



Assessing population-sampling strategies for reducing the COVID-19 incidence[☆]

Miguel Guzmán-Merino^a, Christian Durán^a, Maria-Cristina Marinescu^b,
Concepción Delgado-Sanz^{c,d}, Diana Gomez-Barroso^{c,d}, Jesus Carretero^a, David E. Singh^{a,*}

^a Universidad Carlos III de Madrid, Leganes, Spain

^b Barcelona Supercomputing Center, Barcelona, Spain

^c CIBER en Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

^d National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain

ARTICLE INFO

Keywords:

SARS-CoV-2(COVID-19)
Agent-based simulation
Sampling strategies
Social model
Contact matrices

ABSTRACT

As long as critical levels of vaccination have not been reached to ensure herd immunity, and new SARS-CoV-2 strains are developing, the only realistic way to reduce the infection speed in a population is to track the infected individuals before they pass on the virus. Testing the population via sampling has shown good results in slowing the epidemic spread. Sampling can be implemented at different times during the epidemic and may be done either per individual or for combined groups of people at a time. The work we present here makes two main contributions. We first extend and refine our scalable agent-based COVID-19 simulator to incorporate an improved socio-demographic model which considers professions, as well as a more realistic population mixing model based on contact matrices per country. These extensions are necessary to develop and test various sampling strategies in a scenario including the 62 largest cities in Spain; this is our second contribution. As part of the evaluation, we also analyze the impact of different parameters, such as testing frequency, quarantine time, percentage of quarantine breakers, or group testing, on sampling efficacy. Our results show that the most effective strategies are pooling, rapid antigen test campaigns, and requiring negative testing for access to public areas. The effectiveness of all these strategies can be greatly increased by reducing the number of contacts for infected individual.

1. Introduction

While vaccination has started several months ago in a significant part of the world, it is not progressing at the fast pace that was expected. In parts of the world outside North America and Europe, vaccination rates are still abysmally low [39]. New variants are adding complexity and insecurity about the effectiveness of the different vaccines available. In the absence of a vaccine, emphasis should be placed on policies that protect the most-vulnerable population while herd immunity is hoped to be achieved in the less vulnerable people [16]. Mitigation strategies that can slow the propagation of the virus in the susceptible population are the target of continuous modification. An effective way to this is to test

part of the population that is not (yet) symptomatic, during a process called sampling. Reducing virus transmission is directly proportional to the percentage of the population that is tested. In terms stated by the ECDC, “*Ideally, all people with COVID-19 symptoms should be tested as soon as possible after symptom onset*” [7]. This practice has been implemented in hospitals and first care centers, so symptomatic individuals are tested immediately when they decide to get medical attention. In case of positive test results, infected individuals are quarantined to limit the virus spread. This measure has a high impact limiting the virus propagation because it helps distinguish COVID-19 infections from common sickness with similar symptoms such as fever or cough produced by influenza. However, not all individuals who develop symptoms seek

[☆] This work has been supported by the Carlos III Institute of Health under the project grant 2020/00183/001, the project grant BCV-2021-1-0011, of the Spanish Supercomputing Network (RES) and the European Union’s Horizon 2020 JTI-EuroHPC research and innovation program under grant agreement No 956748. The role of all study sponsors was limited to financial support and did not imply participation of any kind in the study and collection, analysis, and interpretation of data, nor in the writing of the manuscript.

* Corresponding author. Departamento de Informática, Universidad Carlos III de Madrid, Avda. de La Universidad 30, 28911, Leganés, Spain.

E-mail address: dexposit@inf.uc3m.es (D.E. Singh).

<https://doi.org/10.1016/j.combiomed.2021.104938>

Received 3 July 2021; Received in revised form 8 October 2021; Accepted 9 October 2021

Available online 12 October 2021

0010-4825/© 2021 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

medical attention. Additionally, a seemingly large percentage of infected people are asymptomatic, and they add to the ranks of pre-symptomatic individuals who are, unknowingly, transmission vectors. Identifying these individuals before they further infect other susceptible people can have a very large positive impact on the development of the epidemics [27]. A more precise understanding of the effects of different possible sampling strategies would contribute to design policies that can slow down the propagation and reduce the number of affected individuals, while trying to optimize the resources of the health system.

This work builds on EpiGraph [23], a fully distributed agent-based simulator for influenza and COVID-19 diseases. Agent-based approaches have the potential to model each individual's characteristics and interaction patterns, which can result in much more realistic simulations compared to other approaches [5,15]. One of the distinguishing features of EpiGraph is that it relies on realistic data for both individuals and their interaction patterns, which we extract by scaling from existing social networks and contact matrices. Interconnections are time-dependent in order to realistically capture the temporal nature of interactions and the specifics of each profession or social activity. This paper makes the following main contributions:

- Design and implementation of a more refined social interaction model which reflects profession-dependent connections to increase the realism of population mixing. We also use contact matrices extracted from public surveys to reflect age-dependent interactions of each individual. Lastly, we make the code and related data available as open source in [42].
- Development of different COVID-19 sampling (i.e. testing) strategies, which we compare with a basic scenario in which a baseline of

2. Materials and methods

Our simulator's structure is detailed in Algorithm 1. Each city simulates a population mix based on the Spanish census data [25], and defines social connections between its individuals. For each simulation step (equal to 10 minutes - line 1) and every city in the simulated territory (line 2), the algorithm updates the health status of each infected individual as indicated by EpiGraph's Epidemic Model (line 5). The next step *ComputeSpreadGraph* (line 6) computes how the infectious agent spreads through the Social Model. We call non-pharmacological interventions applied at individual-level (*Individual_NPIs*, line 8) a non-pharmacological action taken by an individual to mitigate the virus propagation like using face masks. An example of this kind of intervention is when an individual starts using a surgical face masks at work, but not during family time.

In line 10, dynamic transmissions are evaluated - as opposed to the permanent graph-based transmissions (line 6). Dynamic connections are generated for individuals belonging to certain collectives that have short-duration interactions with different people. Examples of these collectives are health professionals and catering workers that are in contact with different patients or customers. Note that momentary connections change over the time, i.e. some of the individuals involved may be different every time.

Algorithm 1. EpiGraph transmission algorithm. Variable *simulation_time* represents the simulation duration, *simulated_territory* is the simulated area including several cities, each one of them, denoted as *city* with a social interaction model for the population. *Individual* contains characteristics and health status of each individual belonging to each city.

Algorithm 1 EpiGraph transmission algorithm. Variable *simulation_time* represents the simulation duration, *simulated_territory* is the simulated area including several cities, each one of them, denoted as *city* with a social interaction model for the population. *Individual* contains characteristics and health status of each individual belonging to each city.

```

1: for timestep = 1 → simulation_time do
2:   for city ∈ simulated_territory do
3:     for individual ∈ city do
4:       if individual is infectious then
5:         UpdateStatus(individual)
6:         ComputeSpreadGraph(individual, city)
7:       end if
8:       Individual_NPIs(individual)
9:     end for
10:    ComputeSpreadDynamic(city)
11:    Collective_NPIs(city)
12:    Transportation(city, simulated_territory)
13:    Vaccination(city)
14:  end for
15: end for

```

sampling and mitigation strategies are applied. The scenario includes the 62 largest cities of Spain, and testing strategies are applied over the whole population. In total, we evaluate 12 different sampling strategies including several variants changing configuration parameters, such as the testing frequency, quarantine time, and the percentage of quarantine breakers.

The rest of the paper is organized as follows. After the simulation framework introduction, in Section 2 we describe the new features developed in EpiGraph. Section 3 shows how EpiGraph was calibrated using the available data of the Spanish Third wave, at early of 2021. After that, the results of the impact of different sampling policies are shown. Section 4 provides a discussions of the findings as well as the limitations of the work. Finally, Section 5 presents the main conclusions of our work.

