every period t , agent j chooses a degree of compliance $\phi_j(t)$ which is a function of a public signal and observations (via De Groot learning) of the degree of compliance of other agents in the agent's network as described in [Algorithm 11](#). The probability of a meeting between agent j and $i \in \mathcal{N}_j^{oc}$ is specified as:

$$\pi_{ij}^M(t) = (\omega + (1 - \omega)(1 - \phi_i(t))(\omega + (1 - \omega)(1 - \phi_j(t)))$$

Thus, if both agents are fully compliant (i.e. $\phi_i(t) = \phi_j(t) = 1$), the probability of a physical meeting is ω^2 . If both agents are fully non-compliant (i.e. $\phi_i(t) = \phi_j(t) = 0$), the probability of a physical meeting is 1. This encodes two features: (i) even with full compliance with policy, some physical meetings may happen during the course of everyday life, and (ii) individuals can always coordinate to ensure a meeting should they wish to do so strongly enough.

D6. Update compliance using lexicographic learning

If the alternative lexicographic learning scenario is active, a sub-algorithm follows [Algorithm 11](#). Following this algorithm, an agent will update their compliance to $\tilde{\phi}_j(t) = 1$ if they have a neighbour who is critically ill or has died. Specifically, when agent j has a neighbour who is critically ill or dead, they will set their desired degree of contact reduction to the maximum $\tilde{\phi}_j(t) = 1$. Otherwise, they ignore all restrictions and do not choose to reduce their contacts at all, $\tilde{\phi}_j(t) = 0$. Otherwise, the model remains unchanged.

```

for each  $j \in \mathcal{N}$  do
    if  $P_j(t-1) = E$  then  $T_j^E(t) \leftarrow T_j^E(t-1) + 1$ 
        if  $T_j^E(t) \leq \tau^E$  then  $P_j(t) \leftarrow E$ 
        else draw  $\tilde{\pi}^s \sim U[0, 1]$ 
            if  $\tilde{\pi}^s < \pi^s$  then
                 $P_j(t) \leftarrow I^s, T_j^s(t) \leftarrow 1$ 
            else
                 $P_j(t) \leftarrow I^{as}, T_j^{as}(t) \leftarrow 1$ 
            end if
        end if
    end if
    if  $P_j(t-1) = I^{as}$  then  $T_j^{as}(t) \leftarrow T_j^{as}(t-1) + 1$ 
        if  $T_j^{as}(t) \leq \tau^{as}$  then  $P_j(t) \leftarrow I^{as}$ 
        else  $P_j(t : T) \leftarrow R$ 
        end if
    end if
    if  $P_j(t-1) = I^s$  then  $T_j^s(t) \leftarrow T_j^s(t-1) + 1$ 
        if  $T_j^s(t) \leq \tau^s$  then  $P_j(t) \leftarrow I^s$ 
        else draw  $\tilde{\pi}^C \sim U[0, 1]$ 
            if  $\tilde{\pi}^C < \pi^{C,a_j}$  then
                 $P_j(t) \leftarrow C, T_j^C(t) \leftarrow 1$ 
            else
                 $P_j(t : T) \leftarrow R$ 
            end if
        end if
    end if
    if  $P_j(t-1) = C$  then  $T_j^C(t) \leftarrow T_j^C(t-1) + 1$ 
        if  $T_j^C(t) \leq \tau^c$  then  $P_j(t) \leftarrow C$ 
        else draw  $\tilde{\pi}^D \sim U[0, 1]$ 
            if  $\tilde{\pi}^D < \delta^L \pi^{D,a_j}$  then
                 $P_j(t : T) \leftarrow D$ 
            else
                 $P_j(t : T) \leftarrow R$ 
            end if
        end if
    end if
end for

```

Algorithm 9. Update infection status.

Appendix E. Data Sources

All data used in this paper is publicly available. In this section, we outline the data we used and provide information on how the data can be downloaded.

- F^{ca} : Observed cases per district
 - Source: Western Cape Government
 - Each day, the Western Cape government releases a pdf report documenting the number of cases by sub-district. These reports can be found [here](#)
 - We transcribe these cases for each sub-district for each day. A sub-district is a larger spatial definition than the wards we use in this paper.
 - We then overlay our wards on the sub-districts and assign wards to sub-districts.
 - Finally, we assign sub-district level infection cases to wards based on a probability calculated as the ward level population normalised by the sub-district level population.
- F^{hc} : Observed age group household contacts

```

for each  $j \in \mathcal{N}$  do
  if  $P_j(t) \in [I^s, I^{as}]$  then
    for each  $i \in \mathcal{N}_j$  do
      if  $P_i(t) = S$  then
        if  $i \in \mathcal{N}_j^{hc}$  then
           $\pi_{i,j}^M(t) \leftarrow 1$ 
        else (i.e.  $i \in \mathcal{N}_j^{oc}$ )
           $\pi_{i,j}^M(t) \leftarrow (\omega + (1 - \omega)(1 - \phi_i(t))(\omega + (1 - \omega)(1 - \phi_j(t)))$ 
        end if
        draw independently  $\tilde{\pi}^E, \tilde{\pi}^M \sim U[0, 1]$ 
        if  $\tilde{\pi}^M < \pi_{i,j}^M(t)$  and  $\tilde{\pi}^E < \pi^E$  then
           $P_i(t) \leftarrow E$ 
        end if
      end if
    end for
  end if
end for

