UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

# EVALUATING MITIGATION STRATEGIES IN THE COVID-19 PANDEMIC: A NETWORK MODEL WITH STOCHASTIC DYNAMICS

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

SIMÓN EMILIO MATURANA MOLINA

PROFESOR GUÍA:
MARCELO OLIVARES ACUÑA

MIEMBROS DE LA COMISIÓN:
JOSÉ CORREA HAEUSSLER
YERKO MONTENEGRO ORTIZ

SANTIAGO DE CHILE
2021

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN GESTIÓN DE OPERACIONES; INGENIERO CIVIL INDUSTRIAL
POR: SIMÓN EMILIO MATURANA MOLINA
FECHA: 2021
PROF. GUÍA: MARCELO OLIVARES ACUÑA

# EVALUATING MITIGATION STRATEGIES IN THE COVID-19 PANDEMIC: A NETWORK MODEL WITH STOCHASTIC DYNAMICS

**Introducción:** El Coronavirus 2019 (COVID-19) es una enfermedad causada por el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2). Tiene un riesgo significativo debido a su alto número reproductivo básico, que puede traducirse en UCI sobrepobladas en el sistema de salud, siempre que la infección no esté controlada. En la medida en que no hayan vacunas disponibles para todos y todas, la única forma de controlar la pandemia es mediante la prevención: aislamiento de casos, rastreo de contactos, cuarentenas, distanciamiento físico y medidas de higiene.
**Razón fundamental:** En este artículo simulamos cómo el virus se propagó a través de la población de la *Región Metropolitana* en Chile, usando una microsimulación. Utilizamos muchas fuentes de datos, incluido el censo, para recrear las principales características de las personas que viven en la región y sus desplazamientos. Para los desplazamientos utilizamos datos de teléfonos celulares para rastrear el movimiento de las personas, midiendo su cambio en sus desplazamientos con respecto a los mismos antes de la pandemia. Para evaluar las medidas de control, implementamos el mecanismo trazabilidad utilizado en Chile en diferentes niveles, para identificar qué tan dura debe ser la medida para controlar la infección rápida.
**Resultados:** Desarrollamos una microsimulación donde los contagios ocurren agente por agente. Pueden pasar por distintos estados que dependen (y corresponden) a su comportamiento frente a la pandemia. También utilizamos un método de máxima verosimilitud a través del tiempo para calibrar el modelo. El modelo estimó que trazabilidad realizada en agosto-2020 equivale a testear y aislar al 20 % (o 10 %) de las personas y rastrear al 10 % (o 20 %) de sus contactos. Además, descubrimos que la eficacia de la trazabilidad necesaria para reducir a cero las nuevas infecciones es mayor a un 20 % en ambas dimensiones.

# Abstract

**Introduction:** Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It carries significant risk because of its high basic reproductive number, which can translates in overcrowded ICUs in the health system as long as the infection is uncontrolled. While vaccines are not developed the only way to control the pandemic is by prevention: case isolation, contact tracing, quarantines, physical distancing and hygiene measures.
**Rationale:** In this paper we simulated how the virus spread through *Región Metropolitana* population in Chile using a microsimulation. We used many source of data, including census, to recreate the main characteristics of the people living in the region and their commuting behavior. For the commuting we used cellphone data to track the movement of the people, measuring their change in their commuting with respect to the same before to the pandemic. To evaluate control measures we implemented the contact tracing mechanism used in Chile at different levels, to identify how hard the measure need to be in order to control the rapid infection.
**Results:** We developed a microsimulation where the contagions occur agent by agent. They can pass through states which correspond to their behavior in the face of the pandemic. We also used a maximum likelihood through time method to calibrate the model. The model estimated that the contact tracing carried out in August-2020 is equivalent to test and isolate 20 % (or 10 %) of the people and trace 10 % (or 20 %) of their contacts. Moreover, we found that the contact tracing effectiveness necessary to decrease new infections to zero is higher.

# Tabla de Contenido

# Índice de Tablas

# Índice de Ilustraciones

# Introducción

Coronavirus disease 2019 (COVID-19) is a rapidly spreading infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has become a global pandemic. It has a basic reproductive number of 2.2 to 2.5 as determined in Wuhan during the first months of spreading [3] [4].

COVID-19 spread can be fought with both pharmaceutical and nonpharmaceutical interventions (NPIs). The former concern vaccine immunity for the general population and treatments to reduce mortality and critical rates for the infected population. The latter category is broad and can encompass personal protective equipment and hygiene plus social distancing measures for the general population and confinement strategies and contact tracing for the infected population. This work focuses on NPIs since these are the best way to decrease virus spread in the absence of a vaccine.

Since the main goal is to develop a tool to evaluate alternative NPI mitigation strategies, 3 principal modeling approaches were identified. One alternative is to use machine learning [5], which is flexible, maximizes predictive power, and optimally selects among multiple prediction variables but is limited to extrapolating predictions under system interventions. Another alternative are the well-known susceptible-infected-recovered (SIR) models [6] [7] [8], which are parsimonious models to capture the structure of an epidemic and are feasible to estimate with limited data but do not account for the structure of the social/geographic network and the uncertainty and variability of the epidemic. The lasts alternatives are microsimulations [1] [9] [10] that capture details of the patterns of the social/geographic network and incorporate uncertainty and variability but require large datasets and careful construction, and they are computationally intensive.