We call non-pharmacological interventions applied at collective-level (*Collective_NPIs*, line 11) an intervention —such as school closing or a total or partial lockdown— that is imposed (or lifted) by the health authorities at a certain time during the simulation and that involves a specific collective. Line 12 of Algorithm 1 simulates the exchange of individuals between cities by calling the Transportation Model, which is based on the gravity model proposed by Viboud et al. [37]. This model computes the number of travelers depending on the size of the origin and destination cities, as well as the geographical distance between them. The geographical information that EpiGraph takes into account includes latitude, longitude, and distance between urban regions, and was extracted from the Google Maps web service using the Google Distance Matrix API [11]. The Transportation Model captures regular medium and long-distance commute for work and study, together with occasional vacations. Finally, in line 13 the Vaccination Model captures both the COVID-19 vaccine availability and characteristics as well as the

vaccination policies that determine which individuals are vaccinated at any time and with what vaccine type.

2.1. Social interaction model

EpiGraph is an agent-based model that captures individuals with their attributes, groups of individuals, and interactions patterns. We use demographic information to reproduce the social habits of four different main group types: students, workers, stay-at-home people, and elders. Individuals form part of groups, each of a single group type. A group can represent a certain number of individuals that interact during work hours - for instance, groups are the students belonging to the same classroom, workers of the same company, or stay-at-home people or elderly people that perform group activities.

The way the individuals establish social contacts¹ is time-dependent and reflects the temporal nature of the different types of interactions that each individual has throughout the day. For each one of the group types, we consider three different temporal distributions of the individual's activities, those related to weekdays, Saturdays, and holidays (including Sundays). These patterns are specific to the place being modelled to correctly capture typical work hours, school time, family time, and leisure.

EpiGraph creates different graphs to model connections within each work group, school group, stay-at-home (informal meetup) group, and elderly (informal meetups) group. Rather than assuming a distribution or generating synthetic interaction graphs, we use real information from social networks to model the social interaction patterns. We have used the Enron Email Corpus (70,578 nodes and 312,620 edges) for generating the work, elderly and informal-meetup groups while the Facebook (250,000 nodes and 3,239,137 edges) network was used to generate the school groups.

The social interaction model includes two more types of social contacts for leisure and family activities. Leisure contacts are modelled via inter-group contacts. These contacts are between individuals belonging to different groups (for instance, work and school groups) and mostly occur after the main daily activity and before family time, as well as during the weekends. These contacts represent interactions with friends as well as casual contacts with unknown people. The third class of contacts are family contacts, interactions with family members who may or may not be part of the same group. The family connections graph is completely connected. However, connections are time-dependent which means that some of them (related to individuals that are not at home at a given time) are only active during a certain time interval. For a more detailed description of these features, please see Ref. [33].

In this work we have enhanced the social model with information of contact matrices [26]. These matrices, extracted from public surveys, contain statistical information of the average number of contacts between individuals of certain age ranges. There are four matrices corresponding to school, work, household and community contacts. Fig. 1 (left) shows the aggregated contact matrix values that corresponds to Spain. For example, based on this figure, individuals that are between 10 and 20 years old have an approximate average of 2 and 8 contacts per day with other individuals in the range of ages of 0–10 years, and 10–20 years, respectively.

We have developed a new graph scaling algorithm based on the random walk algorithm [20] that starts from as many nodes as the desired group size, chosen randomly, and randomly explores different paths in the graph. This means that the connection pattern of each group is unique, while maintaining certain graph-related properties such as the variable distribution of the number of contacts per individual [22]. In

the newly developed algorithm, the resulting graphs are generated using a similar approach, but with certain average connectivity $\langle k \rangle$ that is specified by the contact matrices. The school contact matrix was used to generate school groups with the same number of contacts per age as this contact matrix. In a similar way, the work sub-matrix was used to define the connectivity for the work, stay-at-home and elders (informal meetup) groups. Finally, the community sub-matrix was used to generate the connectivity related to the leisure contacts reflecting the contacts that are neither related to school nor work connections, for instance, child-worker, child-elderly or worker-elderly connections. Fig. 1(right) shows the resulting contact matrix that has been obtained from the contact model created by the simulator. Note that the real and simulator-generated matrices have a similar structure - although not identical. It turns out that we are able to generate a very good approximation of reality, given that we start from data sources that are very different from real interaction networks - something we cannot possibly hope to have. Our sources are virtual networks (Enron Email Corpus and Facebook graphs), which we scale to obtain matrices that preserve the patterns of the source networks while adapting them to the number of age-based contacts from the contact matrices.

In order to increase the realism of population mixing, the work group has been broken down in different professions. In the same way, we distinguish different sub-collectives for the group of elderly people.² Another new feature is ad-hoc connections for certain profession and collectives that have specific connection patterns. These connections may be *static* or *dynamic*. Static connections are generated during the social model creation and do not change during the simulation. They are created when the connection graph is generated, before the simulator execution. Then, during the simulation these connections are evaluated by `ComputeSpreadGraph()` in Algorithm 1 for a certain time slot. Dynamic connections are not generated offline but during the simulation, and they change every time that they are evaluated. They reflect changing communication patterns that may be in contact with certain individual at a given time, and other individuals at another time. Dynamic connections are evaluated by the `ComputeSpreadDynamic()` function in Algorithm 1. The following list describes the different ad-hoc connections implemented in EpiGraph:

- Ad-hoc school static connections. Each educator is in contact with all the students of a certain class during work hours. These connections are not active during the whole time slot, but are rather sporadic during this period, in order to reflect the fact that educator and students share the same space but are not permanently in contact. Note that the educator also remains in contact with the work, leisure and family groups (each one of them at their respective time slot). In this case, the work group represents the contacts with the other educators belonging to the same center.
- Ad-hoc elderly caregiver static connections. Elderly caregivers are in contact with a certain group of elderly people at a nursing home during work hours in the same way that educators are in contact with students. These contacts are also sporadic, i.e., are not active during the complete work slot but are the same during all the simulation. This means that a certain caregiver has the same related group of elderly people that is connected with.
- Ad-hoc health-care dynamic connections. Each worker belonging to the health sector is in contact with 30 patients per day.³ For non front-line health workers, the patients are chosen at random among the existing population, so the risk of meeting a COVID-19 infected individual is the same as for any other profession. In contrast, front-

¹ We define a social contact (also called contact or interaction) between two individuals based on co-location in time at a distance that is small enough to make disease transmission possible.

² In the supplementary material you can find more details about the collectives belonging to these groups.

³ This value corresponds to the daily average number of patients per doctor in Spain.

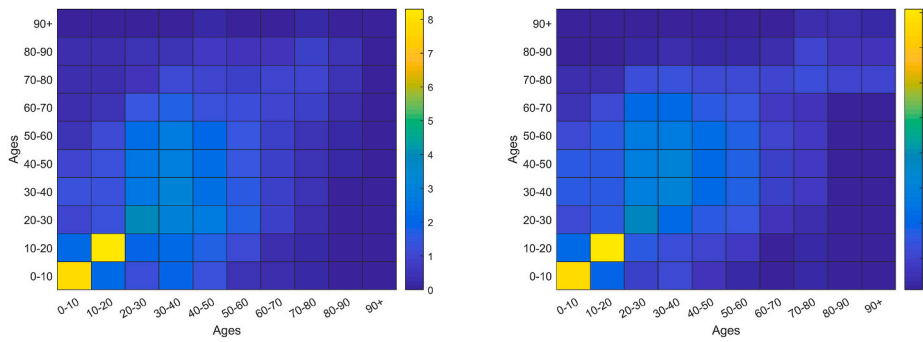


Fig. 1. (Left) Global input contact matrix used in the social model. Each column represents the average number of contacts of the individuals belonging to the age interval shown at the top of the column. These contacts are broken down by age intervals in the different rows. (Right) Global contact matrix related to EpiGraph’s social model used in the experiments. Each column represents the average number of contacts of the individuals belonging to the age interval shown at the top of the column. These contacts are broken down by age intervals in the different rows.

line health workers have also dynamic connections with 3.3 times more risk of meeting a COVID-19 infected patient [31].