```

Algorithm 10. Compute new infections.

```

for each  $j \in \mathcal{N}$  do
  for each  $i \in \mathcal{N}_j$  do
    if  $P_i(t) \in [I^s, C, D]$  then  $\tilde{\phi}_j(t) = 1$ 
    else  $\tilde{\phi}_j(t) = 0$ 
    end if
  end for
end for

```

Algorithm 11. Lexicographic learning.

- Source: Prem et al. (2017)
- Data can be downloaded from the journal website, [here](#)
- F^{oc} : Observed non-household contacts
 - Source: Prem et al. (2017)
 - Data can be downloaded from the journal website, [here](#)
- F^{pop} : Observed district population
 - Source: 2011 Census from Statistics South Africa
 - Data can be downloaded from Statistics South Africa's website, after the creation of a free profile, [here](#). Once you are logged in, navigate to *Community Profiles > Census 2011 (2016 Boundaries)* in the sidebar
- F^{age} : Observed age distribution per district
 - Source: 2011 Census from Statistics South Africa
 - Data can be downloaded from Statistics South Africa's website, after the creation of a free profile, [here](#). Once you are logged in, navigate to *Community Profiles > Census 2011 (2016 Boundaries)* in the sidebar
- F^{TV} : Observed travel matrix
 - Source: 2013 National Household Travel Survey from Statistics South Africa
 - Data is obtained from DataFirst. Data can be found on the DataFirst data portal, [here](#). You will need to create a free account to access the data.
 - We discuss the steps taken in mapping this data to our ward spatial structure in [Appendix B.3](#).
- F^{hs} : Observed district household size distribution
 - Source: 2011 Census from Statistics South Africa
 - Data can be downloaded from Statistics South Africa's website, after the creation of a free profile, [here](#). Once you are logged in, navigate to *Community Profiles > Census 2011 (2016 Boundaries)* in the sidebar
- F^{in} : Informality level
 - Source: 2011 Census from Statistics South Africa
 - Data can be downloaded from Statistics South Africa's website, after the creation of a free profile, [here](#). Once you are logged in, navigate to *Community Profiles > Census 2011 (2016 Boundaries)* in the sidebar
- F^{strin} : Stringency index

- Source: Oxford Covid-19 Government Response Tracker
- Data can be downloaded from Github, [here](#). Download the OxCGRT_latest.csv file and select country code ZAF and the column StringencyIndex

