Modeling Disease Spread: Integrating Mobility, Awareness, and Behavior

Cloe Cortes Balcells [*] Rico Krueger [†] Michel Bierlaire [*]

March 14, 2025

Report TRANSP-OR 240909
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

[*]École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland, {cloe.cortes, michel.bierlaire}@epfl.ch

[†]Department of Technology, Management and Economics, Technical University of Denmark (DTU), {rickr@dtu.dk}

Abstract

The COVID-19 pandemic has underscored the vital link between epidemiology and transportation, particularly in the context of activity-travel behavior. While much research has explored the role of mobility in disease spread, there is a notable gap in understanding how individual behavioral choices, especially regarding testing, influence these dynamics. This paper introduces the Mobility-Aware Behavioral Epidemiological Model (MABEM), an innovative framework that integrates activity-based modeling with latent variables to more accurately represent the impact of testing decisions on disease propagation. Using data from the canton of Vaud, Switzerland, our model simulates the interactions between individual behavior, mobility, and health status, revealing significant underreporting of infections and highlighting the crucial role of testing choices in shaping observed infection rates. The results suggest that testing behaviors vary significantly across different demographic groups and regions, influencing both individual activity patterns and overall disease spread. MABEM model proves to be computationally efficient and offers valuable insights for designing targeted public health interventions. Despite some limitations related to data availability, this research provides a comprehensive approach to understanding the interplay between mobility, individual behavior, and disease dynamics, offering new possibilities for more effective disease containment strategies.

Keywords: Interdisciplinary, Activity-Travel Behavior, Epidemiological Modeling, Activity-Based Model, Discrete Choice Modeling, COVID-19.

1 Introduction

The COVID-19 pandemic has undeniably bridged the epidemiological and transportation communities, highlighting the critical role of activity-travel behavior in the spread of infectious diseases. Governments worldwide have implemented curfews and quarantines to contain the virus. While the link between transportation and epidemiological communities is well-established, less attention has been given to the crucial influence of individual testing choices on mobility patterns and subsequent disease spread. Understanding the behavioral aspect of an individual's decision to get tested is crucial, as it directly impacts their reaction to the disease. One of the main challenges in modeling such behavior is that key behavioral elements are often latent and not directly observable. For instance, one cannot directly observe whether someone is infected; instead, one observes the outcome of a test, which itself depends on the individual's decision to get tested. This creates a layer of complexity, as the model must infer latent states, like infection

status, from observable actions and outcomes, such as test results. For example, if an individual is unaware of being sick, they will continue their regular activities, potentially spreading the disease further. This underscores the importance of modeling individual awareness from a mobility point of view. An example of different behavior regarding testing choices can be found by analyzing Swiss data from the COVID-19 2019 pandemic (Swiss National, 2021). This dataset clearly shows a non-negligible difference in testing habits between French-speaking and German-speaking cantons in Switzerland. Specifically, in the French-speaking canton of Vaud, the average number of contacts per day per person is higher, and the number of tests per 100,000 individuals is lower compared to the German-speaking canton of Zurich. This disparity raises important questions regarding how individual testing choices impact activity-travel behavior and, consequently, disease spread.

Although several studies from the epidemiological and transportation communities have begun to integrate aspects of human mobility and behavior into their epidemiological models (see Müller et al., 2021; Hou et al., 2021; Gozzi et al., 2021; Balcan et al., 2010), there remains a significant gap in current research on how to dynamically include individual behavioral choices, particularly regarding testing and the resulting behavioral adjustments. The epidemiological community, while developing more complex models of disease transmission, often relies on simplified representations of mobility patterns, such as static networks or average contact rates for groups of people. These approaches can overlook the intricate ways in which daily activities contribute to disease spread. On the other hand, the transportation community has made advances by introducing activity-based models to generate synthetic populations and detailed daily schedules, yet they often oversimplify the transmission model by using predefined transition probabilities and neglecting the crucial feedback loop between testing choices and subsequent daily activities. Meanwhile, the medical community has provided valuable insights into the importance of testing for controlling disease spread and has developed strategies to encourage people to get tested (see Hengel et al., 2021; Flynn et al., 2020). While these models are highly effective for policy purposes, they are not typically embedded within an epidemiological framework, nor do they model activity-travel behavior to assess how increasing testing impacts both daily activities and, ultimately, the spread of the disease. There is a clear need for interdisciplinary models that dynamically explore the relationship among testing decisions, activity-travel behavior, and disease spread to fully understand their interconnected impacts. Such models would bridge the gaps between the epidemiological, transportation, discrete choice, and medical communities, offering a comprehensive understanding of how individual behaviors drive disease dynamics and informing more effective public health interventions (see

Tuomisto et al., 2020).

To address these gaps, this paper proposes the Mobility-Aware Behavioral Epidemiological Model (MABEM), an activity-based epidemiological model designed to explore the impact of individual testing decisions on mobility patterns and disease transmission. MABEM consists of two primary components: the epidemiological-behavioral submodel and the mobility restriction submodel. The epidemiological-behavioral submodel includes latent variables for exposure and propensity for testing, allowing us to separately model the health state of individuals (susceptible, infected, or recovered) and their testing choices, along with the outcomes of these tests. The mobility restriction submodel uses the state of awareness, triggered by a positive test result, to modify individual schedules. This comprehensive approach allows for a detailed analysis of how variations in testing behavior influence not only an individual's daily schedules but also the broader dynamics of disease spread, providing essential insights for policymakers aiming to craft more targeted and effective public health interventions.

The remainder of this paper is organized as follows: Section 2 reviews the related literature. Section 3 details the methodology of our model. Section 4 presents the findings from a case study of 810,486 individuals in the canton of Vaud, Switzerland. Finally, Section 5 offers concluding remarks and outlines future research directions.

2 Literature Review

2.1 Epidemiological Models Overview and the Rise of Individual-Based Models during COVID-19

Traditional epidemiological models can be classified into three categories: compartmental, network-based, and individual-based. Compartmental models, such as SIR or SEIR models, simplify the population into distinct categories (for instance, Susceptible, Exposed, Infected, and Recovered), where the transition of each category is represented by differential equations (see Kelman, 1985). Despite their mathematical simplicity and ease of analysis, these models often overlook the "imperfect mixture" of populations, neglecting individual heterogeneity and complex behaviors (Smieszek, 2009; Edmunds et al., 1997). Network-based models, on the other hand, offer a more complex representation of social interactions, as demonstrated in studies such as Hou et al., 2021 and Cui et al., 2021. These models use graph theory to map individual interactions (Mancastroppa et al., 2020). Despite their ability to simulate social heterogeneity, network models face chal-

lenges in densely populated environments and often lack dynamic representations. On the contrary, activity-based models, which are a subset of individual-based models, present a detailed representation of individual behaviors. These models have been especially valuable for understanding the spread of disease in complex real-world scenarios, as seen in studies such as Kerr et al., 2020, Hackl and Dubernet, 2019, Müller et al., 2021, and Eubank et al., 2004. Activity-based epidemiological models have been widely used to guide public health responses during the pandemic (Kerr et al., 2020). With the onset of the COVID-19 crisis, the importance of individual-based models has become more pronounced, since they take into account the crucial role of individual behavior in the spread of the virus.

2.2 Advanced Epidemiological Models and Limitations

The transportation community has experienced a significant increase in literature integrating mobility with epidemiological models (Kerr et al., 2020, Müller et al., 2021) to help policymakers during the COVID-19 emergency. Müller et al., 2021 use MATSim, an activity-based model, to simulate individual mobility and incorporate a mechanistic infection model and a person-centric disease progression model. Similarly, Kerr et al., 2020 developed CovaSim, an agent-based model that simulates the spread of COVID-19 by incorporating individual behaviors, contact networks, and intervention strategies. CovaSim can evaluate the impact of various public health interventions, such as social distancing, testing, and contact tracing, on the spread of the virus. While these models provide valuable insights into how various public health measures impact disease transmission, they do not fully address the complexity of behavioral modeling, particularly the challenge of inferring unobservable factors like infection status from testing choices. Additionally, they overlook the critical feedback loop between an individual's choice to get tested and their subsequent behaviors, which is essential for a more comprehensive understanding of infection dynamics.

On the other hand, the epidemiological community explores the importance of testing to contain the spread of the disease (Cui et al., 2021, Aronna et al., 2021). For instance, Cui et al., 2021 propose a network-based model that examines different testing strategies, such as random testing and contact-tracking testing, and their impact on epidemic control. Their study highlights that higher testing priority for individuals in close contact with confirmed cases significantly reduces the infection scale. Aronna et al., 2021 introduce a compartmental model focusing on the effectiveness of testing and quarantine measures, finding that testing asymptomatic cases is crucial for controlling the epidemic. Similar conclusions are reached in the field of medical research (see Hengel et al., 2021; Flynn et al., 2020). Flynn et al., 2020 present a drive-through testing model designed to

increase the volume and optimize the use of limited resources for testing. Similarly, Hengel et al., 2021 propose a decentralized point-of-care testing model aimed at improving accessibility and making testing equally available across the population, especially in underserved communities.

While all these studies capture how testing impacts disease spread, they fail to properly model the activity-travel behavior of individuals and how it connects to disease spreading. Addressing this gap requires an integrated approach that considers the full spectrum of individual behaviors and choices, providing a more comprehensive understanding of the dynamics of the disease.v

2.3 Literature Gaps

All in all, Sections 2.1 and 2.2 identify several gaps in existing epidemiological models, especially regarding the integration of individual activity-travel behavior through activity-based models and decision-making processes like testing choices into the modeling of disease spread. This paper aims to address these gaps by incorporating individual choices, like infectious disease testing decisions, and capturing the complexity of human behavior and interaction. This is crucial for understanding and managing public health outcomes, particularly in global health crises. The proposed model advances the field by dynamically tracking individual health states and the impact of personal decisions on disease spread and control, highlighted in the detailed literature review comparison visible in Table 1 . This summary effectively outlines the current state of epidemiological research, identifying key areas that need more research and improvement, especially in terms of behavioral modeling.

3 Methodology

This research aims to develop MABEM, which includes activity-travel behavior, health state transitions, and testing choices. To provide a better understanding of disease spread MABEM includes two submodels: the mobility restriction submodel, and the epidemiological behavioral choice submodel. The model considers the infection dynamics, testing outcomes, the influence of a positive test on the individual's mobility, and the impact of awareness of a positive test on mobility at the individual level. The proposed framework simulates individual behaviors over discrete time intervals. MABEM receives as input the schedule of each individual from an activity-based model, together with information about the individuals, and the facilities in the network, and returns not only whether the individual is susceptible (s), infected (i), or recovered (r) at each time interval, but also tracks

Ref	Title	Mobility				Epidemiological Model			
	×	Model	Type of Model	Heterogeneity	Mobility Pattern	Model	Type of Model	Heterogeneity	Epidemiological Pattern
Aleta et al., 2022	Modelling the impact of testing, contact tracing, and household quarantine on second waves of COVID-19	×	×	✓	Integration of mobility contact data per age group (child/adult)	✓	Agent-based model, S-LA-LS-PS-IA-IS-H-ICU-R	×	Data-driven
Aronna et al., 2021	A model for COVID-19 with isolation, quarantine and testing as control measures	×	×	×	×	✓	SEI-AQ-RD compartmental model	×	Focus on isolation, quarantine, and testing to control epidemic through the rates between compartments
Balcan et al., 2010	Modeling the spatial spread of infectious diseases: The Global Epidemic and Mobility model	×	×	×	High-resolution data, Voronoi decomposition	✓	Network, SLIR	✓	Infection dynamics within regions
Brotherhood et al., 2020	An Economic Model of the Covid-19 Epidemic: The Importance of Testing and Age-Specific Policies	×	×	×	The economy is populated by two types of agents. Agents work, enjoy leisure outside the home and home hours	✓	Agent-based, SIRD, individual probability function	✓	Impact of testing, age-based probabilities
Cui et al., 2021	A network-based model to explore the role of testing in the epidemiological control of the COVID-19 pandemic	×	×	×	×	✓	Network-based model,SEIR	×	Testing strategies through the rates between compartments
Eubank et al., 2004	Modelling disease outbreaks in realistic urban social networks	✓	ABM	✓	TRANSIMS-based social network estimation	✓	Agent-based, SIRD, individual probability function	×	Fully homogenous epidemiological model
Gozzi et al., 2021	Estimating the effect of social inequalities on the mitigation of COVID-19	×	×	×	Extended Detail Records (XDR) from mobile phones in regions	✓	Network, SLIR, inspired by Balcan et al., 2010	✓	Spread of COVID-19 within regions
Hackl and Dubernet, 2019	Epidemic spreading in urban aras using agent-based transportation models	✓	ABM	✓	MATSim	✓	Agent-based, SIR, individual probability function	×	Fully homogenous epidemiological model
Hou et al., 2021	Intracounty modeling of COVID-19 infection with human mobility	×	×	×	Walktrap network-based community detection method	✓	Network, Stochastic SEIR	✓	Spread of COVID-19 in different communities, heterogeneity based on regional mobility
Kerr et al., 2020	Covasim: an agent-based model of COVID-19 dynamics and interventions	×	×	✓	Random individual-level interactions through networks	✓	Agent-based, SEP(MSC)A-RD, individual state transitions	✓	Probability of infection based on age group
Müller et al., 2021	Using mobile phone data for epidemiological simulations of lockdowns	✓	ABM	✓	MATSim	✓	SEIR, individual probability function from Smieszek, 2009	✓	Age probability modification
Perez and Dragicevic, 2009	An agent-based approach for modeling dynamics of contagious disease spread	×	×	×	Random generated rban mobility and daily activities, GIS-integrated	✓	Agent-based, SEIRD, individual probability function	×	Focus on public places for disease transmission
Reich et al., 2020	Modeling COVID-19 on a network: super-spreaders, testing and containment	×	×	×	×	✓	Network-based model,SEIR	×	Testing strategies through the rates between compartments
Tuomisto et al., 2020	An agent-based epidemic model REINA for COVID-19 to identify destructive policies	×	×	✓	Random contacts per individual, contact matrices	✓	Agent-based, SEI-HC-RD, individual state transitions	✓	Probability of infection based on age group
Our approach	Enhancing Activity-Based Modeling: Dynamics of Mobility, Awareness, and Individual Choices in Disease Spread	✓	ABM	✓	MATSim	✓	Agent-based model S-I-R	✓	Individual-choices on testing and impact on mobility

Table 1: Summary of relevant epidemiological studies and their characteristics

whether they have been tested (q), if the test was positive ($+$), and whether they are aware of being sick (a).