- Ad-hoc catering dynamic connections. Each worker belonging to this sector is in contact during work hours with 10 other individuals per hour. The contacts (that represent the customers) are chosen at random among the population, excluding the ones that are in bed, quarantined or hospitalized. We consider three levels of catering contacts: pre-pandemic, pandemic with a more reduced number of contacts per hour, and lockdown with catering services closed.
- Ad-hoc public security dynamic connections. They represent the contacts of police and other security forces, in which each officer is in contact during work hours with other individuals per hour, that represent the general public. The contacts are chosen at random among the population, excluding the ones that are in bed, quarantined or hospitalized.
- Ad-hoc occasional meeting dynamic connections. They represent meetings between different groups related to social events. We consider three different classes of occasional meetings: (1) occasional work meetings represent meeting between people belonging to different work groups. Here, once per day we select a varying number of work groups (between 2 and 4) at random and we connect them during a 4-h time period. Note that only a small fraction of the existing work groups are connected by this procedure; the remaining ones do not participate in occasional meetings; (2) occasional school meetings are similar to work meetings, where student groups are chosen instead of work groups. They represent occasional social gathering between students; (3) occasional leisure meetings represent groups of friends that gather. Like for the rest of the occasional meetings, these connections are created once a day from only a small fraction of the existing contacts.

Note that all these ad-hoc contacts are complementary to the existing individual graph-based contacts. This allows an individual to have two different types of interactions during work hours: the ones related to the graph - that connect him with other work colleagues (for instance, educators belonging to the same school, health professional belonging to the same hospital or catering employees working at the same restaurant) but also the ad-hoc ones that connects the individual with other individuals in other group types, for instance, educators with students, health professional with patients and catering workers with customers.

2.2. COVID-19 model

The epidemic model is a compartmental stochastic SEIR model extended with latent, asymptomatic, dead, hospitalized and vaccinated states. Rather than the more common analytic models based on differential equations, EpiGraph probabilistically decides the duration of the different compartments and the transitions between them. In addition, the basic reproduction numbers R_{0s} are different for each compartment. Fig. 2 is an extended version of the figure presented in Ref. [33]. It shows the different infection phases, which are described below:

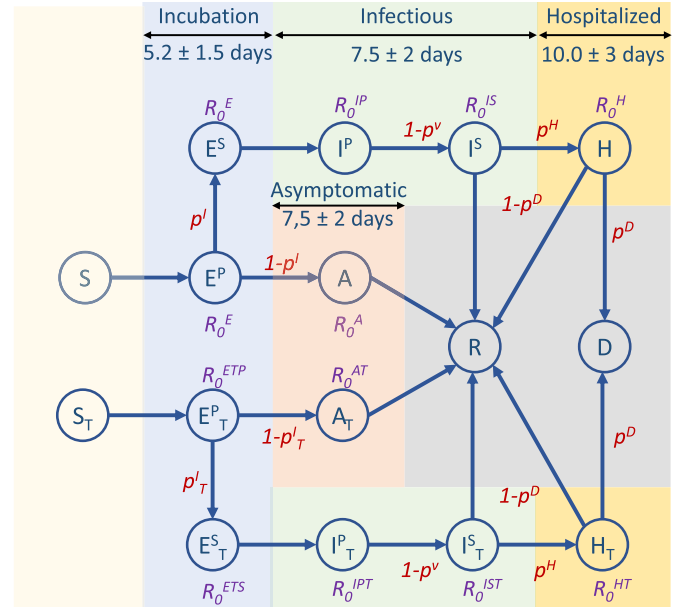


Fig. 2. Compartmental model used by EpiGraph. It consists of the following states: susceptible (S), primary exposed (E^P), secondary exposed (E^S), asymptomatic (A), primary infected (I^P), hospitalized (H), recovered (R) and dead (D) individual. Each state shows the basic reproduction number of the state (non-existing R_0 s means that are not applicable). The edges show the transition probabilities (which are normalized) between the compartments. Duration of the main infection stages consists of an incubation that includes E^P and E^S ; infectious includes I^P and I^S ; hospitalized is represented as H ; and asymptomatic is A . Note that the asymptomatic stage starts after the primary exposed stage (E^P), which in this approximation lasts only one day. States S_T , E_T^P , E_T^S , A_T , I_T^P , I_T^S and H_T are related to *treated individuals* i.e. individual that have been vaccinated.

- **Incubation stage.** At the beginning of this stage individuals are infected but they have no symptoms and are not yet able to transmit the virus. This stage is represented as primary exposed E^P . From this stage the infection can enter one of two phases, based on a probability P^{E^S} : a secondary exposed stage E^S where slight symptoms appear and the individual becomes infectious with a certain $R_0^{E^S}$, or an asymptomatic stage (described below).
- In the **asymptomatic stage** (compartment A), infected individuals do not notice symptoms but are able to transmit the disease with a certain R_0^A reproduction number. After a certain time, they pass to the recovered compartment in which the subject acquires viral immunity.
- In the first **symptomatic stage** - called primary infection state I^P - symptoms appear. Individuals will then transition to phase I^S , where

symptoms persist. I^P , I^S and I^V have associated basic reproduction numbers of R_0^P , R_0^S and R_0^V .

- A certain fraction of the individuals are hospitalized (**hospitalized stage**). The probability of entering this stage is given by the parameter $P^H(\text{age})$, which increases with age. From this state, an individual may transition to either the recovered or the dead stage. During hospitalization, we use R_0^H for modelling the transmission in hospitals.
- The individuals that reach the **dead stage** are removed from the simulation. The transition probability, $P^D(\text{age})$, is also age-dependent and is applied over the portion of hospitalized individuals.
- The **treated stages** shown in the lower part of the figure represent the infection stages for vaccinated individuals. A non-infected vaccinated individual has associated the treated Susceptible state (S_T). In case of being infected, the first related state is the treated Exposed primary (E_T^P). Then, in case of a vaccination failure, the transition will include the states E_T^S , I_T^P , I_T^S and H_T , in a similar way that non-vaccinated individuals. Note that the probability of vaccination failure is p_T^I , which depends on the type of vaccine that has been used, the virus variant that has infected the individual and in some cases other factors, like the age of the individual. On the contrary, when there is no vaccination failure, then the individual transits to the Asymptomatic treated state A_T .

The time spent in a given state is generated following a normal distribution to simulate the time ranges specific to each stage of the infection and the fact that each individual may go through phases of different lengths. We also consider that a percentage of the sick individuals stay in bed, thus reducing the number of people that they interact with. We have used the same COVID-19 parameters (R_0 values, transition probabilities, etc.) as the ones previously depicted in Refs. [33].

2.3. Sampling strategies

In this work we evaluate the effect of new sampling strategies with the aim of identifying the strategy that achieves the larger reduction in COVID-19 cases. Some strategies are specially focused on certain collectives. We choose the health, social-health, defense and catering collectives as target for these strategies given that they have a higher incidence of COVID-19 cases. All the sampling strategies are parameterizable by the number of daily tests, the minimum testing frequency (i.e. the minimum time between two tests carried out for the same individual), the quarantine time, and the percentage of quarantine breakers (i.e. the fraction of people who do not comply with social distancing during quarantine time). Works like [8,12] points out the relevance of the testing frequency. By properly setting this parameter, health authorities avoid testing the same individual within a short period of time (wasting tests that could be used with other individuals) or extending too much the time between successive tests, which would increase the risk of not detecting positive cases. Other determinant factor in the virus spread is quarantine time. Individuals that are COVID-19 positive in Spain are recommended to quarantine until at least three days after symptoms disappear and during at least a 10 days [24]. On the other hand, in case of contact with a positive, Spanish health authorities recommend a 10-day self-isolation period, starting since the last contact with the confirmed case [24]. The WHO (World health organization) recommends that all contacts of individuals with a confirmed or probable COVID-19 be quarantined in a designated facility or at home for 14 days from their last exposure [40]. In our experiments, we used a default value of quarantine of 10 days and a sensitivity value for the PCR tests of 100%. The different sampling strategies we consider in this work are enumerated below.