References

- Acemoglu, D., Chernozhukov, V., Werning, I., Whinston, M.D., 2020. Optimal targeted lockdowns in a multi-group SIR model. NBER Working Paper 27102. National Bureau of Economic Research.
- Acuña-Zegarra, M.A., Santana-Cibrian, M., Velasco-Hernandez, J.X., 2020. Modeling behavioral change and Covid-19 containment in mexico: a trade-off between lockdown and compliance. *Math Biosci* 325, 108370.
- Adepoju, P., 2020. Africa's struggle with inadequate covid-19 testing. *The Lancet Microbe* 1 (1), e12.
- Akbarpour, M., Cook, C., Marzuoli, A., Mongey, S., Nagaraj, A., Saccarola, M., Tebaldi, P., Vasserman, S., Yang, H., 2020. Socioeconomic network heterogeneity and pandemic policy response. Technical Report. National Bureau of Economic Research.
- Almagor, J., Picascia, S., 2020. Exploring the effectiveness of a Covid-19 contact tracing app using an agent-based model. *Sci Rep* 10 (1), 1–11.
- Ambrosio, B., Aziz-Alaoui, M., 2020. On a coupled time-dependent sir models fitting with new york and new-jersey states covid-19 data. *Biology (Basel)* 9 (6), 135.
- Arndt, C., Davies, R., Gabriel, S., Harris, L., Makrelow, K., Robinson, S., Levy, S., Simbanegavi, W., van Seventer, D., Anderson, L., 2020. Covid-19 lockdowns, income distribution, and food security: an analysis for south africa. *Glob Food Sec* 26, 100410.
- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A., 2018. Social connectedness: measurement, determinants, and effects. *Journal of Economic Perspectives* 32 (3), 259–280.
- Bailey, M., Johnston, D., Koenen, M., Kuchler, T., Russel, D., Stroebel, J., 2021. Social networks shape beliefs and behavior: evidence from social distancing during the Covid-19 pandemic. NBER Working Paper 28234. National Bureau of Economic Research.
- Basurto, A., Dawid, H., Harting, P., Hepp, J., Kohlweyer, D., 2020. Economic and epidemic implications of virus containment policies: insights from agent-based simulations. GROWINPRO Working Paper. GROWINPRO.
- Birge, J.R., Candogan, O., Feng, Y., 2020. Controlling epidemic spread: reducing economic losses with targeted closures. University of Chicago, Becker Friedman Institute for Economics Working Paper (2020-57).
- Boppert, T., Harmenberg, K., Krusell, P., Olsson, J., 2022. Integrated epi-econ assessment of vaccination. *Journal of Economic Dynamics and Control* 104308.
- Bubar, K.M., Reinholz, K., Kissler, S.M., Lipsitch, M., Cobey, S., Grad, Y.H., Larremore, D.B., 2021. Model-informed Covid-19 vaccine prioritization strategies by age and serostatus. *Science* 371 (6532), 916–921.
- Calafiore, G.C., Novara, C., Possieri, C., 2020. A time-varying SIRD model for the covid-19 contagion in Italy. *Annu Rev Control*.
- Centre for Disease Control and Prevention, 2021. Science Brief: SARS-CoV-2 Infection-induced and Vaccine-induced Immunity. Report.
- Charoenwong, B., Kwan, A., Pursiainen, V., 2020. Social connections with Covid-19-affected areas increase compliance with mobility restrictions. *Sci Adv* 6 (47), eabc3054.
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. *The Lancet* 395 (10223), 507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
- Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V., Vernon, M.C., 2011. Networks and the epidemiology of infectious disease. *Interdiscip Perspect Infect Dis* 2011.
- Dasaratha, K., 2020. Virus Dynamics with Behavioral Responses. arXiv preprint arXiv:2004.14533.
- Dasaratha, K., Golub, B., Hak, N., 2020. Learning from neighbors about a changing state. *SSRN Working Paper* 3097505.
- DeGroot, M.H., 1974. Reaching a consensus. *J Am Stat Assoc* 69 (345), 118–121.
- Delli Gatti, D., Reissl, S., 2020. ABC: An Agent Based Exploration of the Macroeconomic Effects of Covid-19. CESifo Working Paper. CESifo.
- Duffie, D., Singleton, K.J., 1990. Simulated moments estimation of Markov models of asset prices. NBER Working Paper t0087. National Bureau of Economic Research.
- Eichenbaum, M.S., Rebelo, S., Trabandt, M., 2020. The macroeconomics of epidemics. NBER Working Paper 26882. National Bureau of Economic Research.
- Ellison, G., 2020. Implications of heterogeneous SIR models for analyses of COVID-19. NBER Working Paper 27373. National Bureau of Economic Research.
- Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., et al., 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Report 6. WHO Collaborating Centre for Infectious Disease Modelling.
- Franke, R., Westerhoff, F., 2012. Structural stochastic volatility in asset pricing dynamics: estimation and model contest. *Journal of Economic Dynamics and Control* 36 (8), 1193–1211.
- Glaeser, E.L., Gorback, C., Redding, S.J., 2020. Jue insight: how much does Covid-19 increase with mobility? evidence from new york and four other us cities. *J Urban Econ* 103292.
- Glover, A., Heathcote, J., Krueger, D., 2022. Optimal age-based vaccination and economic mitigation policies for the second phase of the Covid-19 pandemic. *Journal of Economic Dynamics and Control* 140, 104306.
- Golub, B., Jackson, M.O., 2010. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics* 2 (1), 112–149.
- Golub, B., Sadler, E., 2016. Learning in social networks. *The Oxford Handbook of the Economics of Networks*.
- Hale, T., Petherick, A., Phillips, T., Webster, S., 2020. Variation in government responses to COVID-19. Blavatnik school of government working paper.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet* 395 (10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Ing, A.J., Cocks, C., Green, J.P., 2020. Covid-19: in the footsteps of ernest shackleton. *Thorax*. <https://doi.org/10.1136/thoraxjnlg-2020-215091>.<https://thorax.bmjjournals.org/content/early/2020/05/27/thoraxjnlg-2020-215091.full.pdf>
- Ioannidis, J.P., Cripps, S., Tanner, M.A., 2020. Forecasting for Covid-19 has failed. *Int J Forecast*.
- Keeling, M.J., Eames, K.T., 2005. Networks and epidemic models. *Journal of the Royal Society Interface* 2 (4), 295–307.
- Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115 (772), 700–721.
- Krueger, D., Uhlig, H., Xie, T., 2020. Macroeconomic dynamics and reallocation in an epidemic. NBER Working Paper 27047. National Bureau of Economic Research.
- Lee, B.-S., Ingram, B.F., 1991. Simulation estimation of time-series models. *J Econom* 47 (2–3), 197–205.
- Leon, D.A., Shkolnikov, V.M., Smeeth, L., Magnus, P., Pechholdová, M., Jarvis, C.I., 2020. Covid-19: a need for real-time monitoring of weekly excess deaths. *The Lancet* 395 (10234), e81.
- Makridis, C., Wang, T., 2020. Learning from friends in a pandemic: Social networks and the macroeconomic response of consumption. *SSRN Working Paper* 3601500.
- McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57 (5), 995–1026.