3.1 Modeling Elements for MABEM

Time Discretization We discretize time to capture the dynamic behavior of individuals across the days. Specifically, we divide each day ℓ into P time periods, where a typical duration for each period p can be 30 minutes. The simulation spans L days, resulting in a total of $T = PL$ time intervals, where $t = 1, \cdots, T$, $p = 1, \cdots, P$ with $P = 48$, and $\ell = 1, \cdots, L$. For this reason, we can define t as:

$$t = p + P\ell. \tag{1}$$

As a consequence of Equation (1), each timestep t corresponds to a couple (p, ℓ) such that:

$$p(t) = t \bmod P \tag{2}$$

$$\ell(t) = \left\lfloor \frac{t}{P} \right\rfloor \tag{3}$$

where $\lfloor\rfloor$ and $\bmod$ represent the quotient and the remainder of the integer division t/P, respectively[1]. We say that a specific time interval t' belongs to a specific day ℓ', i.e., $t' \in \ell'$, if and only if:

$$\ell(t') = \ell'$$

Similarly, we define a total number of weeks, where $w = 1, \cdots, \lfloor \frac{T}{7} \rfloor$. A week w can be linked with t by defining:

$$w = w(t) = \left\lfloor \frac{\ell(t)}{7} \right\rfloor = \left\lfloor \frac{t}{(7P)} \right\rfloor \tag{4}$$

and we say that t' belongs to a week w', or $t' \in w'$ if and only if:

$$\left\lfloor \frac{t'}{(7P)} \right\rfloor = w'.$$

[2]

[1]For example, if $t = 50$, then $\ell(50) = 1$ and $p(50) = 2$, i.e. the simulation is at hour 1 of day 2.

[2]While this level of detail might seem excessive, it is necessary for accurately mapping time intervals to specific days and weeks, given that the schedule varies daily. The equations ensure precise synchronization within the simulation.

Space and Facilities The space is represented by facilities f which belong to a discrete set $\mathcal{F}$ of locations. Each individual n visits a facility f at each time t. Each facility f has a set of $\mathcal{J}$ characteristics constant over time $\widehat{y}_f^j$ [3]. The different characteristics include information like the square meters of the facility, ventilation level, or number of floors. We also consider a discrete list of A activities that individuals can perform during the day. Each activity a is associated with a set of locations or facilities $\mathcal{F}_a$.

Population Characteristics We consider a generic population $\mathcal{N}$ consisting of N individuals. For each individual n, we include a set of $\mathcal{E}$ socioeconomic characteristics $\widehat{y}_n^e$, which remain constant over time (e.g., income, education level, age, or political orientation). Additionally, we include a set of $\mathcal{H}$ health characteristics $\widehat{y}_n^h$, also constant over time (e.g., body mass index, previous lung diseases, or smoking habits).

Based on their socioeconomic characteristics, the population can be segmented into various groups, such as age, gender, or residence location. We introduce the index g to represent these groups, where $g(n)$ denotes the specific group to which an individual n belongs. It is important to note that these groups form a partition of the population, meaning every individual belongs to exactly one group, and no group overlaps with another. Finally, for each individual n, we define the timestep t_n^+ as the last timestep they test positive, and a recovery time γ_n, representing the duration from the time of infection t_n^i until the individual recovers.

Restriction Policies We define a measure u which involves activating a set of parameters from a set $\mathcal{U}$. Each element in $\mathcal{U}$ is defined as a set of binary parameters, that activate a specific restriction, i.e. not participating in a specific activity, curfew, or delaying the starting time and shortening the end time, for a given activity a.

Individual Schedules We begin by importing a schedule for a reference day which provides the location of each individual at each 30-minute interval p. For each individual $n \in \mathcal{N}$, facility $f \in \mathcal{F}$, and $p \in [1, \ldots, 48]$, we assume that each individual is assigned to a specific location during each time interval. This is represented as:

$$\widehat{y}_{fnp} = \begin{cases} 1 & \text{if individual } n \text{ is at location } f \text{ during time interval } p, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

[3]Note that we use a hat notation (e.g., $\widehat{y}_f^j$) for quantities that remain constant throughout the simulation, while quantities without a hat are those manipulated by the simulator.

This reference schedule serves as the baseline input to our simulation model and repeats every day ℓ, i.e., every 48 time-steps if no restrictions are modified. To obtain the reference schedule, we can use the output of a microscopic activity-based model, such as such as MATSim, or TASHA (Axhausen, 2016; Yasmin et al., 2015), which generate detailed activity schedules for N individuals. Alternatively, GPS tracking data from mobile phones or other mobility datasets could be used, as long as they provide information about individuals' locations at each timestep and their encounters with others. The choice of data source depends on the availability of data and the level of detail required for the study.

This approach is used because we are not employing a multi-day activity-based model, and it also optimizes computational performance, enabling us to integrate the simulation with optimization tools for policy decision-making. However, the framework is flexible and allows for different schedules to be preloaded for each day if needed, accommodating more complex activity patterns. The actual schedule y_{fnt} for individual n at facility f during time interval t:

$$y_{fnt} = \begin{cases} 1 & \text{if individual } n \text{ is at location } f \text{ during time interval } t, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

is then determined by both the reference schedule and the new schedule given restriction u for individual n, during day ℓ. This can be expressed as:

$$y_{fnt} = \begin{cases} \widehat{y}_{fnp(t)} & \text{if no restriction } u \text{ is applied,} \\ \widehat{y}^{u}_{fnp(t)} & \text{otherwise.} \end{cases} \tag{7}$$

where $\widehat{y}^{u}_{fnp(t)}$ is the new reference schedule for individual n given restriction u.

From Facility to Individual Characteristics The definition of variable y_{fnt} in Equation (7) allows us to associate the characteristics of the facility to the individuals present in it at a given time t. We define y^{j}_{nt} as the characteristic j of the facility that n visits at time t as:

$$y^{j}_{nt} = \sum_{f} y_{fnt} \widehat{y}^{j}_{f}, \quad \forall j \in \mathcal{J}. \tag{8}$$

Health States The health state of individual n at time t can be susceptible (s), infected (i), or recovered (r), and is captured by three binary variables. This is a classical representation in the epidemiology literature, commonly referred to as the SIR model (see Kelman, 1985). We define them as:

$$Z^{s}_{nt} = \begin{cases} 1, & \text{if individual } n \text{ is susceptible at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$Z^{i}_{nt} = \begin{cases} 1, & \text{if individual } n \text{ is infected at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

$$Z^{r}_{nt} = \begin{cases} 1, & \text{if individual } n \text{ is recovered at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

such that it satisfies the constraint:

$$Z^{s}_{nt} + Z^{i}_{nt} + Z^{r}_{nt} = 1 \quad \forall n, t. \tag{9}$$

Testing Choices and Outcomes Moreover, to model the choice of testing and the positive outcome of the test, we define Z^{q}_{nt}, and Z^{+}_{nt}, where:

$$Z^{q}_{nt} = \begin{cases} 1 & \text{if individual } n \text{ decides to test at timestep } t, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

$$Z^{+}_{nt} = \begin{cases} 1 & \text{if individual } n \text{ tests positive at timestep } t, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Awareness We define a period of ν days during which an individual remains aware of their infection status. This awareness period begins when the individual becomes aware of their infection, modeled by the binary variable Z^{a}_{nt}, defined as:

$$Z^{a}_{nt} = \begin{cases} 1 & \text{if individual } n \text{ is aware of being infected at timestep } t, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

This awareness is determined by the individual's health status and the result of a positive test, represented mathematically as:

$$Z^{a}_{nt} = 1 \quad \forall t : t^{+}_{n} < t < t^{i}_{n+\nu}. \tag{11}$$

After this awareness period of ν days, if no new positive test occurs, the individual's awareness resets and $Z^{a}_{nt} = 0$.

From Infectious Contacts to Individual Characteristic Equation (12) calculates the proportion of infected individuals that individual n encounters at time t by dividing the total encounters with infected individuals across all locations by the total encounters with all individuals, excluding n. The proportion is used instead of the absolute number because it better represents the risk of infection,

considering the overall context. Encountering a higher proportion of infected individuals in a group generally increases the risk of exposure, as the likelihood of contact with an infected person is greater.[4]

$$y^{v}_{nt} = \frac{\sum_{f} \sum_{m \in \mathcal{N} - \{n\}} y_{fmt} \cdot Z^{i}_{fmt}}{\sum_{f} \sum_{m \in \mathcal{N} - \{n\}} y_{fmt}}. \tag{12}$$

Latent States The latent states in our model are continuous variables designed to capture underlying and unobserved processes. Specifically, we define $Q^{\star}_{nt}$ as a latent variable representing the propensity of an individual n to seek testing at time t. Additionally, $E^{\star}_{nt}$ is another latent variable that defines the level of exposure of individual n to the disease at time t, reflecting both environmental factors and interactions that could lead to transmission. These latent states allow the model to account for the variability in testing behavior and exposure risk over time and across individuals, and their unit and scale is arbitrary.

Data In terms of infection information, we define $\widehat{y}^{q}_{nt}$ as a binary variable that defines whether n has tested at t. We assume to have at our disposal, the observed number of positive cases $\widehat{y}^{+}_{gw}$, the observed number of negative cases $\widehat{y}^{-}_{gw}$, and the observed number of tested individuals $\widehat{y}^{q}_{gw}$ for each age group g of the population and each week w. Finally, we assume to have at our disposal for each day ℓ, a variable $\widehat{y}^{+}_{g\ell}$ that contains all positive cases for that day, per group g. These data variables are used to estimate the parameters of the epidemiological behavioral choice model and as initialization values.

3.2 MABEM Input and Outputs

The main input of MABEM is $\widehat{y}_{fnp}$ which defines the facilities that an individual visits across one day. The rest of the inputs include the socioeconomic characteristics $\widehat{y}^{e}_{n}$ of the individuals, their health characteristics $\widehat{y}^{h}_{n}$, and the characteristics of the facility that n is visiting at timestep t, y^{j}_{nt}.

As outputs, MABEM generates the following indicators: Z^{s}_{nt}, Z^{i}_{nt} and Z^{r}_{nt}, and Z^{q}_{nt}, Z^{+}_{nt}, and Z^{a}_{nt}. Z^{s}_{nt}, Z^{i}_{nt}, and Z^{r}_{nt} represent the health state of the individual, Z^{q}_{nt} their choice of testing, Z^{+}_{nt} the test outcome, and Z^{a}_{nt} how the test outcomes impact activity-travel behavior.

[4]For example, being in the same location of one infected individual out of ten (10%) represents the same risk proportionally as encountering 20 infected individuals out of 200 (10%). However, the absolute number of encounters could influence the cumulative risk in scenarios where multiple contacts are likely.

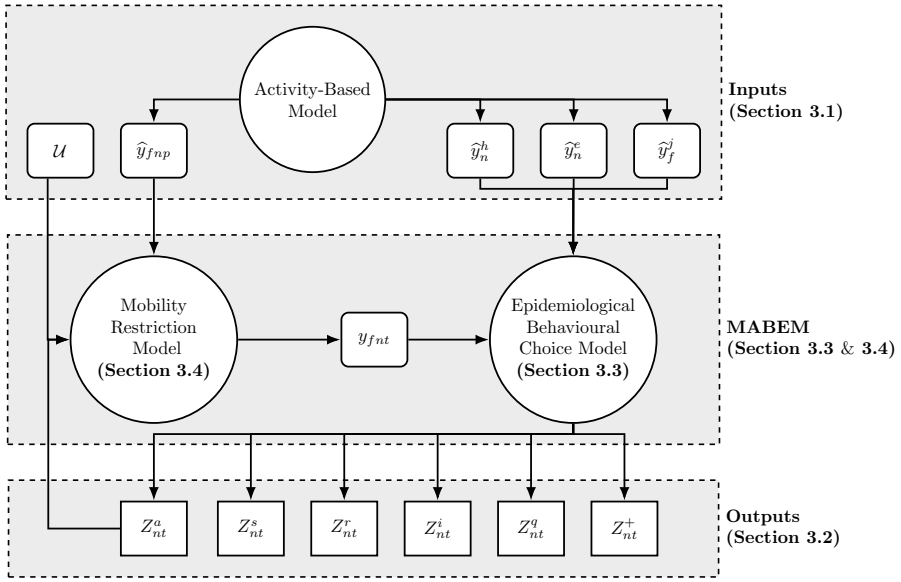

Figure 1: Overview of the methodologies.

In order to provide a clearer understanding of the workflow and interrelationships between the different components of our framework, we include Figure 1 to portray the various models involved visually, the sections where each is detailed, and their respective inputs and outputs.