- **Strategy 1**, baseline strategy. This strategy reproduces the testing strategy applied in Spain: a given number of tests (0.25% of the simulated population) are performed daily and a percentage of these (around 9%) are positive. We model this strategy by a combination of random testing over the simulated population and a selective identification of positive cases. The goal here is to achieve positive testing ratios similar to the existing ones in Spain at this time (around 9%). To achieve this, tests in the simulation are of two types: they are either randomly applied to the population or they are carried out selectively, but exclusively over the existing positive cases, in random fashion. The size of each of the two sets is empirically adjusted to achieve a percentage of positive tests as close to the 9% real ratio. Some of the remaining sampling strategies - described below - are applied in combination with this baseline. Note that when combining strategies, if a certain individual is tested in the context of the baseline approach, he/she cannot be tested again during the same period by any of the other testing strategies.
- **Strategy 2**, random testing without selective identification. In this strategy extra tests are performed in combination with the baseline strategy. More specifically, an additional 0.3% of the population is randomly tested daily. In this strategy a selective identification of positive cases is not performed. This means that, unlike in the baseline strategy, we do not impose any restriction on the fraction of the tests that we assume to return positive. Note that this represents the simplest testing strategy, in which the available extra tests are used with all individuals chosen at random.
- **Strategy 3**, extra tests are targeted to the health, social-health home, and defense workers in combination with the baseline strategy. These collectives represent 7.5% of the simulated population and they represent public-sector employees that have a higher risk of being in contact with COVID-19 infected individuals. With this strategy 0.3% of the target group is tested daily.
- **Strategy 4**, extra tests target catering workers in combination with the baseline strategy. The catering group is about 4.3% of the simulated population for Spain. It represents private-sector employees that are also more likely to be infected. As in the previous strategy, base tests are performed over the entire population and, at the same time, extra test focus on this target group with a high incidence of cases, where 0.3% of the workers is tested daily.
- **Strategy 5**. In the previous strategies only positive individuals are quarantined. This strategy, on the other hand, introduces a variation in the baseline strategy in which family contacts of a positive individual are also quarantined during a 10-day period. No extra tests are performed in this strategy; this maintains 0.25% of the simulated population being randomly tested daily, with a positive rate of 9%.
- **Strategy 6** that represents a variation of Strategy 5. All contacts of positive individuals are quarantined instead of only family contacts; this includes household, leisure, and work contacts.
- **Strategy 7** implements the pooling method [34,41] in combination with the baseline strategy. This method consists in testing various samples with a single test. A positive result means that there is at least one infected individual among the set (called pooled group). If that is the case every member of the pooled group is individually tested in order to identify the positives. A negative result means that there are no infected individuals in the pooled group. This group is formed by individuals that are normally in close contact, such as family or colleagues. In this strategy we focus on work contacts and set a maximum pooling size group of 20 individuals with about 1.7% of the workers that are sampled daily. Quarantined individuals are not included in the pooling process. The number of extra daily test is the same as in Strategy 2.
- **Strategy 8** applies the pooling method (Strategy 7) only for catering workers in combination with the baseline strategy. About 1.57% of catering workers are sampled daily by the pooling method; this represents 0.23% of the workers in this profession.

- **Strategy 9** represents a rapid antigen tests that is carried out in combination with the baseline strategy. This additional campaign takes place during one single week in which the additional daily number of tests is 5 times larger than the ones used per week in the Strategy 2. This increases the daily tested workers to 0.86% which are selected at random. This strategy is particularly useful when the percentage of newly infected individuals is raising sharply and allows the quick identification of a larger number of pre-symptomatic and asymptomatic people, which are subsequently quarantined. After the test campaign period (one week), the number of tests is reduced to the default value of 0.17% of workers that are daily tested.
- **Strategy 10** is a variation of the previous one in which the test campaign (of strategy 9) is selectively used on catering workers in combination with the baseline strategy. During an entire week, the additional number of daily tests is 5 times larger than the usual number. This accounts for 1.4% of the catering workers being tested daily, while baseline tests are performed randomly over the population. In a similar way than the previous strategies, an individual cannot be tested twice during a period shorter than the test window time, no matter if the test campaign is active or not.
- **Strategy 11**, avoiding leisure contacts for non-tested individuals in combination with the baseline strategy. In this strategy a negative PRC is required to be allowed access to public places like restaurants, or leisure places [6]. For instance, an individual is not allowed to have social life without a previous negative PRC. This restriction is applied to all leisure contacts, including the ones related to children.
- **Strategy 12** describes the ideal situation in which the baseline scenario targets only infected individuals. There is still a limited number of daily tests (0.3% of the population) but all of them are positive (in case of existing enough infected individuals).

We also tested other sampling strategies, that we discarded mostly because they showed very low impact on the propagation of the infection or, as in the case of individuals with a high number of contacts, because knowing this information is unrealistic.

3. Results

3.1. Model validation

Our experiments consist of a simulated population of 19,574,086 individuals (which lives in the 63 most populated Spanish cities), whose geographical locations are shown in Fig. 3. Sampling strategy 1 is used in this scenario in order to reproduce the existing testing policy in Spain during the simulation period. To set the number of daily tests for the sampling strategies that are carried out in Spain, we assume that 0.25%

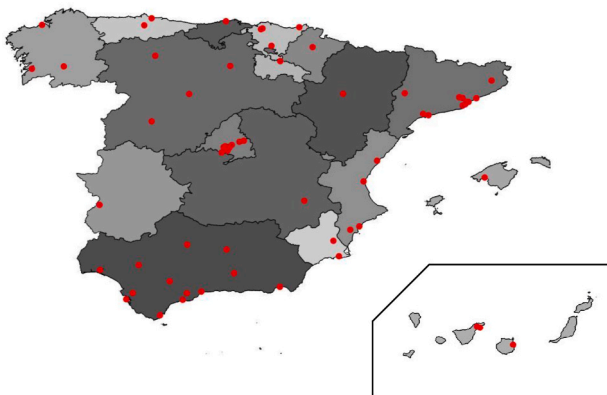


Fig. 3. The 63 most populated cities in Spain are represented in the image. Note that each Spanish province has at least one representative city.

of the Spanish population has been tested daily since March 2020 - on average, with only 9% of these tests returning positive. These percentages are similar to the existing values for Spain. Note that in this baseline scenario the detected individuals that are infected with SARS-CoV-2 are quarantined. In our experiments, we set the testing frequency parameter to 15 days. An individual that has a positive test is originally quarantined for 10 days from work and leisure contacts, although the subject does maintain household contacts. We set the percentage of quarantine breakers to 0 by default, although for some representative strategies we will study the effect of different values for quarantine time and quarantine breaker percentages. In our experiments, all individuals use face masks in work, school and leisure times, but not when they are at home. In addition, we have applied the same social distancing restrictions as the existing in Spain for the simulation period.

EpiGraph was calibrated to set up the conditions of the epidemic in Spain during the third wave, at the beginning of 2021. Simulation starts on December 27th, 2020 with a given percentage of infected individuals, and runs throughout to April 11th, 2021. Concretely, we use the percentage of officially reported infections on December 27th for each of the Spanish provinces. These values, ranging from 0.123% in Canarias province to 0.613% in Extremadura province, only represent a fraction of the real cases, which is what EpiGraph needs as input. Due to this fact, we corrected the starting values by a scale factor that reflects the non-reported cases. According to the Spanish health authorities, only around 70% of the infected individuals were detected, which results in a scale factor of 1.42. Other works [9], estimate a larger scale factor. Fig. 4 shows the aggregated number of infected individuals. Real and simulated data are shown in red and blue, respectively. The real data was obtained from Ref. [17]. Note that in our experiments, each city includes demographic information which corresponds to the region it is located in, and contains the population pyramid, job sector distribution, number of family members per household, etc. The method we use for the calibration manually tunes the scale factor which adjusts the initial number of infected individuals for each region. The rest of the parameters related to the epidemiological model and the NPIs are the same for all the considered cities and were not involved in the calibration process.