- Moein, S., Nickaeen, N., Rooointan, A., Borhani, N., Heidary, Z., Javanmard, S.H., Ghaisari, J., Gheisari, Y., 2021. Inefficiency of sir models in forecasting Covid-19 epidemic: a case study of isfahan. *Sci Rep* 11 (1), 1–9.
- Mwalili, S., Kimathi, M., Ojiambo, V., Gathungu, D., Mbogo, R., 2020. SEIR model for Covid-19 dynamics incorporating the environment and social distancing. *BMC Res Notes* 13 (1), 1–5.
- Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. *Computer Journal* 7, 308–313.
- Pasquariello, P., Stranges, S., 2020. Excess mortality from Covid-19: a commentary on the italian experience. *Int J Public Health* 65, 529–531.
- Prem, K., Cook, A.R., Jit, M., 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* 13 (9), e1005697.
- Rockett, R.J., Arnott, A., Lam, C., Sadsad, R., Timms, V., Gray, K.-A., Eden, J.-S., Chang, S., Gall, M., Draper, J., et al., 2020. Revealing Covid-19 transmission in australia by sars-cov-2 genome sequencing and agent-based modeling. *Nat. Med.* 26 (9), 1398–1404.
- Shen, C., Taleb, N.N., Bar-Yam, Y., 2020. Review of ferguson et al “impact of nonpharmaceutical interventions”. *New England Complex Systems Institute* 17.
- Silal, S., Pulliam, J., Meyer-Rath, G., Nichols, B., Jamieson, L., Kimmie, Z., Moultrie, H., 2020. Estimating cases for COVID-19 in South Africa Update: 19 May 2020. Presentation. Innovation Pharmaceutical Association South Africa.
- Squazzoni, F., Polhill, J.G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., Chappin, E., Borit, M., Verhagen, H., Giardini, F., et al., 2020. Computational models that matter during a global pandemic outbreak: a call to action. *JASSS: Journal of Artificial Societies and Social Simulation* 23 (2).
- Stein, M., 1987. Large sample properties of simulations using latin hypercube sampling. *Technometrics* 29 (2), 143–151.
- Thompson, M.G., 2021. Interim estimates of vaccine effectiveness of BNT162b2 and mRNA-1273 Covid-19 vaccines in preventing SARS-CoV-2 infection among health care personnel, first responders, and other essential and frontline workers—eight us locations, december 2020–March 2021. *MMWR Morb. Mortal. Wkly. Rep.* 70.
- Toxvaerd, F., 2020. Equilibrium social distancing. *Covid Economics* 110.
- Verelst, F., Willen, L., Beutels, P., 2016. Behavioural change models for infectious disease transmission: a systematic review (2010–2015). *Journal of The Royal Society Interface* 13 (125), 20160820.
- Verity, R., Okell, L., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P., Fu, H., Dighe, A., Griffin, J., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z.M., FitzJohn, R., Gaythorpe, K., Ferguson, N., 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).
- Voysey, M., Clemens, S.A.C., Madhi, S.A., Weckx, L.Y., Folegatti, P.M., Aley, P.K., Angus, B., Baillie, V.L., Barnabas, S.L., Bhorat, Q.E., et al., 2021. 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* 397 (10269), 99–111.
- World Health Organization and others, 2020. Fair allocation mechanism for COVID-19 vaccines through the COVAX Facility. Final working version-9 September.
- Yang, R., Gui, X., Xiong, Y., 2020. Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in wuhan, china. *JAMA Network Open* 3 (5). <https://doi.org/10.1001/jamanetworkopen.2020.10182>-e2010182-e2010182