3.3 Epidemiological-Behavioral Choice Submodel

3.3.1 Model Overview

Figure 2 illustrates the epidemiological-behavioral choice submodel, providing a visual representation of the two latent states (represented in an oval), the explanatory variables (represented in a rectangle), and the observed variable from data (represented in a square shape). The model is fed with all the inputs presented in Sec. 3.1.

The first latent state presented in Figure 2, i.e. the level of exposure $\mathsf{E}_{\mathrm{nt}}^{\star}$, captures the level of contact an individual has with the infection, which depends on several factors ($\widehat{\mathsf{y}}_{\mathrm{n}}^{\mathrm{h}}$, $\mathsf{y}_{\mathrm{nt}}^{\mathrm{j}}$ and $\mathsf{y}_{\mathrm{nt}}^{\mathrm{v}}$). The inclusion of facility characteristics, such as ventilation or square meters, reflects their critical role in influencing disease transmission. For instance, a well-ventilated, spacious area reduces the risk of infection, while a crowded, poorly ventilated space increases it.

From this exposure state, we model the probability that the individual becomes infected, providing a dynamic and individualized measure of risk. The second

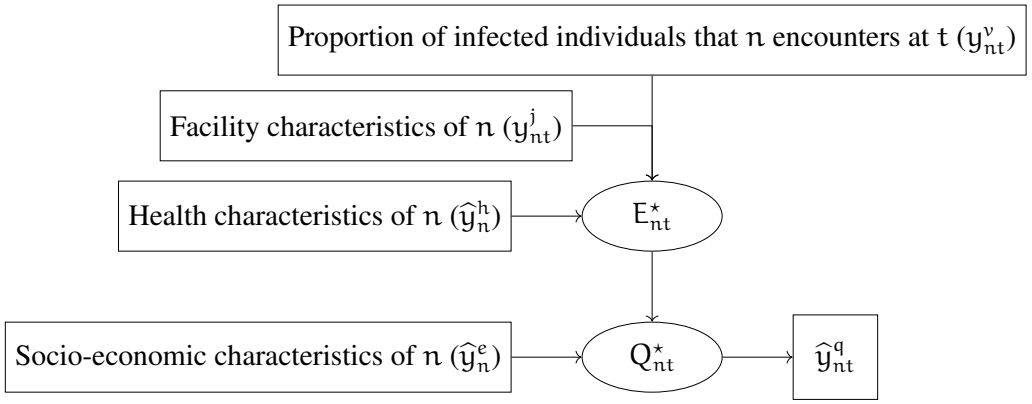

Figure 2: Epidemiological Behavioral Model: Structural Diagram

latent state, propensity to test, is influenced by the individual's socioeconomic characteristics ($\widehat{y}^{e}_{n}$) and their exposure level. This combination allows the model to capture how likely an individual is to seek testing based on their perceived risk. Socioeconomic characteristics are crucial to model the propensity to test, as they strongly impact the likelihood of testing of an individual. This model, therefore, integrates both behavioral and environmental factors to simulate the complex dynamics of disease spread and individual decision-making.

3.3.2 Structural equations

Exposure State Modeling The hidden exposure state $E^{\star}_{nt}$ is a continuous variable that captures the individual's level of exposure to infection. As previously discussed, this latent state takes as exogenous variables the health characteristics of the individual $\widehat{y}^{h}_{n}$, the characteristics of the facility that individual n is visiting y^{j}_{nt}, and the proportion of infectious individuals y^{v}_{nt} that an individual n encounters at timestep t in facility f. We define the structural equation for $E^{\star}_{nt}$ as:

$$E^{*}_{nt} = \beta^{0}_{E^{\star}} + \sum_{h \in \mathcal{H}} \beta^{h} \widehat{y}^{h}_{n} + \beta^{v} y^{v}_{nt} + \sum_{j \in \mathcal{J}} \beta^{j} y^{j}_{nt} + \varepsilon_{E^{\star}}, \tag{13}$$

where β^{h} is the parameter h from $\widehat{y}^{h}_{n}$, $\forall h \in \mathcal{H}$, β^{v} is the parameter of y^{v}_{nt} defined in (12), β^{j} is the parameter j from y^{j}_{nt}, $\forall j \in \mathcal{J}$, defined in (8), $\beta^{0}_{E^{\star}}$ is the intercept, and $\varepsilon_{E^{\star}}$ is the error term. Note that while the current expression of exposure employs a linear term to approximate the relationship between health and facility characteristics and infection risk, we acknowledge the potential benefits of exploring multiplicative or non-linear models in future studies as they might

provide better insights, especially with access to more detailed data.[5]

Propensity to Test To model an individual's propensity to test, we use another latent state defined as:

$$Q^{\star}_{nt} = \beta^{0}_{Q^{\star}} + \sum_{e \in \mathcal{E}} \beta^{e} \widehat{y}^{e}_{n} + \eta_{E^{\star}} E^{\star}_{nt} + \varepsilon_{Q^{\star}}, \tag{14}$$

where β^{e} is the parameter e from $\widehat{y}^{e}_{n}$, $\forall e \in \mathcal{E}$, $\eta_{E^{\star}}$ is the parameter for $E^{\star}_{nt}$, $\beta^{0}_{Q^{\star}}$ is the intercept, and $\varepsilon_{Q^{\star}}$ is the error term. Note that the decision to get tested is not directly tied to an individual's health status; even those infected and aware of their condition might opt for testing, perhaps out of curiosity or to confirm whether they can safely resume their regular activities.

3.3.3 Measurement Equations

As observed data, the choice model presented requires knowing whether an individual n tested or not at each timestep t. However, obtaining such detailed, individual-level data is complicated due to data privacy concerns and the logistical challenges of tracking a large number of individuals over an extended period to monitor their testing behavior. This process would involve monitoring a vast population continuously for months, which is both resource-intensive and potentially invasive.

Given these challenges, we propose a method to use aggregated data instead. To link the tests performed for each individual and timestep, with the observed number of tests $\widehat{y}^{q}_{gw}$ per segment of the population g and per week w, we define N^{q}_{gw} as the number of individuals in group g who get tested in a week w. Similarly, to link the positive tests per individual and timestep, with the observed number of positive tests $\widehat{y}^{+}_{g\ell}$ per group g and per day ℓ, we define $N^{+}_{g\ell}$ as the number of individuals in the segment who get tested in day ℓ, as:

$$N^{q}_{gw} = \sum_{n \in g} \sum_{\forall t \in w} P(Z^{q}_{nt} = 1), \tag{15}$$

$$N^{+}_{g\ell} = \sum_{n \in g} \sum_{\forall t \in \ell} P(Z^{+}_{nt} = 1), \tag{16}$$

where $P(Z^{q}_{nt} = 1)$ is the probability of an individual n to test at timestep t, and $P(Z^{+}_{nt} = 1)$ is the probability of an individual n to test positive at timestep

[5]The lack of data does not allow for including protective measures, such as facemask usage, which are not explicitly modeled. However, the impact of facemasks can be conceptualized as a multiplicative factor in the exposure equation, reducing the overall exposure risk. This factor can decay over time, reflecting varying compliance and effectiveness.

t, defined below in Equations (25) and (26)–(27), respectively. To estimate our model, we need to provide a connection to our data. This connection is established through the log-likelihood functions $\mathcal{L}_1(\theta)$ and $\mathcal{L}_2(\theta)$, where $\theta = [\beta^e, \beta^h, \beta^j, \beta^v, \eta_{E^\star}, \beta_0^{Q^\star}, \beta_0^{E^\star}, r_1, r_2]$ is the vector that contains all the decision variables, with β^e for all $e \in \mathcal{E}$, β^h for all $h \in \mathcal{H}$, and β^j for all $j \in \mathcal{J}$. To describe these functions, we opt to use the negative binomial distribution for modeling the count data represented by $\widehat{y}^q_{gw}$ and $\widehat{y}^+_{g\ell}$ primarily due to its flexibility in handling overdispersed data. The Poisson distribution, traditionally used in epidemiological studies for its simplicity and the property of the mean equaling the variance, is often inadequate when the data exhibit greater variability than the Poisson model can accommodate (Cameron and Trivedi, 2013). This overdispersion in epidemiological data can be attributed to unobserved heterogeneity in the population, where differences in susceptibility, exposure, and reporting rates across demographic segments can lead to variance that significantly exceeds the mean. The negative binomial distribution introduces an extra parameter (r_1, r_2), which allows it to adjust for this variability by letting the variance be greater than the mean, thus providing a more accurate and fitting model for our epidemiological counts (Hilbe, 2011). Moreover, the negative binomial is particularly suited for data like ours, where the count of positive tests and the number of tests conducted can vary widely across different population groups and times. It effectively captures the stochastic nature of disease transmission and the associated observation processes, which are influenced by varying policy interventions and behavioral responses over time. By employing the negative binomial framework within our log-likelihood functions $\mathcal{L}_1(\theta)$ and $\mathcal{L}_2(\theta)$, we ensure that our model is robust to these complexities, enhancing the reliability and validity of our epidemiological inferences. This methodological choice is not only supported by the nature of our data but also aligns with advanced statistical practices in epidemiological research that require accommodating overdispersion (Lawless, 1987). $\mathcal{L}_1(\theta)$ and $\mathcal{L}_2(\theta)$ are described as:

$$
\begin{aligned}
\mathcal{L}_1(\theta) = \sum_{n \in g} \sum_{w \in W} \Big(& \log \Gamma(\widehat{y}^{q}_{gw} + r_1) - \log \Gamma(r_1) \\
& - \log \Gamma(\widehat{y}^{q}_{gw} + 1) + r_1 \cdot \log \left(\frac{r_1}{r_1 + N^{q}_{gw}} \right) \\
& + \widehat{y}^{q}_{gw} \cdot \log \Big(\frac{N^{q}_{gw}}{r_1 + N^{q}_{gw}} \Big), \\
\mathcal{L}_2(\theta) = \sum_{n \in g} \sum_{\ell \in L} \Big(& \log \Gamma(\widehat{y}^{+}_{g\ell} + r_2) - \log \Gamma(r_2) \\
& - \log \Gamma(\widehat{y}^{+}_{g\ell} + 1) + r_2 \cdot \log \left(\frac{r_2}{r_2 + N^{+}_{g\ell}} \right) \\
& + \widehat{y}^{+}_{g\ell} \cdot \log \Big(\frac{N^{+}_{g\ell}}{r_2 + N^{+}_{g\ell}} \Big).
\end{aligned}
\tag{17, 18}
$$

where r_1 and r_2 are the parameters of the negative binomial. These parameters control the shape of the distribution, particularly how much the variance of the data exceeds the mean, accounting for overdispersion in the observed counts $\widehat{y}^{q}_{gw}$ and $\widehat{y}^{+}_{g\ell}$.

3.3.4 Epidemiological Model

A hidden Markov chain is employed to model the transitions between health states. The model is characterized by the following state transition matrix $\mathbf{B}$:

$$
\mathbf{B}_{n,t} = \begin{bmatrix} B^{s,s} & B^{s,i} & 0 \\ 0 & B^{i,i} & B^{i,r} \\ B^{r,s} & 0 & B^{r,r} \end{bmatrix}_{n,t},
$$

where $B^{k_1,k_2}_{n,t}$ represents the probability of transitioning from state k_1 to state k_2 at any timestep t. Each row of the matrix sums to one, ensuring that the probabilities cover all possible state transitions for an individual at time t. In particular, the transition probability from the susceptible toward the infected health state is calculated using:

$$
B^{s,i}_{n,t} = P(Z^{i}_{n(t+1)} = 1 | Z^{s}_{nt} = 1) = \frac{1}{1 + e^{-\mu E^{\star}_{nt}}}, \tag{19}
$$

where μ is set to 1, reflecting the non-estimable nature of the scale parameter in this model context , and $E^{\star}_{nt}$ is the hidden state of exposure defined in Equation

(13). As a consequence, the probability of remaining susceptible is the complement to one of the probability defined in Equation (19):

$$B^{s,s}_{n,t} = P(Z^{s}_{n(t+1)} = 1 | Z^{s}_{nt} = 1) = 1 - B^{s,i}_{n,t}. \tag{20}$$

If an individual is infected, the probability of them transitioning to the recovered state, accounting for the recovery time γ_n, can be defined as[6]:

$$B^{i,r}_{n,t} = P(Z^{r}_{n(t+1)} = 1 | Z^{i}_{nt} = 1) = t - t^{i}_{n} > \gamma_n. \tag{21}$$

and, similarly to Equation (20):

$$B^{i,i}_{n,t} = P(Z^{i}_{n(t+1)} = 1 | Z^{i}_{nt} = 1) = 1 - B^{i,r}_{n,t}. \tag{22}$$

Finally, we assume that an individual who recovers stays recovered until the end of the simulation. Following this logic, we define:

$$B^{r,r}_{n,t} = P(Z^{r}_{n(t+1)} = 1 | Z^{r}_{nt} = 1) = 1, \tag{23}$$

$$B^{r,s}_{n,t} = P(Z^{s}_{n(t+1)} = 1 | Z^{r}_{nt} = 1) = 0. \tag{24}$$

3.3.5 Behavioral Model

At each timestep t, an individual can decide to get tested. As previously discussed, we assume that the propensity of an individual to get tested depends on their socio-economic characteristics, and their exposure, described in Equation (14). We model the probability of an individual deciding to get tested through a logit model:

$$P(Z^{q}_{n(t+1)} = 1) = \frac{1}{1 + e^{-\mu Q^{\star}_{nt}}}, \tag{25}$$

μ is set to 1, with the same reasoning as in Equation (19). For individuals who are tested, that is when $Z^{q}_{nt} = 1$, we compute the outcome of the test. To accurately reflect the test results, we consider the possibilities of true positives in Equation (26) and false positives in Equation (27) from the literature (see Pecoraro et al., 2022 and American Society, 2022).

$$P^{\text{true positive}}_{nt} = P(Z^{+}_{nt} = 1 | Z^{q}_{nt} = 1 \text{ and } Z^{i}_{nt} = 1) = 0.88 \pm (0.86 - 0.90), \tag{26}$$

$$P^{\text{false positive}}_{nt} = P(Z^{+}_{nt} = 1 | Z^{q}_{nt} = 1 \text{ and } Z^{i}_{nt} = 0) = 0.02 \pm (0.01 - 0.03). \tag{27}$$

[6]This probability can be calculated using the cumulative distribution function (CDF) of the distribution representing the time to recovery γ_n. In this case, we assume that γ_n follows a log-normal distribution as explained in Appendix B.