For the beginning of the simulation, on December 27th, 2021, we use the prevalence values of 11% for workers, 9.1% for students, 8.6% for unemployed and 1.01% for elderly people. These values, collected from Ref. [29], correspond to the percentage of population that was already

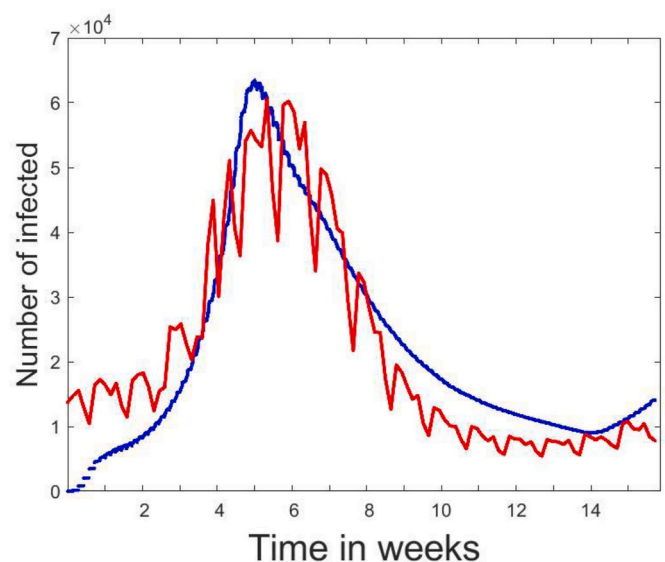


Fig. 4. Model validation: daily real (in red color) and simulated (in blue color) data related to the number of infections of the COVID-19 spread in Madrid metropolitan area for the Spanish Third Wave. Simulation starts on December 27th of 2020.

infected with COVID-19 before the start of the third wave. In this work, we assume that these individuals have become immune to COVID-19 and they will not be able to be reinfected during the entire simulation time. In order to provide more precise and realistic results, EpiGraph uses the vaccination model that reproduces the vaccination campaign started in Spain on December 27th and takes into account the different vaccines, their efficacy to each COVID-19 variant, and the number of doses employed with each collective and profession during each simulated day.

EpiGraph uses stochastic processes to perform the simulations, which may result in differences between results every time the simulation runs. In order to quantify the deviation in the results, we have repeated the same simulation 20 times obtaining a median percentage of 5.3% of the population that was infected at the end of the simulation, with a variance (of 0.477% for Spain). Note that this value is similar to the reported number of 6.7%. Fig. 5 shows the breaks down by groups for the percentage of infections at the end of the simulation. Note that some collectives with a high degree of contacts (like catering workers and defense professionals) have a higher proportion of infection cases. Other collectives that are at high risk of being in contact with infected individuals (like front-line health professionals) have a smaller proportion of infections because they have been prioritised in the vaccination program and they have been vaccinated at early stages of the simulation. The rest of this section summarizes the main features of the most important modules in EpiGraph: the social interaction and virus evolution models.

3.2. Sampling policy evaluation

Fig. 6 shows the results for all the considered strategies, in the context of the third COVID-19 wave in Spain between January and March 2021. Each value is the average of 20 simulations. This figure shows with blue bars the percentage of infected population. The related values are shown on the left y axis. Samples and PCR tests are respectively represented by the orange (solid) and green (dotted) lines, with values shown on the right y axis. These values represent the total number of tests carried out during the entire simulation. Note that both lines take the same value in all cases except in strategies 7 and 8, in which pooling is applied and several samples are tested with a single PCR test. Fig. 7 also shows the final percentage of infected individuals, but here this information is broken down by groups.

The baseline scenario (Strategy 1), corresponds to the scenario shown in Fig. 4. We can observe that as expected, Strategy 12 is the one

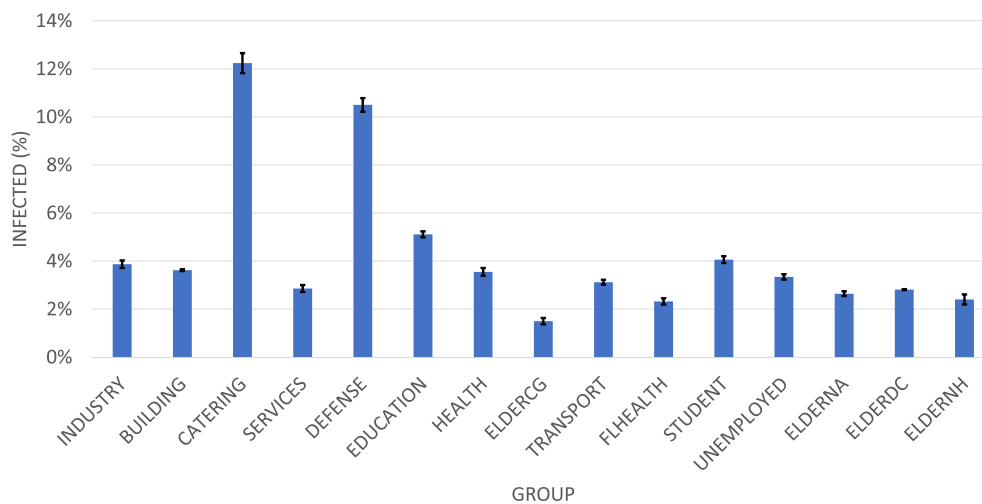


Fig. 5. Percentage of infected population broken down by groups. The acronyms stand for: ELDERCG: care giver for elderly people, HEALTH: non-front-line health professionals, FLHEALTH: front-line health professionals, ELDERNA - ELDERDC - ELDERNH: the elderly people that live by themselves, are attended in daily centers, or live in nursing homes.

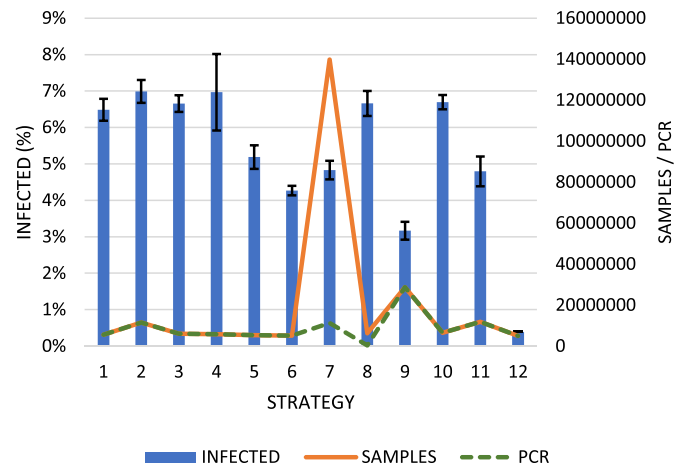


Fig. 6. Simulation outcome for each of the sampling strategy. The percentage of infected population is represented by blue bars and the related values are shown on the left y axis. Samples and PCR tests are represented by the orange and green lines, with values shown on the right y axis.