For simplicity, we assume that these probabilities remain constant throughout the simulation, across all individuals n and timesteps t. The ranges in parentheses correspond to the 95% confidence intervals. Note that after deciding to get tested, individuals continue with their scheduled activities until they receive test results. This approach assumes no immediate change in behavior during the waiting period.

3.4 Mobility Restriction Submodel

Positive test results on infected individuals trigger a user-defined awareness period of ν days, during which individuals modify their behavior, influencing the overall infection spread. For those who test positive and are infected (see Equation (11)), awareness (Z^a_{nt}) is activated as defined in Equation (10) and remains active for ν days. The mobility restriction submodel defines the activity-travel behavior measures u which sets user-defined rules on how aware individuals should behave. This triggers an activity-based model (for example, the activity-based model defined in Cortes Balcells et al., 2024) passing u as an input in order to obtain the new schedule for each aware individual $\widehat{y}^u_{fnp}$.

3.5 Simulation

The schematic in Figure 3 provides an overview of the dynamics in our framework, illustrating the sequential process each individual undergoes during the simulation. The process begins with the Initialization phase, where each individual's daily schedule and initial health state, whether Susceptible, Infected, or Recovered, are established.

Following initialization, the model calculates each individual's exposure level based on location and interactions, subsequently influencing their propensity to get tested. For each susceptible individual n at time t, their infection status at the next time step $t+1$ is determined by sampling from a Bernoulli distribution, using the calculated probability of becoming infected:

$$Z^i_{n(t+1)} \sim \text{Bernoulli}(B^{s,i}_{n,t}). \tag{28}$$

If the individual is infected (as opposed to being susceptible or recovered), we determine their recovery status at time $t+1$ by sampling from another Bernoulli distribution:

$$Z^r_{n(t+1)} \sim \text{Bernoulli}(B^{i,r}_{n,t}). \tag{29}$$

Also, for each individual n at time t, the decision to get tested is modeled as a Bernoulli random variable::

$$Z^{q}_{n(t+1)} \sim \text{Bernoulli}(P(Z^{q}_{n(t+1)} = 1)), \tag{30}$$

The outcome of the test $Z^{+}_{n(t+1)}$ for each tested individual is then determined by:

$$Z^{+}_{n(t+1)} \sim \text{Bernoulli}\left(Z^{i}_{nt} P^{\text{true positive}}_{nt} + (1 - Z^{i}_{nt}) P^{\text{false positive}}_{nt}\right), \tag{31}$$

which captures the test result for an individual at time $t + 1$, determining whether the result is positive or negative based on the individual's infection status and the associated probabilities of true and false positives.

In Figure 3, decision points are represented by diamonds, which guide the flow based on conditions such as whether an individual is selected for testing or if their health state transitions from susceptible to infected. These decisions lead to different outcomes, including testing results and subsequent actions like confinement, which are depicted as rectangular blocks. The model also incorporates feedback mechanisms, where outcomes like self-isolation feed back into recalculating exposure. Overall, the figure captures the iterative and conditional nature of the model, showing how individual states are updated over time through interactions.

3.6 Estimation of the parameters

To estimate the parameters of the model, we need to maximize the sum of the likelihoods defined in Equations (17) and (18), as follows:

$$\max \mathcal{L}(\theta) \tag{32}$$

subject to:

$$\beta^{e}, \beta^{h}, \beta^{j}, \beta^{v}, \eta_{E^{\star}}, \beta_{0}^{Q^{\star}}, \beta_{0}^{E^{\star}} \in \mathbb{R} \quad \forall e \in E, \forall h \in H, \forall j \in J \tag{33}$$

$$0 < r_1, r_2 < 1 \tag{34}$$

where the objective function is defined as $\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \mathcal{L}_2(\theta)$, $\mathcal{L}_1(\theta)$ is defined in (17), and $\mathcal{L}_2(\theta)$ is defined in (18). θ represents the vector of decision variables, and the constraints include boundaries for each element of θ.

As seen in Equation (12), y^{v}_{nt} is endogenous over time, complicating the model estimation since it depends on parameters that evolve throughout the simulation. This interdependence makes it challenging to estimate the model's parameters without running the full simulation. For this reason, we develop an algorithm to estimate the model parameters to tackle this issue. The algorithm, as defined in Algorithm 1, relies on iteratively solving the optimization problem defined in

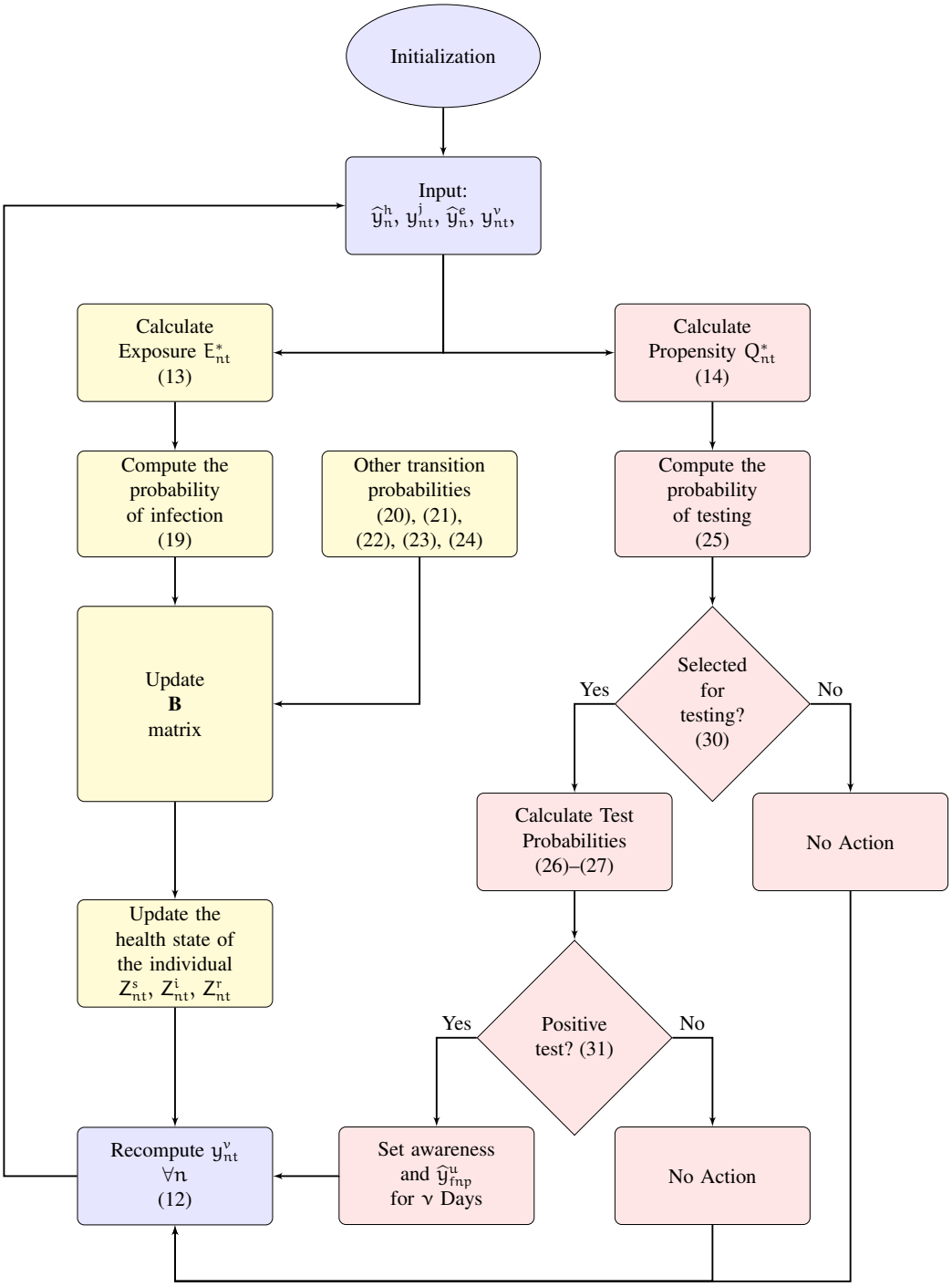

Figure 3: Model Dynamics

Equations (32)–(34). Initially, the algorithm starts with an arbitrary set of parameter values, which may result in a high Mean Squared Error (MSE) between two sets of simulated values of y_{nt}^{ν}, obtained under varying parameter sets or simulation conditions. Through successive iterations, it refines the parameters (θ^*) to better align these simulated outcomes, effectively calibrating the model. This iterative process continues until the MSE falls below a predefined threshold, indicating an adequate estimation of the model parameters. The approach is based on the Banach fixed-point theorem (see Rudin, 1976), which assumes that the mobility and infection variables will converge to a stable equilibrium during the estimation process. According to the theorem, if the iterative mapping is a contraction, the sequence will converge to a unique fixed point, ensuring that the model parameters stabilize. In our case study, we confirm that the conditions are met, allowing for consistent and reliable parameter estimation as the MSE decreases and convergence is achieved.

Algorithm 1 Model Calibration

Require: Simulation model, Optimization solver, MSE threshold
Ensure: Calibrated parameter values
1: Initialize MSE $= \infty$
2: Run simulation to obtain y_{nt}^{ν} $\forall n$ and t
3: Compute (15) and (16) using y_{nt}^{ν}
4: **while** MSE > 0.05 **do**
5: Solve optimization problem (32) with constraints (33)–(34) to find θ^*
6: Run simulation with θ^* to obtain $y_{nt}^{\nu'}$
7: Update MSE between $y_{nt}^{\nu'}$ and y_{nt}^{ν}
8: **end while**
9: Output calibrated parameter values θ^*

To solve this problem, we use the `'trust-constr'` solver from SciPy, suitable for non-convex bounded optimization problems. At each iteration, we start the optimization from five different initial points for θ to ensure that the solution is unique and to avoid local minima.

4 Case Study

The proposed framework is validated on a case study concerning the population of Vaud in Switzerland, comprising 810,486 individuals. The discretization of time is fixed at 30 minutes, which was chosen to balance the computational cost of the simulations and the accuracy in capturing activity-travel behavior dynamics. A finer resolution, while potentially offering more precise behavioral insights, would significantly increase the computational burden, making it impractical for the scale of our study covering 810,486 individuals over 77 days. Conversely, a lower frequency discretization could overlook important patterns in how individuals schedule and conduct their activities, potentially leading to less accurate predictions of disease spread patterns. The 30-minute interval represents a compromise allowing detailed behavior modeling without overextending computational resources, thereby maintaining the feasibility of extensive simulations while still providing robust analytical value. The length of the simulation is 77 days, i.e. 11 weeks. In this section, we present: i) the data used for our case study, ii) the final model specification given the available data, iii) the calibrated output analysis, iv) the behavioral changes before and after testing positive, v) the geographical distribution of infection spots, vi) and the impact of self-quarantine duration on the disease spread.

4.1 Data

For the epidemiological data, we use two primary sources: Google Cloud Platform (CloudPlatform, 2021) and the Swiss Federal Office of Public Health (FOPH) Riou et al., 2021. The data from the Google Cloud Platform provides the total number of tests and positive tests per week, disaggregated by 10 age classes g, each spanning a range of 10 years. Additionally, the data from FOPH includes the total number of positive tests, further broken down by age (from 0 to 100 years), gender (male and female), and municipality. Both datasets cover the period starting from February 2020, until September 2021. This data is used to calibrate the parameters of the epidemiological behavioral choice model and to initialize the simulations. It is important to note that these figures represent officially reported cases and may not fully account for the prevalence of self-testing, specifically self-testing at home, which can lead to discrepancies in the observed and actual incidence rates. This consideration is important in interpreting the data used to calibrate the parameters of the epidemiological behavioral choice model and to initialize the simulations.

As mobility data, we use the population input of MATSim that includes the socioeconomic characteristics of the population of Switzerland, such as age and gender (see Table 3 in Appendix A). Secondly, we run the Switzerland MATSim

scenario provided by Horl and Balac, 2021, and obtain the output of MATSim, which contains the schedule of each individual for a generic day ℓ. The schedule includes detailed information about each individual's activities throughout the day, specifying the type of activity (home, work, education, secondary activities, shop, and public transportation), geographic coordinates, start and end times, and the facility id or transport mode at which the activity takes place. This data allows us to obtain two main inputs for the model: the binary variable y^{ν}_{fnt}, which defines the facilities that an individual visits for every time t, and the vector of socioeconomic characteristics $\widehat{y}^{e}_{n}$ for each individual. However, it does not provide information on the health characteristics $\widehat{y}^{h}_{n}$ or the detailed facility characteristics y^{j}_{nt}. This limitation restricts the model's ability to fully capture the complexity of exposure, thereby limiting its potential to comprehensively explain the factors influencing disease transmission.

4.2 Calibration of the Epidemiological-Behavioral Choice Submodel

Structural equations Given the unavailability of disaggregated data to train our model, which ideally would include individual socio-economic and health characteristics, information on positive and negative tests, as well as their daily schedule, we develop a simplified model with a few characteristics to be able to calibrate it. For this reason,E^{*}_{nt} from Equation (13) only depends on the proportion of infected individuals per timestep, y^{ν}_{nt}. The individual's propensity to test ($Q^{\star}_{nt}$) from Equation (14) includes $\widehat{y}^{\text{age}}_{n}$ and $\widehat{y}^{\text{employed}}_{n}$ as socioeconomic characteristics, see Table 3, where β^{ν}, β^{age}, β^{employed},$\beta^{Q^{\star}}_{0}$,and $\eta_{E^{\star}}$ are parameters to be estimated, and $\varepsilon_{Q^{\star}}$ is set to 0. Also, there are two extra parameters to be calibrated, which are the parameters of the negative binomial r_1 and r_2, see Equation (17) and (18).

Calibrated parameter values The parameters of the model are calibrated with data from June 2020 to the end of August 2020, and their values are shown in Table 2, ensuring that the model accurately reflects the observed data.

Specifically, the calibrated infection parameter is $\beta^{\nu} = 0.018$. The positive sign indicates that a higher proportion of infected individuals increases the exposure and therefore the probability of infection. For the testing propensity, the model includes an age coefficient (β^{age}) of 0.03 and an employment status coefficient (β^{employed}) of 0.01 for the employed individuals. Additionally, the model includes an eta parameter ($\eta_{E^{\star}}$) set at 185 and an alternative specific constant for testing ($\beta^{Q^{\star}}_{0}$) set at -8.9. Analyzing the signs of these parameters:

Latent Variable	Parameter	Variable Defined	Value	t-statistic	p-value
E^*_{nt}	β^{ν}	$\widehat{y}^{\nu}$	0.018	11.3	$< 1.00 \times 10^{-3}$
Q^*_{nt}	β^{age}	$\widehat{y}^{e}_{n}$	0.031	-28.6	$< 1.00 \times 10^{-3}$
	β^{employed}	$\widehat{y}^{e}_{n}$	0.010	5.43	$< 1.00 \times 10^{-3}$
	$\beta_0^{Q^\star}$	Q^*_{nt}	-8.91	-48.9	$< 1.00 \times 10^{-3}$
	$\eta_{E^\star}$	E^*_{nt}	185	-95.3	$< 1.00 \times 10^{-3}$

Table 2: Values of the calibrated parameters, their categories, t-statistics, and p-values.

1. The age parameter ($\beta^{\text{age}} = 0.03$) is positive, indicating that older individuals have a higher propensity to get tested.

2. The employment parameter ($\beta^{\text{employed}} = 0.01$) is also positive, suggesting that being employed increases the likelihood of getting tested.

3. The eta parameter ($\eta_{E^\star} = 185$) is positive, showing a strong influence of the exposure level on the propensity to get tested. This high value indicates that as exposure increases, the propensity to test rises significantly.

4. The alternative specific constant for testing ($\beta_0^{Q^\star} = -8.9$) is negative, reflecting a generally low baseline propensity for testing, which means that in the absence of other factors, especially exposure, individuals are not likely to seek testing.

It is worth mentioning that the calibrated parameters reveal a low willingness to get tested ($Q^\star$), consistent with observed data where the maximum weekly positive tests were around 8,000 in a population of 800,000, during a period with no restrictions. For example, even with a high infected fraction of 0.1, the contribution of exposure ($E^\star = 0.0018$) adds only 0.33 to $Q^\star$, keeping it on the far-left side of the logistic function and resulting in very low testing probabilities.

To further validate the robustness of these calibrated parameters, where θ represents the vector of decision variables including β^{ν}, β^{age}, β^{employed}, $\eta_{E^\star}$, and $\beta_0^{Q^\star}$, we perform a sensitivity analysis using bootstrapping. The t-statistics are computed using the formula:

$$\text{t-statistic}_b = \frac{\bar{\theta}_b - \theta_b}{\text{SE}_{\theta_b}}$$

where $\bar{\theta}_b$ is the mean of the bootstrap estimates for the b-th parameter in the vector θ, θ_b is the initial calibrated value of that parameter, and SE_{θ_b} is the standard error of the estimates, calculated as the standard deviation of the bootstrap

samples divided by the square root of the number of samples. The corresponding p-values are obtained using the cumulative distribution function (CDF) of the t-distribution with $c - 1$ degrees of freedom, where c is the number of bootstrap samples. These calibrated values provide insights into how different factors influence testing behaviors. Age and employment status increase the likelihood of getting tested, while exposure significantly raises the propensity for testing, emphasizing the model's sensitivity to perceived infection risk.

4.3 Simulation

We simulate the epidemic incidence from September 21, 2020, to December 13$^{\text{th}}$ 2020, using the calibrated parameters from the previous section. Regarding the initialization process, we force the first week of the simulation to have the same number of tested individuals ($\widehat{y}^{q}_{gw}$) and test positive ($\widehat{y}^{+}_{gw}$) individuals as observed in the real data from Google CloudPlatform, 2021. We use the data from Google containing the number of tested and test-positive individuals per age group (grouped in 10-year intervals) and per week. For the first week, we randomly select individuals from each age group, randomizing the process per timestep, ensuring that the chosen number of individuals tested matches the real data. We compare the predictions of our epidemiological model against observed data to evaluate its accuracy and validity. In the simulation for this study case, MABEM considers an 80% probability of self-isolation at home in case of awareness [7]. If the individual does not self-isolate, they continue their usual activities, potentially contributing to ongoing disease transmission. Now, for the individuals that are aware, their schedule is modified so that they stay home $\forall t \in [\ell(t^{i}_{n}) + 1, \ldots, \ell(t^{i}_{n}) + 1 + \nu]$. We assume that individuals complete their schedule and begin self-isolating the day after becoming aware. Since the model does not capture whether an individual is symptomatic, we assume that all infected individuals who test positive are likely to self-isolate. We impose a closure restriction $\forall a$, where $a \neq$ home, which means swapping all the activities of $\widehat{y}_{fnt}$ and replacing them for home. This assumption improves computational efficiency and aligns with the framework's primary focus on the effects of self-isolation on mobility dynamics and disease spread, rather than on activity substitution due to restriction policies, even though the framework is flexible enough to accommodate other implementations.

[7] This percentage reflects the high compliance rates observed in Switzerland during the COVID-19 pandemic. This rate is supported by analysis from the KOF Swiss Economic Institute (see Pleninger et al., 2021), which highlights the effectiveness of containment measures and the substantial public adherence to self-isolation guidelines.

4.4 Validation

4.4.1 Infection Rates vs. Positive Tests

In order to compare the aggregated simulated data with the observed data, we present Figure 4. Figure 4 shows the epidemic incidence and COVID-19 testing trajectory including five trajectories: "Simulation - Daily New Infections," "Simulation - Daily Positive Tests," "Test Data - Daily Positive Tests," "Simulation - Daily Tests," and "Test Data - Daily Tests." The x-axis represents the weeks, while the y-axis represents the count of individuals. In addition, the filled areas around the simulation curves represent the 95% confidence intervals, which are computed using a rolling standard deviation with a window of 3 weeks. These intervals provide a visual representation of the uncertainty in the simulation results, offering insight into the potential variability in the epidemic incidence and testing trajectories.

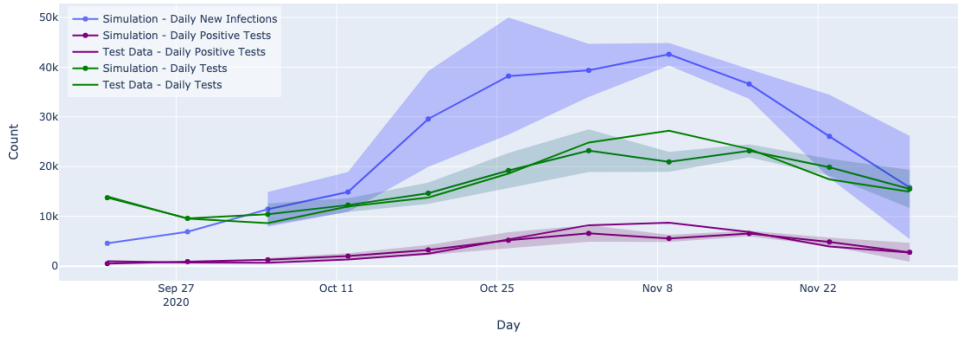

Figure 4: Epidemic Incidence and COVID-19 Testing Trajectory: A comparison between simulated data and actual test data over time.

In Figure 4, the simulated data gives an estimate of the unobserved number of total cases, which is significantly higher than the reported cases. This indicates that many infections go undetected due to limited testing, asymptomatic cases, or individuals choosing not to get tested. Indeed, the infection curve is approximately four times higher than the curve for positive tests at the weekly peak, highlighting the significant impact of individual testing choices on the observed data. To further quantify the discrepancy between the reported and predicted values, we calculate the Root Mean Squared Error (RMSE), which measures the average error between actual and predicted data points. The RMSE between the reported number of tested individuals and the simulated estimates is 2067.78, which if divided by the total tests, represents the 0.01%. These results indicate that the prediction

for total tests aligns closely with the reported data, providing a preliminary validation of our results with respect to the total number of tests. Similarly, for the positive tests, we obtain an RMSE of 1114.94, which represents the 0.02% of the total positive tests. This result provides a first validation regarding the number of tests and positive tests, suggesting that our model's predictions are consistent with the reported data. Our calibrated model, therefore, validates the hypothesis that testing choices significantly affect observed infection rates and can distort the true spread of the disease. This has important implications for public health policies, emphasizing the need for widespread testing and encouraging individuals to get tested to understand better and control the epidemic's spread.

4.4.2 Epidemiological Data and Comparative Analysis

To further validate our model, we propose a comparative analysis with established epidemiological models to substantiate the robustness and accuracy of our simulations. We use data from Scire et al., 2023 which offers estimates of the effective reproduction number $\widehat{y}^{R_e}$ for Switzerland over the same study period, complete with confidence intervals. Their results address underreporting by estimating the actual $\widehat{y}^{R_e}$, acknowledging that not all infections are documented due to discrepancies in testing practices. The estimates from Scire et al., 2023 are derived using advanced deconvolution techniques to refine raw epidemiological data. Deconvolution techniques are mathematically straightforward methods to infer the incidence curve of an epidemic from a recorded daily death curve and time-to-death distribution. This approach allows for more accurate representations of the infection dynamics, as detailed in Goldstein et al., 2009. To be able to establish a comparison with these values, we compute the y^{R_e} values for our MABEM. This requires determining the aggregated force of infection λ, and the aggregated recovery rate γ, thus defining the y^{R_e} as the ratio between λ and γ. We employ a simple, fully aggregated SIR model as outlined in Kermack et al., 1927, to derive λ and γ. The model initialization is based on the reported number of positive tests from the first week, mirroring the approach used for MABEM. We use a Minimum Square Error (MSE) objective function to fine-tune λ and γ to ensure the simulated trajectory of infections (green line in Figure 5) aligns closely with the MABEM simulated data (blue line in Figure 5). Finally, we juxtapose the reproduction number $\widehat{y}^{R_e}$ from Scire et al., 2023 (dashed red line in Figure 5) against our model's reproduction number, y^{R_e}, achieving a difference in percentage of 0.004. This low RMSE corroborates the accuracy of our model, particularly in capturing underreporting rates that are remarkably consistent with those reported by Scire et al., 2023, further validating our approach.

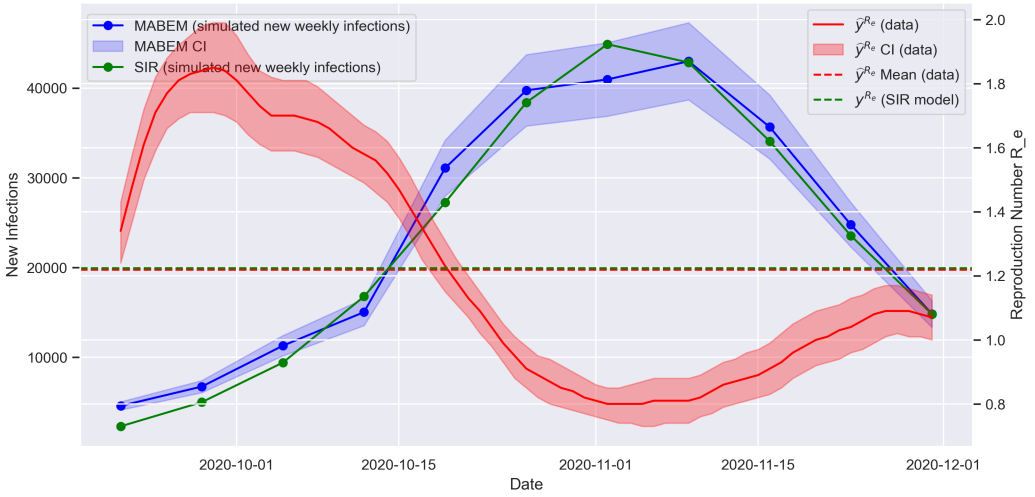

Figure 5: Comparison of MABEM simulated new weekly infections with SIR model outputs and observed $\widehat{\mathrm{y}}^{\mathrm{R_e}}$ values, illustrating the alignment with empirical data and model predictions.