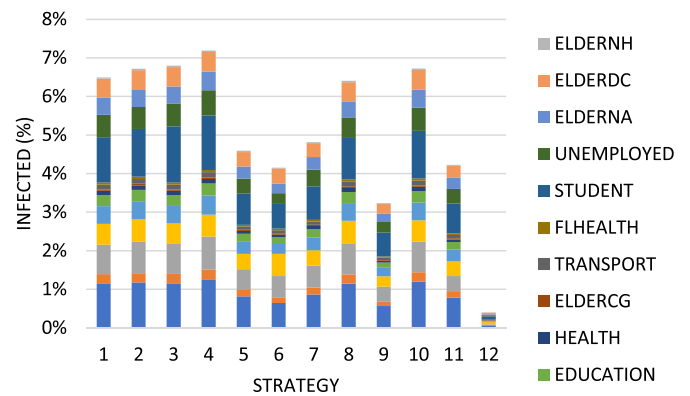


Fig. 7. The percentage of infected population related to each sampling strategy broken down by groups. ELDERCG stands for care giver for elderly people, HEALTH represents the non-front-line health professionals, FLHEALTH stands for the front-line health professionals, and ELDERNA, ELDERDC and ELDERNH represent the elderly people that are non-attended, attend daily centers and live in nursing homes, respectively.

that achieves a larger reduction in total number of contagious because it is ideally targeted to infected individuals. When considering the remaining strategies, the best results are achieved with strategies 6 and 9. Self-isolation of infected individual's contacts is the best way to reduce the virus spread. When considering Strategy 6, self-isolation does not require extra testing, which makes it the most economic strategy. However, tracing infected individuals contacts is not easy to implement, and a perfect self-isolation is usually not a realistic option. On the other hand, Strategy 9 needs more tests than the rest of the strategies, which makes it the most expensive in terms of resources. The test campaign implemented here takes place when the number of infected people starting to increase, which in our simulation occurs between week 2 and 4 (see Fig. 4). Strategy 5 also shows a significant reduction in the percentage of infected individuals while only family contacts of infected individuals are required to isolate. Note that according to our study, specially targeted strategies (3 and 4) do not achieve a significant reduction in the overall number of contagious although they slightly reduce the COVID-19 incidence in the collectives. For Strategy 3, the contagious in health, social-health home, and defense professional are reduced in 6.75%, 4.83% and 6.86%, respectively. In Strategy 4 the number of infections in catering workers does not significantly change. The reason is that the main reason of contagious is between the catering workers and the clients, and there is no testing among the client collective. Similar results are obtained with Strategy 10. In this figure we can observe that the pooling method (Strategy 7) and the requirement of negative PCRs for leisure activities (Strategy 11) are also effective in reducing the COVID-19 incidence, although by different mechanisms: Strategy 7 increases the number of tests while Strategy 11 reduces the transmission risk during leisure time.

Fig. 8 evaluates, for Strategy 1, the effect of the percentage of quarantine breakers on the final number of infections. We can observe that, even for a small increment in the percentage of quarantine breakers, the final percentage of infected individuals experiences a sharp increase until reaching a maximum value that corresponds to the maximum level in disease propagation among all strategies. The effect in the change the test window size (defined as the minimum time required between two tests applied to the same individual) is shown in Fig. 9. We observe that in our experiments, this parameter does not seem to have a strong impact on the infection reduction.

In this work we evaluate when test campaigns proposed in Strategy 9 that occur at different time of the infection spread. In Fig. 10 shows the distribution of the number of infected cases for different campaign starting times. Note that the test campaign only last one week. Fig. 11 shows final infected population percentage for each case. We can observe that the efficacy of this strategy is strongly dependent of the time that is applied. An early use of this strategy provides the maximum contention in the infection spread.

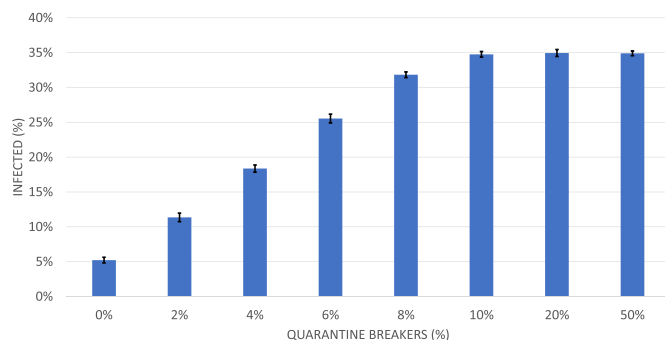


Fig. 8. Percentage of infected population while increasing the Quarantine breakers percentage. While changing quarantine breakers percentage, the rest of parameters remain as before: Test Window is 15 days and Quarantine Period is 10 days.

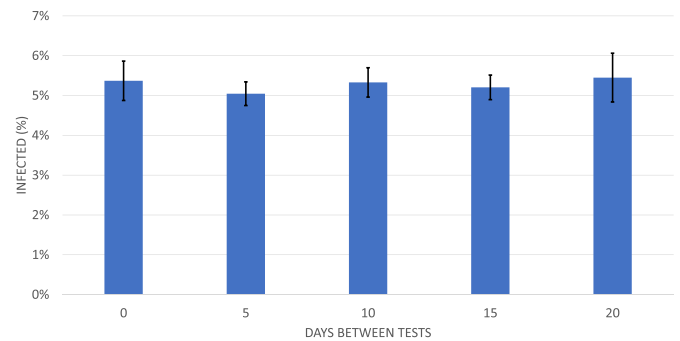


Fig. 9. Percentage of infected population while increasing the test window period. While changing test window period, the rest of the parameters remain as before: Quarantine Breakers is 0%, and Quarantine Period equal to 10 days.

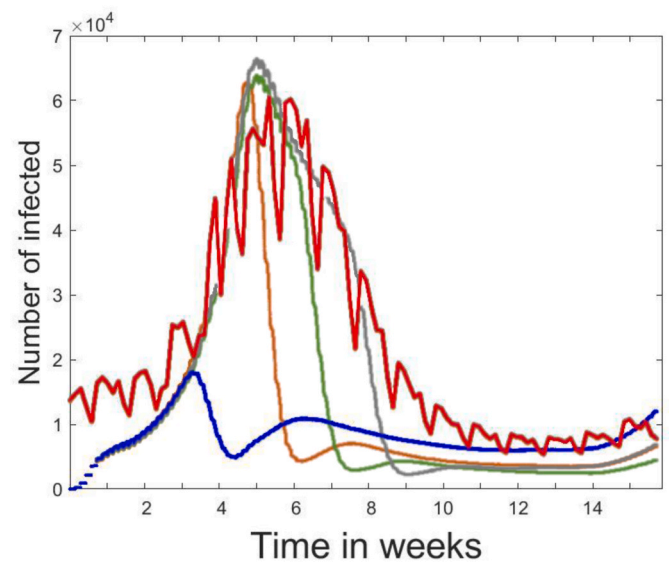


Fig. 10. Infected wave comparison when changing test campaign week in Strategy 9. The test campaign starting on week 3 is represented with the blue line, week 5 with the brown line, week 6 with the green and 7 with the grey. The red line represents the real infection curve for Spain.

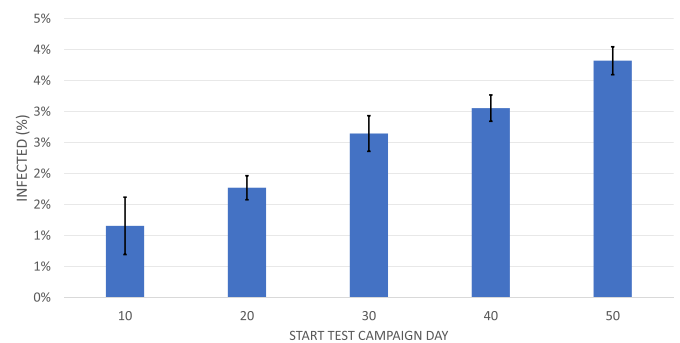


Fig. 11. Infected comparison when changing test campaign week in Strategy 9. The rest of parameters remain as the previous ones: Test Window is 15 days, Quarantine Breakers is 0%, and Quarantine Period is 10 days.

4. Discussion

In this section we will first discuss the suitability of different models for simulating the sampling strategies. We then focus on the effectiveness and suitability of the sampling strategies. Simulation tools proved

to be a useful tool to support decision-making on part of health authorities and have been broadly used for modelling COVID-19 propagation in different scenarios. Forecasting models such as the COFFEE from Los Alamos National Laboratory [3] do not explicitly model the effects of interventions or other ‘what-if’ scenarios; they are therefore unable to set hypothesis and test them. This is necessary to be able to evaluate and find the best sampling strategies. Other works are based on learning over existing data, which is not a good option when only scarce data is available. For instance, DELPHI (Differential Equations Leads to Predictions of Hospitalizations and Infections) [21] models the potential impact of various policies on future infections by estimating the average effect of each measure as implemented across states, via training. Youyang Gu’s COVID-19 model [13] applies machine learning to derive the basic reproduction number (R_0) from data published by Johns Hopkins University’s Center for Systems Science and Engineering (CSSE), and hooks this to a compartmental model. Their infection estimates include all infected individuals of the SARS-CoV-2 virus, not just those that took a COVID-19 test and tested positive.