Figure 5 illustrates the comparison of the simulated infection rates from our model against the benchmark values and the corresponding confidence intervals, highlighting the alignment of our model with the empirical data reported by ETH Zurich, affirming the credibility of our simulation approach. Moreover, this validation demonstrates that our model not only aligns with the reported $\widehat{\mathrm{y}}^{\mathrm{R_e}}$ values but also reflects realistic disease spread dynamics as captured by advanced epidemiological studies. The consistency of our model's outputs with established data underscores its potential utility in policy-making contexts, particularly in scenarios requiring quick and accurate assessment of public health strategies.

4.4.3 Qualitative Positive Test Analysis

Finally, to further validate our findings, we plot Figure 6 to study the correlation between the difference in new infections and newly tested positives from the simulation compared to the percentage of positive tests from CloudPlatform, 2021. The upward trend observed in this correlation plot substantiates the model's ability to capture the real dynamics of infection transmission, highlighting a consistent increase in the proportion of newly tested positives alongside rises in simulated new infections. This trend is particularly significant given that over 35% of tests are positive during the peak period, which starkly contrasts with the new infections comprising less than 1% of the total population each week, indicating substantial underreporting (Lau et al., 2021; Albani et al., 2021). This underreporting is consistent with findings from diverse regions, where the rate of positive tests is

used as an indicator of adequate testing and epidemic control, varying from below 10% in places like Chicago and New York City to 2% in Denmark, as reported in Albani et al., 2021. This correlation not only confirms the effectiveness of our model in reflecting actual COVID-19 transmission trends but also emphasizes the model's utility in providing reliable data for policy formulation and epidemic management.

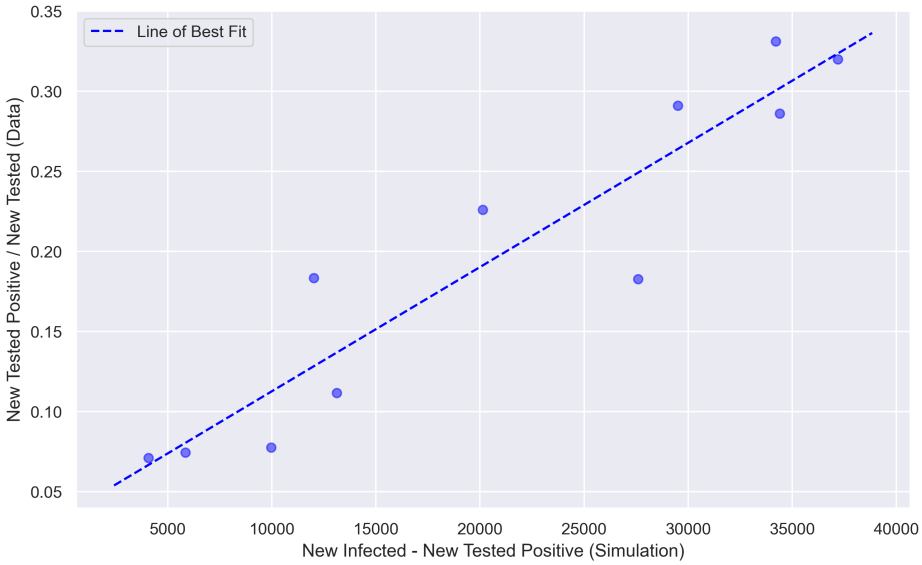

Figure 6: Correlation between the difference in new infections and newly tested positives from simulation and the percentage of positive tests, validating the model's capability to capture real infection dynamics.

4.5 Behavioral Changes in Mobility Patterns Before and After Testing Positive

This section investigates the different possible behavioral responses and scenarios related to COVID-19 testing and infection awareness. We focus on how individuals' awareness of their infection status, influenced by testing results, can affect their mobility patterns and adherence to public health guidelines. We illustrate various cases showing how different individuals might behave based on their infection status and test results.

Figure 7 shows the diversity of behavioral responses to testing and awareness of infection status. Each subfigure describes a specific behavioral pattern of a specific individual from the simulation, reflecting the complex relationship between individual awareness, infection, and testing choices. The subfigure on the left (Id 14239) shows the ideal scenario where an individual, who is already infected on

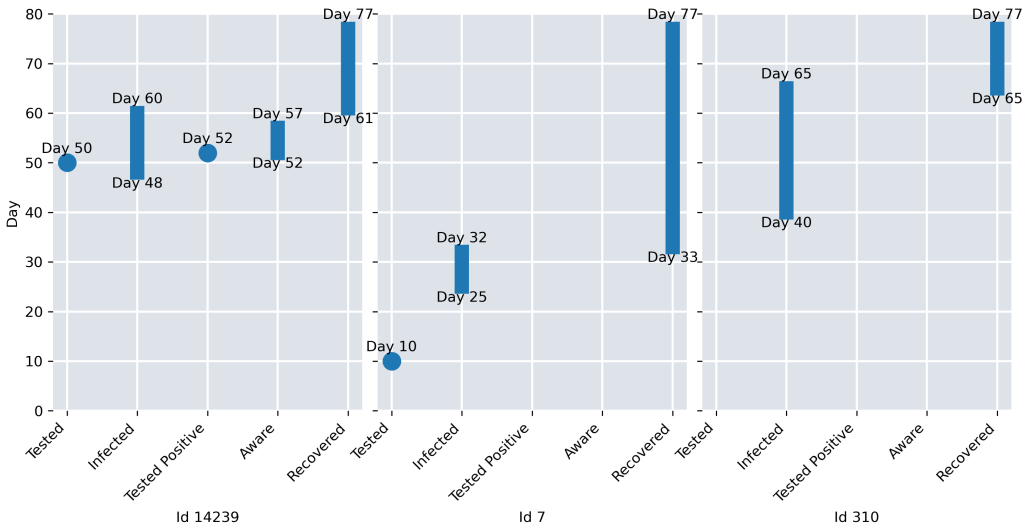

Figure 7: Diversity of behavioral responses to testing and awareness of infection status among individuals.

day 48, tests on day 50, and upon receiving a positive test result on day 52, conscientiously adjusts their travel habits (becomes aware) to mitigate the spread of the virus. This is indicated by $Z^{i}_{nt} = 1$, $Z^{q}_{nt} = 1$, $Z^{+}_{nt} = 1$, and $Z^{a}_{nt} = 1$. Finally, on day 61, they become recovered. The central subfigure (Id 7) presents a scenario where an individual is tested and found not to be infected at that moment (day 10). Later on, it becomes infected but never tested. For this reason, this individual's behavior does not comply with public health guidelines since they are never aware of their infection status. The subfigure on the right (Id 310) shows an individual who is infected but not tested, thus remaining unaware of their infection status and never changing their activity-travel behavior. This scenario underscores the risks associated with undetected infections and the critical role of widespread testing in controlling the spread of the virus.

Through this figure, we observe how the model captures the varying behavior and mobility patterns among individuals. These differences underscore the importance of accurate and widespread testing to inform individuals and enable them to take appropriate actions to mitigate the spread of COVID-19. Our analysis highlights the impact of testing choices on observed infection rates and the actual spread of the disease, demonstrating the need for comprehensive public health strategies to encourage testing and ensure effective disease control.

4.6 Geographical Distribution of Infection Hotspots

This section provides a visual analysis of the geographical spread of COVID-19 infection hotspots over the course of the pandemic, using heatmaps overlaid on a map to represent the intensity and movement of outbreaks. The heatmap, as shown in Figure 8, dynamically displays the local density of cases throughout the landscape, revealing areas of higher transmission risk.

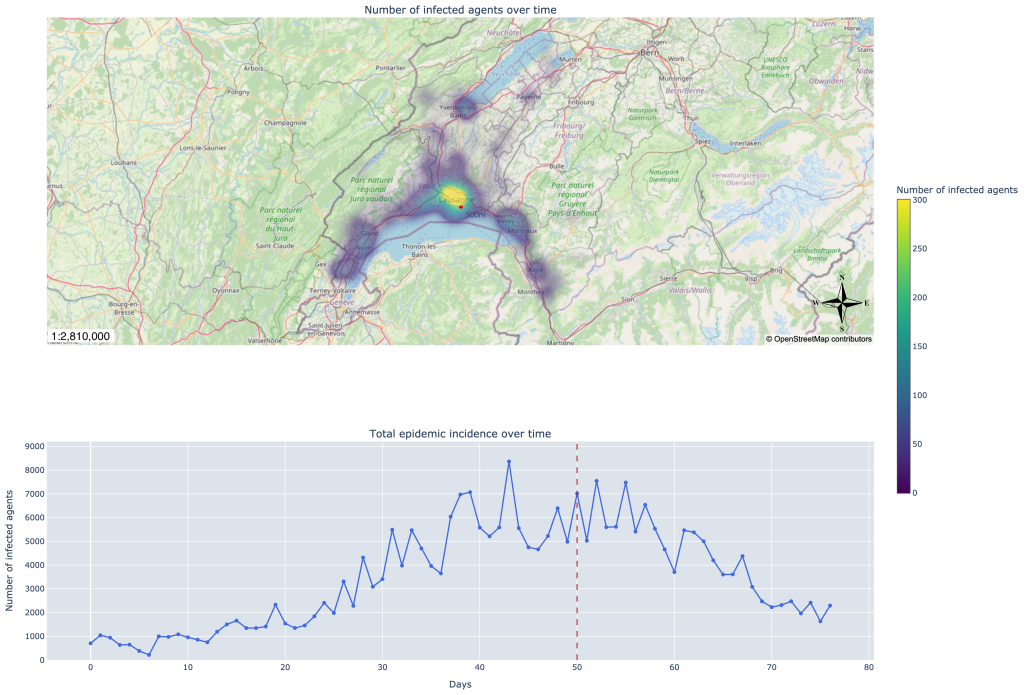

Figure 8: Heatmap visualization of COVID-19 infection hotspots: the upper map highlights the different number of cases throughout the region over the observed period, with color intensities indicating the level of new infections from low (deep purple) to high (yellow). The lower graph tracks the total epidemic incidence over time, displaying the number of new infections each day. The vertical dashed line corresponds to a specific point in time within the epidemic curve, providing a cross-reference between the geographic and temporal data.

The figure indicates that areas with high population density, such as Lausanne, not only show the highest absolute numbers of infections but also have elevated infection rates per capita. Other cities, like Nyon and Vevey, also show significant infection hotspots. This coherence between population density and infection values underlines the importance of focusing public health efforts on densely populated areas to mitigate the spread of the virus, showing the activity-based aspect

of our model.

Beneath the geographical representation, the graph titled 'Total epidemic incidence over time' illustrates the trajectory of the outbreak. The x-axis of the graph measures the days since the initial case was reported, while the y-axis shows the number of new infections. The dashed vertical line marks a specific day in the epidemic's progression, serving as a temporal reference point between the map and the graph.

This visualization is essential for understanding how activity-travel behavior influences disease spread, as more densely populated areas exhibit higher infection intensity. It provides public health authorities and policymakers with a valuable tool to monitor the effectiveness of interventions over time and allocate resources to areas in most urgent need. Additionally, it offers the general public a clear understanding of how the epidemic spreads geographically, underlining areas where increased vigilance and adherence to public health measures are necessary.

4.7 Impact of Quarantine Duration on Total Number of Infections

This subsection examines the impact of different quarantine durations on the total number of infected individuals. By varying the length of the quarantine period, we evaluate the effectiveness of this containment strategy in reducing disease transmission. It is important to remember that only individuals who are aware of being infected, due to testing positive, undergo quarantine.

According to the findings in Ashcroft et al., 2021, a 7-day quarantine combined with a negative PCR test on day 5 results in minimal additional risk of onward transmission compared to longer quarantines without testing. This approach significantly reduces the person-days spent in quarantine, thus lowering the social, psychological, and economic impacts while maintaining epidemiological safety. Our model simulations explored varying quarantine durations (5, 7, 9, 18, 25 days) to assess their effectiveness in controlling the spread of the disease.

Figure 9 illustrates how these varying durations impact the total number of infected individuals. Our results show that while longer quarantine periods effectively reduce the number of infections, shorter quarantines, especially a 7-day period as described by Ashcroft et al., 2021, lead to a 30% reduction in infections compared to a 5-day quarantine. Moreover, we observe that a 7-day quarantine can provide a nearly equivalent epidemiological benefit compared to 9, 18, and 25 with less disruption, allowing individuals to resume normal activities sooner.

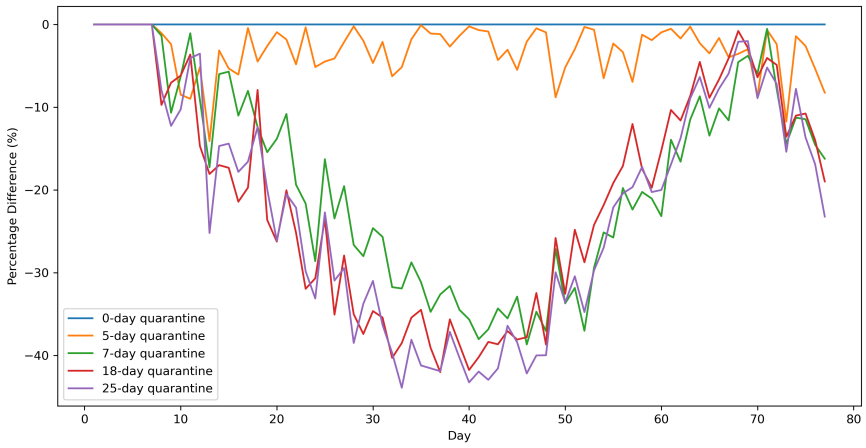

Figure 9: Impact of different quarantine durations on the total number of infections, illustrating the effectiveness of shorter quarantine periods combined with testing.

Our findings advocate for informed decision-making in public health policy, emphasizing the benefits of modeling to optimize quarantine durations over assuming standard durations that might not be optimal for all situations. These results align with our research question, which explores how individual testing choices and subsequent behaviors impact disease spread. The findings demonstrate that compliance with quarantine measures can significantly reduce infection rates, underscoring the importance of adhering to public health guidelines. This analysis offers valuable insights for policymakers to design effective containment strategies and emphasizes the critical role of quarantine in controlling the pandemic.

4.8 Key Insights

The Vaud case study yields several important insights that inform the effectiveness of the model and its implications for public health strategies. Section 4.2 shows that socioeconomic characteristics impact testing choices, although challenges in data granularity limited the model's ability to capture the full complexity of disease transmission. Section 4.4 demonstrates that the model accurately reflects the significant underreporting of COVID-19 infections due to limited testing, with the simulation revealing that the true number of infections is much higher than the reported cases, emphasizing the role of testing in understanding the actual spread of the disease. Section 4.5 highlights significant variability in individual choices to

COVID-19 tests, with infection awareness being crucial for containment. Section 4.6 displays that high-density areas like Lausanne exhibit the highest infection rates, underscoring the need for targeted testing in these regions as a containment method. Finally, 4.7 shows that while extended quarantine periods effectively reduced infection rates, they also posed social and economic challenges, emphasizing the need for a balanced approach.

5 Conclusion

This study investigates the impact of individual testing choices on the spread of disease, with a focus on the COVID-19 pandemic. By developing and validating an activity-based epidemiological behavioral model (MABEM), we address the critical question of how individual testing choices influence activity-travel behavior and, consequently, affect disease spread. Using detailed mobility schedules through an activity-based model combined with epidemiological data, we capture the interactions between testing behaviors and the transmission of disease within the population of Vaud, Switzerland. The calibrated model reveal a significant disparity between the actual number of infections and reported cases, underscoring the extent of underreporting due to limited testing and asymptomatic cases. This finding highlights the need of integrating testing choices into pandemic models to more accurately reflect the true dynamics of disease spread. MABEM also demonstrates its computational efficiency, simulating 810,486 individuals over 90 days with 30-minute timesteps in just 1 minute and 11 seconds on a MacBook Pro (14-inch, November 2023) with an Apple M3 Pro chip and 36 GB of memory, running macOS version 14.6.1. This performance makes it a practical tool for large-scale simulations and public health strategy development.

The key findings of our results include: i) Socioeconomic factors, such as age and employment status, influence testing propensity. This suggests that public health interventions should be tailored to improve testing rates and compliance with health guidelines. ii) There is significant variability in how individuals respond to positive COVID-19 test results, with awareness of infection status being crucial for disease spread containment and changes in daily activities. iii) In regions with high population density, like Lausanne, both the total number of infections and the rate of infection per capita are significantly higher. This finding emphasizes the need for targeted testing policies and health measures in densely populated areas to control and reduce the spread of the disease effectively. iv) While longer quarantine periods significantly reduce infection rates, they can also be highly disruptive. Striking a balance between the effectiveness of quarantine duration and its impact on individual well-being is essential. These

findings emphasize the importance of individual testing choices in understanding the spread of disease and offer valuable insights for policymakers to design effective public health strategies.

Nonetheless, our study comes with some limitations. Firstly, the lack of disaggregated data challenges the model calibration, which impacts the accuracy of our predictions. The lack of epidemiological individual data, including whether an individual tested positive, negative, or did not test each day, makes it very hard to calibrate the model, so adding more parameters becomes a very difficult task. To improve the calibration process of our model, we would need more individual-based data, particularly through surveys that track individual testing behaviors, including the frequency of tests, test results (positive or negative), and associated individual characteristics. This approach would not only improve the precision of our predictions but also allow for a deeper understanding of the factors influencing testing choices. Additionally, the lack of data concerning facility characteristics, such as square meters, ventilation, floors of the building, or rooms in the building, prevents us from fully understanding the most relevant variables to include in the exposure latent state. Future work should include the use of OpenStreetMap and similar resources to enhance the representation of facility characteristics in our model. Moreover, due to these limitations in available data, we do not model specific visits to healthcare facilities for testing purposes. Instead, we streamline the simulation process by assuming that tests occur in the home setting, which aligns with the data's capability to provide detailed location and time information without the granularity needed to model distinct healthcare visits. This is further complicated by the absence of health characteristics of individuals, which would significantly impact the probability of getting infected given the level of exposure. Finally, the lack of symptom data penalizes us in two ways: first, it hinders model calibration since symptoms could be used as observed data for the exposure latent state; and second, it forces us to assume that people are aware of their infection only if they test positive and have symptoms, at which point they will self-isolate. These limitations highlight the need for more comprehensive data collection to improve the robustness and accuracy of the model.

Future research includes more detailed health characteristics and further calibrating the model could enhance its accuracy. In addition to improving the model's precision through more detailed health characteristics and extensive calibration, future versions of the model will integrate with the policy optimization framework from Cortes Balcells et al., 2021, which uses individual activity-travel behaviors to simulate optimal policy outcomes. This integration aims to provide targeted public health strategies that balance health impacts with economic considerations. For example, the model could propose policies that show mortality

with economic trade-offs, offering practical insights for policy decision-making and preparation before epidemic outbreaks.Additionally, future iterations of this model will aim to extend its geographical applicability and relevance by incorporating case studies and scenarios from various global contexts, as suggested in studies such as Anupriya et al., 2022; Kucharski et al., 2021. This expansion will allow us to explore how regional differences in travel patterns, particularly air travel, affect disease spread and public health response effectiveness. Understanding these dynamics will be crucial for developing more universally applicable public health intervention strategies. Furthermore, inspired by recent studies such as Singh et al., 2023, which evaluate the impact of different public health interventions, we plan to add to our model the ability to optimize and evaluate various intervention strategies on top of activity restrictions. This addition will include different policies, such as social distancing mandates, testing frequency, and quarantine measures, giving us insight into how these policies impact disease transmission within different demographic groups and community settings. Also, recognizing the influence that joint decisions, particularly within family units, have on the effectiveness of public health interventions, future iterations of this model will investigate mechanisms for incorporating familial relationships into our modeling framework. This development will include modeling the decision-making processes of parents regarding testing and activity-travel behavior for their children, particularly for school-related activities, which have been identified as critical in the spread of diseases. Finally, applying the model to different cantons and analyzing how testing behavior changes in different sociodemographic populations, and how it impacts the spread of the virus, could offer valuable insights for regional public health policies. These adaptations will enable policymakers to craft strategies that are not only scientifically sound but also culturally and socially informed, enhancing the overall effectiveness of public health interventions.

In conclusion, MABEM emphasizes the importance of individual testing choices and subsequent behaviors in influencing the spread of diseases. This approach opens up new possibilities for developing more effective strategies for disease spread containment by deeply exploring the interplay between individual behaviors, public health interventions, and disease dynamics.

Competing Interests

The authors declare that they have no competing interests.

References

Albani, V., Loria, J., Massad, E. and Zubelli, J. (2021). COVID-19 underreporting and its impact on vaccination strategies, BMC infectious diseases **21**(1): 1111.

Aleta, A., Martín-Corral, D., Bakker, M. A., Pastore Y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini, I. M., Pentland, A., Vespignani, A., Moreno, Y. and Moro, E. (2022). Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas, Proceedings of the National Academy of Sciences **119**(26): e2112182119.
URL: *https://pnas.org/doi/full/10.1073/pnas.2112182119*

American Society, f. M. (2022). Boosting Accuracy, Reducing False Positives of PCR COVID-19 Tests.
URL: *https://asm.org:443/Press-Releases/2022/December/Boosting-Accuracy,-Reducing-False-Positives-of-PCR*

Anupriya, Bansal, P. and Graham, D. J. (2022). Modelling the propagation of infectious disease via transportation networks, Scientific Reports **12**(1): 20572. Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41598-022-24866-3*

Aronna, M. S., Guglielmi, R. and Moschen, L. M. (2021). A model for COVID-19 with isolation, quarantine and testing as control measures, Epidemics **34**: 100437.
URL: *https://www.sciencedirect.com/science/article/pii/S1755436521000025*

Ashcroft, P., Lehtinen, S., Angst, D. C., Low, N. and Bonhoeffer, S. (2021). Quantifying the impact of quarantine duration on COVID-19 transmission, eLife **10**: e63704.

Axhausen, K. W. (2016). The Multi-Agent Transport Simulation MATSim, Ubiquity Press.
URL: *http://www.ubiquitypress.com/site/books/10.5334/baw/*

Balcan, D., Goncalves, B., Hu, H., Ramasco, J. J., Colizza, V. and Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model, Journal of Computational Science **1**(3): 132–145.
URL: *https://www.sciencedirect.com/science/article/pii/S1877750310000438*

Brotherhood, L., Kircher, P., Santos, C. and Tertilt, M. (2020). An Economic Model of the Covid-19 Epidemic: The Importance of Testing and Age-Specific Policies.
URL: *https://papers.ssrn.com/abstract=3618840*

Cameron, A. C. and Trivedi, P. K. (2013). Regression Analysis of Count Data, Econometric Society Monographs, 2 edn, Cambridge University Press, Cambridge.
URL: *https://www.cambridge.org/core/books/regression-analysis-of-count-data/2AB83B406C5798030F7C91ECC99B1BE4*

CloudPlatform, G. (2021). Google Covid data.
URL: *https://github.com/GoogleCloudPlatform/covid-19-open-data/blob/main/docs/table-epidemiology.md*

Cortes Balcells, C., Krueger, R. and Bierlaire, M. (2021). Activity-based modeling and simulation of epidemics.
URL: *http://www.strc.ch/2021.php*

Cortes Balcells, C., Torres, F., Krueger, R. and Bierlaire, M. (2024). Modeling the Influence of Restriction Policies and Perceived Risk due to COVID-19 on Daily Activity Scheduling, Technical Report TRANSP-OR 240703, Transport and Mobility Laboratory, EPFL, Lausanne, Switzerland.

Cui, Y., Ni, S. and Shen, S. (2021). A network-based model to explore the role of testing in the epidemiological control of the COVID-19 pandemic, BMC Infectious Diseases **21**(1): 58.
URL: *https://doi.org/10.1186/s12879-020-05750-9*

Edmunds, W. J., O'callaghan, C. J. and Nokes, D. J. (1997). Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections, Proceedings of the Royal Society of London. Series B: Biological Sciences **264**(1384): 949–957.
URL: *https://royalsocietypublishing.org/doi/10.1098/rspb.1997.0131*

Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczkai, Z. and Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks, Nature **429**(6988): 180–184. Number: 6988 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/nature02541*

Flynn, E. F., Kuhn, E., Shaik, M., Tarr, E., Scattolini, N. and Ballantine, A. (2020). Drive-Through COVID-19 Testing During the 2020 Pandemic: A

Safe, Efficient, and Scalable Model for Pediatric Patients and Health Care Workers, Academic Pediatrics **20**(6): 753–755. Publisher: Elsevier.
URL: *https://www.academicpedsjnl.net/article/S1876-2859(20)30205-9/fulltext*

Goldstein, E., Dushoff, J., Ma, J., Plotkin, J. B., Earn, D. J. D. and Lipsitch, M. (2009). Reconstructing influenza incidence by deconvolution of daily mortality time series, Proceedings of the National Academy of Sciences **106**(51): 21825–21829. Company: National Academy of Sciences Distributor: National Academy of Sciences Institution: National Academy of Sciences Label: National Academy of Sciences Publisher: Proceedings of the National Academy of Sciences.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.0902958106*

Gozzi, N., Tizzoni, M., Chinazzi, M., Ferres, L., Vespignani, A. and Perra, N. (2021). Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile, Nature Communications **12**(1): 2429. Number: 1 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41467-021-22601-6*

Hackl, J. and Dubernet, T. (2019). Epidemic Spreading in Urban Areas Using Agent-Based Transportation Models, Future Internet **11**(4): 92. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/1999-5903/11/4/92*

Hengel, B., Causer, L., Matthews, S., Smith, K., Andrewartha, K., Badman, S., Spaeth, B., Tangey, A., Cunningham, P., Saha, A., Phillips, E., Ward, J., Watts, C., King, J., Applegate, T., Shephard, M. and Guy, R. (2021). A decentralised point-of-care testing model to address inequities in the COVID-19 response, The Lancet Infectious Diseases **21**(7): e183–e190.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1473309920308598*

Hilbe, J. M. (2011). Negative Binomial Regression, 2 edn, Cambridge University Press, Cambridge.
URL: *https://www.cambridge.org/core/books/negative-binomial-regression/12D6281A46B9A980DC6021080C9419E7*

Horl, S. and Balac, M. (2021). Synthetic population and travel demand for Paris and Ile-de-France based on open and publicly available data, Transportation Research Part C: Emerging Technologies **130**: 103291.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0968090X21003016*

Hou, X., Gao, S., Li, Q., Kang, Y., Chen, N., Chen, K., Rao, J., Ellenberg, J. S. and Patz, J. A. (2021). Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race, Proceedings of the National Academy of Sciences **118**(24): e2020524118. Publisher: Proceedings of the National Academy of Sciences.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.2020524118*

Kelman, A. (1985). Compartmental models and their application, International Journal of Bio-Medical Computing **16**(3-4): 294–295.
URL: *https://linkinghub.elsevier.com/retrieve/pii/0020710185900637*

Kermack, W. O., McKendrick, A. G. and Walker, G. T. (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. Publisher: Royal Society.
URL: *https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118*