In [30], the authors use a deterministic SEIR framework to model the propagation of the virus and the effect of non-pharmaceutical interventions (social distancing mandates and mask use) until the Spring of 2021. Some of the limitations of this approach are the absence of age structure and mixing within location (assumption of a well-mixed population), and the inability to model super-spreader-like events. These are very important factors when applying sampling strategies - and the measures following the detection of positive cases. Flaxman et al. [10] describes an extension of a semi-mechanistic Bayesian hierarchical model that infers the impact of interventions and estimates the number of infections over time. The working assumption is that changes in R_0 are an immediate response to interventions rather than broader gradual changes in behavior, and are calculated backward from temporal data. Covasim [18] includes demographic information about age structure and population size; realistic transmission networks in different social layers, including households, schools, workplaces, and communities; age-specific disease outcomes; and intra-host viral dynamics, including viral-load-based transmissibility. Different from our work, the contacts are not based on existing patterns; scalability issues are partly side-stepped by dynamic scaling.

In [4], the authors claim that sampling strategies must be focused on target groups to increase the effectiveness of the sampling policies; these may be an age group [14,36] or a high-risk group such as those of people in nursing homes, where both residents and staff are in close contact [8, 12]. According to our experiments, these policies may contribute to reducing the incidence within certain collectives, but to reduce the overall incidence it is necessary to continue testing the population from other collectives. While there is agreement about quarantine times for those infected and their contacts, various studies such as [2,19,38] suggest introducing additional testing to reduce this time, approaches known as test-and-release strategies. In Ref. [2], the authors propose that when an individual that has been in contacts with a confirmed SARS-CoV-2 case gets a negative on day 5, then it is possible to the lift the of quarantine two days later (on day 7), reducing the social-distancing time imposed to the individual. They also show that performing test-and-release on day 6 has almost the same benefit as a 10 day quarantine for returning travelers.

One of the drawbacks of RT-PCR testing is that produces false results in a significant percentage of the cases. In order to increase the diagnosis accuracy, complementary techniques [35] can be used in combination with sampling. Due to the economic cost of the testing campaigns, other alternatives have been explored to reduce these costs with minimum detection loss of positive cases. In this context, the pooling method seems to be a promising approximation [28]. According to our experiments, this pooling (Strategy 7) is able to achieve a significant reduction in the percentage of infections because of a much larger number of samples taken compared to other strategies. In Ref. [1], the authors describe sojourn time, the duration before clinical symptoms become

apparent but during which it is detectable by a screening test. Its clinical relevance is that it represents the duration of the temporal window of opportunity for early detection. Via a simple sensitivity analysis, they determine the most important parameters in the model, which turn out to be the fraction of cases that are asymptomatic. This model allows to consider infection by asymptomatic individuals.

In this work we have made several assumptions which we summarize here. First, the NPIs are the same for all the cities, i.e. they are homogeneously applied to the entire country despite the fact that there may be some differences between the regions. Given that these differences are small, we believe that they should not have a big impact on the simulation outcome. Secondly, the contact matrices (for population mixing) apply to a pre-COVID-19 scenario. Given that the social model of the simulator reflects distancing policies that restrict contacts between certain population groups, population mixing is reduced to what could realistically happen in a pandemic scenario. Thirdly, we assume that the sensitivity value of the PCR tests is 100%. A smaller sensitivity would produce a larger number of undetected cases. However, given that current testing achieves very high values, the effect of this assumption should not be significant.

EpiGraph has several limitations, some of them related to the disease model. In the current version of EpiGraph, a recovered individual acquires indefinite immunity to the virus, which makes it impossible to be re-infected for the duration of the simulation, which can be inaccurate, specially for long-term simulations. EpiGraph does not consider attributes such as previous pathologies, and other factors not well understood for now, that may also come into play when we evaluate risk of developing COVID-19 severe symptoms. Adding such attributes is straightforward. Our transmission model does not include other transmissions models such as surface contact, which are relevant for COVID when the population is not using masks or not washing their hands often.

Another limitation is related to the transportation model. In the model, the movement of individuals between cities depends on the distance between the cities and the population size. Having real knowledge about mobility patterns, for instance about those individuals using public transportation means, would provide a much more realistic approximation in this model. In our experiments we only model the largest urban regions in Spain; we could add more information related to smaller cities and towns, including rural regions.

5. Conclusion

This work evaluates the efficacy of sampling to reduce virus propagation by implementing strategies to study the effect of different parameters such as target groups, testing frequency, quarantine time, or group testing. Our evaluation is based on simulation to measure the quantitative and qualitative effect that these strategies and parameters have on slowing down the propagation of COVID-19. The first contribution of this work is the introduction of a social model that reflects social mixing patterns that are crucial when modeling interactions in a realistic environments. These patterns include breaking down some collectives (i.e. elderly people and workers) into sub-groups and professions, some of them with specific interaction structure (e.g. teacher-students, doctor-patient, etc). We also apply a new technique which uses contact matrices to determine the number of connections of each individual with others, depending on their ages. Contact matrices are specific to each country (Spain in our experiments).

We perform the evaluation of each of the sampling strategies by simulating the propagation over a network of 19,574,086 people (from the 63 most populated Spanish cities) that has realistic social and demographic characteristics, representative of Spain. The simulator was initially calibrated based on the existing prevalence values at the beginning of the third wave in Spain (at the end of 2020). The results we presented show that the most effective strategies are either pooling, rapid antigen test campaigns, or negative test requirements for access to

public areas, when followed by policies for reducing the number of contacts for infected individuals.

Declaration of competing interest

All the authors of the work: None Declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbimed.2021.104938>.