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Hart, G., Rosenfeld, K., Selvaraj, P., Nunez, R. C., Hagedorn, B., George, L., Izzo, A., Palmer, A., Delport, D., Bennette, C., Wagner, B., Chang, S., Cohen, J. A., Panovska Griffiths, J., Jastrzebski, M., Oron, A. P., Wenger, E., Famulare, M. and Klein, D. J. (2020). Covasim: an agent-based model of COVID-19 dynamics and interventions. Pages: 2020.05.10.20097469.
URL: *https://www.medrxiv.org/content/10.1101/2020.05.10.20097469v1*

Kucharski, R., Cats, O. and Sienkiewicz, J. (2021). Modelling virus spreading in ride-pooling networks, Scientific Reports **11**(1): 7201. Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41598-021-86704-2*

Lau, H., Khosrawipour, T., Kocbach, P., Ichii, H., Bania, J. and Khosrawipour, V. (2021). Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters, Pulmonology **27**(2): 110–115.
URL: *https://www.sciencedirect.com/science/article/pii/S253104372030129X*

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression, Canadian Journal of Statistics **15**(3): 209–225. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/3314912.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.2307/3314912*

Mancastroppa, M., Burioni, R., Colizza, V. and Vezzani, A. (2020). Active and inactive quarantine in epidemic spreading on adaptive activity-driven networks, Physical Review E **102**(2): 020301.
URL: *https://link.aps.org/doi/10.1103/PhysRevE.102.020301*

Müller, S. A., Balmer, M., Charlton, W., Ewert, R., Neumann, A., Rakow, C., Schlenther, T. and Nagel, K. (2021). Predicting the effects of COVID-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data, PLOS ONE **16**(10): e0259037. Publisher: Public Library of Science.
URL: *https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0259037*

Pecoraro, V., Negro, A., Pirotti, T. and Trenti, T. (2022). Estimate false negative RT PCR rates for SARS CoV 2. A systematic review and meta analysis, European Journal of Clinical Investigation **52**(2): e13706.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8646643/*

Perez, L. and Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread, International Journal of Health Geographics **8**(1): 50.
URL: *https://doi.org/10.1186/1476-072X-8-50*

Pleninger, R., Streicher, S. and Sturm, J.-E. (2021). Do COVID-19 Containment Measures Work? Evidence from Switzerland, KOF Working Papers **494**. Accepted: 2021-07-07T11:36:01Z Publisher: KOF Swiss Economic Institute, ETH Zurich.
URL: *https://www.research-collection.ethz.ch/handle/20.500.11850/493408*

Reich, O., Shalev, G. and Kalvari, T. (2020). Modeling COVID-19 on a network: super-spreaders, testing and containment, medRxiv . Publisher: Cold Spring Harbor Laboratory Press _eprint: https://www.medrxiv.org/content/early/2020/05/05/2020.04.30.20081828.full.pdf.
URL: *https://www.medrxiv.org/content/early/2020/05/05/2020.04.30.20081828*

Riou, J., Panczak, R., Althaus, C. L., Junker, C., Perisa, D., Schneider, K., Criscuolo, N. G., Low, N. and Egger, M. (2021). Socioeconomic Position and the Cascade From SARS-CoV-2 Testing to COVID-19 Mortality: Population-Based Analysis of Swiss Surveillance Data, SSRN Electronic Journal .
URL: *https://www.ssrn.com/abstract=3845990*

Rudin, W. (1976). Principles of Mathematical Analysis, McGraw-Hill. Google-Books-ID: kwqzPAAACAAJ.

Scire, J., S. Huisan, J., Angst, D. and Li, J. (2023). COVID-19 Reproductive Number.
URL: *https://ibz-shiny.ethz.ch/covid-19-re-international/*

Singh, R., Hörcher, D. and Graham, D. J. (2023). An evaluation framework for operational interventions on urban mass public transport during a pandemic, Scientific Reports **13**(1): 5163. Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41598-023-31892-2*

Smieszek, T. (2009). A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread, Theoretical Biology and Medical Modelling **6**(1): 25.
URL: *https://tbiomed.biomedcentral.com/articles/10.1186/1742-4682-6-25*

Swiss National, S. T. (2021). COVID-19 Info Switzerland.
URL: *https://www.corona-data.ch*

Tuomisto, J. T., Yrjola, J., Kolehmainen, M., Bonsdorff, J., Pekkanen, J. and Tikkanen, T. (2020). An agent-based epidemic model REINA for COVID-19 to identify destructive policies, preprint, Infectious Diseases (except HIV/AIDS).
URL: *http://medrxiv.org/lookup/doi/10.1101/2020.04.09.20047498*

Wolfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brunink, S., Schneider, J., Ehmann, R., Zwirglmaier, K., Drosten, C. and Wendtner, C. (2020). Virological assessment of hospitalized patients with COVID-2019, Nature **581**(7809): 465–469. Number: 7809 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/s41586-020-2196-x*

Yasmin, F., Morency, C. and Roorda, M. (2015). Assessment of spatial transferability of an activity-based model, TASHA, Transportation Research Part A Policy and Practice **78**: 200–213.

A Tables

Name	Description
x_n^{age}	Age of the individual (in years).
x_n^{sex}	The gender of the individual, 1 = Female; 0 = Male; -1 = others.
$x_n^{\text{bike availability}}$	Binary variable, where: 1 = Yes; 0 = No.
$x_n^{\text{car availability}}$	Binary variable, where: 1 = Yes; 0 = No.
x_n^{employed}	Binary variable, where: 1 = Yes; 0 = No.
$x_n^{\text{has license}}$	Binary variable, where: 1 = Yes; 0 = No.
$x_n^{\text{home\_x}}$	Latitude geographic coordinate as decimal fraction of a degree
$x_n^{\text{home\_y}}$	Longitude geographic coordinate as decimal fraction of a degree
$x_n^{\text{household income}}$	Income of the individual (in CHF).
$x_n^{\text{is car passenger}}$	Binary variable, where: 1 = Yes; 0 = No.
$x_n^{\text{municipality type}}$	The municipality type, based on the Swiss Federal Statistical Office: 1 = Rural; 2 = Urban; 3 = Others.
$x_n^{\text{individual has GA}}$	Binary variable indicating whether the individual has a GA Travelcard, which offers unlimited travel across the Swiss public transportation network, where: 1 = Yes; 0 = No.
$x_n^{\text{civil status}}$	Binary variable, where: 1 = Married; 0 = Single, divorced, or widowed.

Table 3: The agents' social, economic, and demographic characteristics

B Recovery Time

The recovery time γ_n represents the duration an individual takes to transition from being infected to a recovered state. The first timestep t in which individual n changes the health state from $Z_{nt}^s = 1$ to $Z_{nt}^i = 1$ is t_n^i, defined as:

$$t_n^i = \sum_{t=1}^{T} (Z_{nt}^i - Z_{n(t-1)}^s)t,$$

and the timestep in which individual n transitions from $Z_{nt}^{i} = 1$ to $Z_{nt}^{r} = 1$, is denoted as t_n^r. Where γ_n is the time interval between t_n^r and t_n^i:

$$\gamma_n = t_n^r - t_n^i.$$

This time interval is assumed to follow a log-normal distribution as in Wolfel et al., 2020, a probabilistic model widely used to describe variables that are the product of many small independent factors. For our specific case, the mean and standard deviation of the underlying normal distribution are 8 days and 2 days, respectively.

$$\gamma_n \sim \text{lognormal}(384, 96), \quad \forall n, \tag{35}$$

where the mean and the standard deviation are expressed in timesteps t. To implement this distribution in our model, we generate recovery times for individual n by computing:

$$\gamma_n = e^{384+96\pi}, \tag{36}$$

where π is a random number drawn from a standard normal distribution and 384 and 96 are respectively the mean and standard deviation of the logarithm of recovery time. This process ensures that the recovery times vary in accordance with the log-normal distribution parameters. We compute γ_n for each individual, ensuring heterogeneity for the recovery time within the population.

C Derivation of Negative Binomial Model Equations

To derive the log-likelihood function for the Negative Binomial distribution as shown, let's start by recalling the probability mass function (PMF) of the Negative Binomial distribution and then proceed step by step to transform it into the given log-likelihood function format.

1. **PMF of the Negative Binomial Distribution** The Negative Binomial distribution models the number of failures before achieving a specified number of successes in a sequence of independent Bernoulli trials. If Y is a Negative Binomial random variable representing the number of failures before the r_1-th success, the PMF of Y is given by:

   $$P(Y = y) = \binom{y + r_1 - 1}{y} p^{r_1} (1 - p)^y$$

   where r_1 is the number of successes, y is the number of failures, p is the probability of success in each trial

2. **Logarithm of the PMF** To find the log-likelihood function, we take the logarithm of the PMF. Let r_1 represent the number of successes, $\widehat{y}^q_{gw}$ the observed number of failures, and N^q_{gw} the total number of trials minus successes (failures plus successes until the r_1 successes are achieved). The PMF then becomes:

$$\log P(\widehat{y}^q_{gw}) = \log \binom{\widehat{y}^q_{gw} + r_1 - 1}{\widehat{y}^q_{gw}} + r_1 \log p + \widehat{y}^q_{gw} \log(1-p)$$

3. **Simplifying the Binomial Coefficient** The binomial coefficient $\binom{\widehat{y}^q_{gw}+r_1-1}{\widehat{y}^q_{gw}}$ can be expressed using gamma functions:

$$\binom{\widehat{y}^q_{gw} + r_1 - 1}{\widehat{y}^q_{gw}} = \frac{\Gamma(\widehat{y}^q_{gw} + r_1)}{\Gamma(r_1)\Gamma(\widehat{y}^q_{gw} + 1)}$$

Then, the log of this expression is:

$$\log \binom{\widehat{y}^q_{gw} + r_1 - 1}{\widehat{y}^q_{gw}} = \log \Gamma(\widehat{y}^q_{gw} + r_1) - \log \Gamma(r_1) - \log \Gamma(\widehat{y}^q_{gw} + 1)$$

4. **Relating p to N^q_{gw} and r_1** Express p and $1-p$ in terms of r_1 and N^q_{gw} assuming $p = \frac{r_1}{r_1 + N^q_{gw}}$, then:

$$p = \frac{r_1}{r_1 + N^q_{gw}} \quad \text{and} \quad 1 - p = \frac{N^q_{gw}}{r_1 + N^q_{gw}}$$

Plug these into the log-likelihood equation:

$$\log P(\widehat{y}^q_{gw}) = \log \Gamma(\widehat{y}^q_{gw} + r_1) - \log \Gamma(r_1) - \log \Gamma(\widehat{y}^q_{gw} + 1) \tag{37}$$

$$+ r_1 \cdot \log\left(\frac{r_1}{r_1 + N^q_{gw}}\right) \quad + \widehat{y}^q_{gw} \cdot \log\left(\frac{N^q_{gw}}{r_1 + N^q_{gw}}\right) \tag{38}$$

5. **Summation for Log-Likelihood** The total log-likelihood $\mathcal{L}_1(\theta)$ for all observations across all groups g and weeks w (or whatever the indexing parameters represent) is the sum of the log probabilities for each observation:

$$\mathcal{L}_1(\theta) = \sum_{n \in g} \sum_{w \in W} \Big(\log \Gamma(\widehat{y}^q_{gw} + r_1) - \log \Gamma(r_1) - \log \Gamma(\widehat{y}^q_{gw} + 1) + r_1 \cdot \log\left(\frac{r_1}{r_1 + N^q_{gw}}\right) + \widehat{y}^q_{gw} \cdot \log\left(\frac{N^q_{gw}}{r_1 + N^q_{gw}}\right) \Big)$$

Notations

Variable	Concept	Type
$+$	positive	index
a	awareness	index
b	age	index
c	gender	index
d	municipality	index
e	parameter socioeconomic characteristics	index
f	facility	index
g	age group	index
h	parameter health characteristics	index
i	infected	index
j	parameter facility characteristics	index
ℓ	day	index
m, n	individual	index
p	time interval per day	index
q	choice of testing	index
r	recovered	index
s	susceptible	index
t	time interval throughout days	index
v	visit	index
w	week	index
E	number of socioeconomic characteristics	discrete
J	number of facility characteristics	discrete
H	number of health characteristics	discrete
L	number of days	discrete
N	number of individuals	discrete
P	number of timesteps within a day	discrete
T	number of timesteps throughout days	discrete
W	number of weeks	discrete
$\mathcal{F}$	set of locations	set
$E^{\star}_{nt}$	level of exposure for individual n at time t	latent
$Q^{\star}_{nt}$	propensity to test for individual n at time t	latent
Z^{+}_{nt}	individual n is positive at time t	binary
Z^{a}_{nt}	individual n is aware at time t	binary
Z^{i}_{nt}	individual n is infected at time t	binary
Z^{q}_{nt}	individual n has tested at time t	binary
Z^{r}_{nt}	individual n is recovered at time t	binary
Z^{s}_{nt}	individual n is susceptible at time t	binary

Z^{v}_{fnt}	restriction of a facility f for an individual n at time t	binary
y_{fnt}	presence of individual n in facility f at timestep t	variable
y^{j}_{nt}	characteristic j of the facility where n is at time t	variable
y^{v}_{nt}	proportion of infected individuals that individual n encounters at the facility where they are at time t	variable
$\widehat{y}^{e}_{n}$	socioeconomic characteristic e of an individual n, $\forall e \in E$	data
$\widehat{y}^{h}_{n}$	health characteristic h of an individual n, $\forall h \in H$	data
$\widehat{y}^{j}_{f}$	characteristic j of a facility f, $\forall j \in j$	data
$\widehat{y}^{0}_{fnt}$	presence of individual n in facility f at period p in day 0	data