References

- [1] J. Arino, S. Portet, A simple model for COVID-19, *Infectious Disease Modelling* 5 (2020) 309–315.
- [2] Peter Ashcroft, Sonja Lehtinen, Daniel C. Angst, Nicola Low, Sebastian Bonhoeffer, Quantifying the impact of quarantine duration on covid-19 transmission, *Elife* 10 (2021), e63704.
- [3] Lauren Castro1, Geoffrey Fairchild, Isaac Michaud, Dave Osthus, COFFEE: COVID-19 Forecasts Using Fast Evaluations and Estimation, Los Alamos National Laboratory, 2020. LA-UR-20-28630.
- [4] Matthew Cleveley, Daniel Susskind, David Vines, Louis Vines, Samuel Wills, A workable strategy for covid-19 testing: stratified periodic testing rather than universal random testing, *Oxf. Rev. Econ. Pol.* 36 (Supplement_1) (2020) S14–S37.
- [5] Erik Cuevas, An agent-based model to evaluate the covid-19 transmission risks in facilities, *Comput. Biol. Med.* 121 (2020) 103827.
- [6] Deutsche Welle (DW), Covid: testing times for Germany's retail sector. <https://www.dw.com/en/covid-testing-times-for-germanys-retail-sector/a-57209054>, 2021. (Accessed 3 July 2021) accessed.
- [7] ECDC, European centre for disease prevention and control european centre for disease prevention and control (ECDC). <https://www.ecdc.europa.eu/en/covid-19/surveillance/testing-strategies>, 2020. (Accessed 3 July 2021) accessed.
- [8] Daniel J. Escobar, Maria Lanzi, Pouné Saberi, Ruby Love, Darren R. Linkin, John J. Kelly, Darshana Jhala, Valerianna Amorosa, Mary Hofmann, Jeffrey B. Doyon, Mitigation of a Covid-19 Outbreak in a Nursing Home through Serial Testing of Residents and Staff, *Clinical Infectious Diseases*, 2020.
- [9] Amanda Fernández-Fontelo, David Moríña, Alejandra Cabaña, Argimiro Arratia, Pere Puig, Estimating the real burden of disease under a pandemic situation: the sars-cov2 case, *PLoS One* 15 (12 December) (2020). December.
- [10] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Azra C. Ghani, Christl A. Donnelly, Steven Riley, A. Michaela, C. Vollmer, Neil M. Ferguson, Lucy C. Okell, Samir Bhatt, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Nature* 584 (2020) 257–261.
- [11] Google LLC. <https://developers.google.com/maps/>, 2021. (Accessed 3 July 2021). Google Maps API.
- [12] Neil SN. Graham, Cornelia Junghans, Rawlida Downes, Catherine Sendall, Helen Lai, Annie McKirdy, Paul Elliott, Robert Howard, David Wingfield, Miles Priestman, et al., Sars-cov-2 infection, clinical features and outcome of covid-19 in United Kingdom nursing homes, *J. Infect.* 81 (3) (2020) 411–419.
- [13] Youyang Gu, COVID-19 projections using machine learning. <https://covid19-projections.com>, 2020. (Accessed 3 July 2021) accessed.
- [14] Vincent Guilamo-Ramos, Benzekri Adam, Marco Thimm-Kaiser, Andrew Hidalgo, David C. Perlman, Reconsidering Assumptions of Adolescent and Young Adult Severe Acute Respiratory Syndrome Coronavirus 2 Transmission Dynamics, *Clinical Infectious Diseases*, 2020, p. ciaa1348, 09.
- [15] Robert Hinch, William JM. Probert, Anel Nurtay, Michelle Kendall, Chris Wymatt, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, et al., Openabm-covid19-an Agent-Based Model for Non-pharmaceutical Interventions against Covid-19 Including Contact Tracing, *medRxiv*, 2020.
- [16] Nicolas Hoertel, Martin Blachier, Carlos Blanco, Mark Olsson, Marc Massetti, Marina Sánchez Rico, Frédéric Limosin, Henri Leleu, A stochastic agent-based model of the sars-cov-2 epidemic in France, *Nat. Med.* 26 (9) (2020) 1417–1421.
- [17] ISCIII, Red nacional de vigilancia epidemiológica (RENAVE). <https://cneecovid.isciii.es/covid19>, 2020. (Accessed 3 July 2021) accessed.
- [18] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Romesh G. Abeyesuriya, Gregory Hart, Katherine Rosenfeld, Prashanth Selvaraj, Rafael C. Nunez, Brittany Hagedorn, Lauren George, et al., Covasim: an Agent-Based Model of Covid-19 Dynamics and Interventions, *medRxiv*, 2020.
- [19] Karl F. Krupp, Purnima Madhivanan, Carlos M. Perez-Velez, Should qualitative RT-PCR be used to determine release from isolation of covid-19 patients? *J. Infect.* 81 (3) (2020).
- [20] L. Lovasz, Random walks on graphs: a survey, *Volyai Soc. Math. Stud.* 2 (46) (1993).
- [21] Michael Lingzhi Li, Hamza Tazi Bouardi, Omar Skali Lami, Nikolaos Trichakis, Thomas Trikalinos, Mohammad Fazel Zarandi, Dimitris Bertsimas, Overview of DELPHI Model V3 - COVIDAnalytics, 2020.
- [22] G. Martín, M.C. Marinescu, D.E. Singh, J. Carretero, Leveraging social networks for understanding the evolution of epidemics, *BMC Syst. Biol.* 5 (S3) (2011).
- [23] Gonzalo Martín, David E. Singh, Maria-Cristina Marinescu, Jesús Carretero, Towards efficient large scale epidemiological simulations in epigraph, *Parallel Comput.* 42 (2015) 88–102 (Parallelism in Bioinformatics).
- [24] Instituto de salud Carlos III (ISCIII) Ministerio de sanidad de España. Estrategia de detección precoz, vigilancia y control de covid-19..
- [25] Ministry of economic affairs and digital transformation (MINECO), National Statistics Institute (INE), 2021. <http://www.ine.es/>. (Accessed 3 July 2021). accessed.
- [26] D. Mistry, M. Litvinova, A. Pastore, y Piontti, et al., Inferring high-resolution human mixing patterns for disease modeling, *Nat. Commun.* 12 (2021) 323.
- [27] Seyed M. Moghadas, Meagan C. Fitzpatrick, Pratha Sah, Abhishek Pandey, Affan Shoukat, Burton H. Singer, Alison P. Galvani, The implications of silent transmission for the control of covid-19 outbreaks, *Proc. Natl. Acad. Sci. Unit. States Am.* 117 (30) (2020) 17513–17515.
- [28] Leon Mutesa, Pacifique Ndishimye, Yvan Butera, Souopgui Jacob, Annette Uwineza, Robert Rutayisire, Ella Larissa Ndiricimpaye, Emile Musoni, Nadine Rujeni, Thierry Nyatanyi, et al., A pooled testing strategy for identifying sars-cov-2 at low prevalence, *Nature* 589 (7841) (2021) 276–280.
- [29] M. Pollán, B. Pérez-Gómez, R. Pastor-Barriuso, et al., Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study, *Lancet* 396 (10250) (2020) 535–544.
- [30] R.C. Reiner, R.M. Barber, et al., Modeling COVID-19 scenarios for the United States, *Nat. Med.* (2020).
- [31] Anoop S.V. Shah, et al., Risk of hospital admission with coronavirus disease 2019 in healthcare workers and their households: nationwide linkage cohort study, *BMJ* 371 (2020).
- [32] David E. Singh, Maria-Cristina Marinescu, Miguel Guzmán-Merino, Christian Durán, Concepción Delgado-Sanz, Diana Gomez-Barroso, Jesus Carretero, Simulation of covid-19 propagation scenarios in the madrid metropolitan area, *Front. Public Health* 9 (2021) 172.
- [33] Istvan Szapudi, Efficient Sample Pooling Strategies for Covid-19 Data Gathering, *medRxiv*, 2020.
- [34] Seda Arslan Tuncer, Hakan Ayyıldız, Mehmet Kalaycı, Taner Tuncer, Scat-net: covid-19 diagnosis with a cnn model using scattergram images, *Comput. Biol. Med.* 135 (2021) 104579.
- [35] Ranjit Kumar Upadhyay, Sourin Chatterjee, Satvik Saha, Rajeev K. Azad, Age-group-targeted testing for covid-19 as a new prevention strategy, *Nonlinear Dynam.* 101 (3) (2020) 1921–1932.
- [36] Cécile Viboud, Ottar N. Bjørnstad, David L. Smith, Lone Simonsen, Mark A. Miller, Bryan T. Grenfell, Synchrony, waves, and spatial hierarchies in the spread of influenza, *Science* 312 (5772) (2006) 447–451.
- [37] Chad R. Wells, Jeffrey P. Townsend, Abhishek Pandey, Seyed M. Moghadas, Gary Krieger, Burton Singer, Robert H. McDonald, Meagan C. Fitzpatrick, Alison P. Galvani, Optimal covid-19 quarantine and testing strategies, *Nat. Commun.* 12 (1) (2021) 1–9.
- [38] WHO, Who coronavirus (covid-19) dashboard. <https://covid19.who.int/>, 2021. (Accessed 3 July 2021) accessed.
- [39] World Health Organization (WHO), Considerations for quarantine of contacts of covid-19 cases. <https://www.who.int/publications>, 2020. (Accessed 3 July 2021) accessed.
- [40] Idan Yelin, Aharony Noga, Einat Shaer Tamar, Amir Argoetti, Esther Messer, Dina Berenbaum, Einat Shafraan, Areen Kuzli, Nagham Gandali, Omer Shkedi, et al., Evaluation of covid-19 RT-qPCR test in multi sample pools, *Clin. Infect. Dis.* 71 (16) (2020) 2073–2078.
- [41] 2.1, 2021. EpiGraph git repository. (Accessed 17 October 2021).