This work is a microsimulation with the objective of assessing different NPIs in *Región Metropolitana* to mitigate COVID-19 propagation. Spread occurs from individual to individual with a probability that depends on the contagion force and the regimen followed by the infected, which changes according to the average mobility of their zone. A key aspect to consider in the development of this model is the dynamics of contagion. The attributes used to characterize this were incubation time [11], infectiousness over time and the serial interval [12].

Given that the compliance of the population with lockdown policies is low and depends on the income level of the people of each zone [2], cellphone data were used to quantify mobility from one commune to another (i.e., lockdown policy compliance).

Calibration of the simulation is performed with maximum likelihood methods. To find the parameters that provide the best fit, a single likelihood function is calculated for each of the simulated days, and the sum of all day contributions is maximized. Once the model is calibrated, a contact tracing system is implemented to test how many people would be necessary to trace and how many of their contacts would be necessary to reach.

The principal contribution of this work is the development of a tool that allows to asses the effect that different NPIs, like quarantines, testing and contact tracing measures, have on the pandemic. From this tool, the principal finding is that virus spread can be controlled by tracing 20 % of the infected and reaching 20 % of their contacts. Higher values of any of those metrics lead to a substantial reduction in new cases.

# Capítulo 1

# Transmission network model

The simulation of transmission is modeled through a network of individuals where connections represent physical contacts in the population and the evolution of the contagion process follows Markov processes. States represent the health condition of each individual (susceptible, multiple stages of infection and recovered), and contagion from susceptible to infection is probabilistic. This section describes the details of how the contact network is determined and the probability model used to capture the stochastic behavior of the contagion process.

## 1.1. Network structure

The population is modeled through a graph $(V, E)$, where the vertices (nodes) $V$ represent individuals and the set of *undirected* edges $E$ capture potential physical contacts among them. This contact network is partitioned into four superimposed sets of edges representing different places where contact can occur: (i) household ($E_H$); (ii) school ($E_S$); (iii) work ($E_W$); and (iv) other contacts, referred to as *community* ($E_C$).

Each individual in $V$ is characterized by their (i) age, which affects their severity of infection, and (ii) geographic location, specified by coordinates, which determine the network of contacts.

The population is located in zone $z$, which determines certain characteristics, such as the population density, age and gender of individuals and number of inhabitants in each household. All of this is simulated to match the patterns of the region studied.

Each individual i is assigned a gender, an age, a zone $z$ (matching the number of inhabitants of the corresponding zone) and a household $h$(i) belonging to that zone. Households become a clique, forming a complete subgraph among individuals assigned to the same household. The union of the household cliques forms the set of edges $E_H$. Figure 1.1 shows an example of connections within households.

Each individual associated with a household $h$ can have social contact with his or her neighbors, which are represented by the nodes located at $\bar{\mathrm{d}}$ kilometers as a maximum from the origin node. Community contacts are generated through a random graph following a

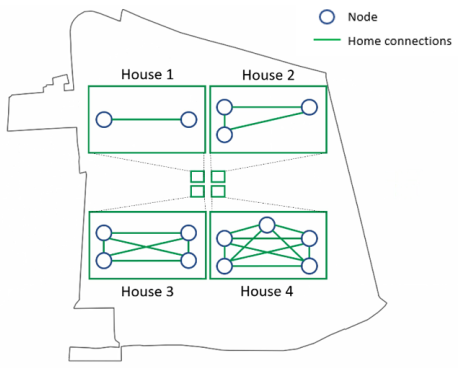



Figura 1.1: An example of four households belonging to a zone and having different numbers of inhabitants. Nodes inside the same household are connected to one another, forming a clique.

power law. Two nodes are connected with probability $f_{pow}(\mathrm{d})$, specified in Section 1.2, which depends on the geographic distance d between two nodes. These community connections form the set of edges $E_C$, with each node having an average connectivity of 10 edges. Figure 1.2 illustrates how nodes can be connected to one another regardless of the node an individual inhabits.

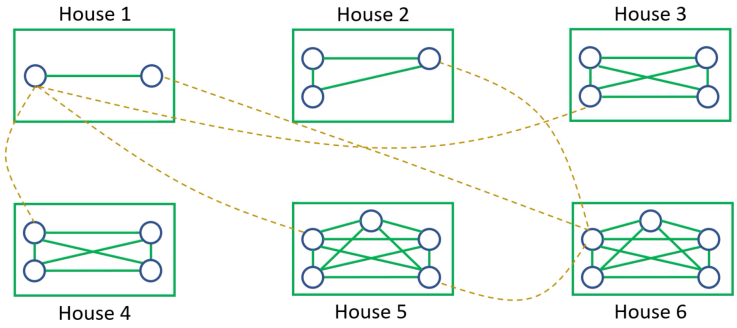
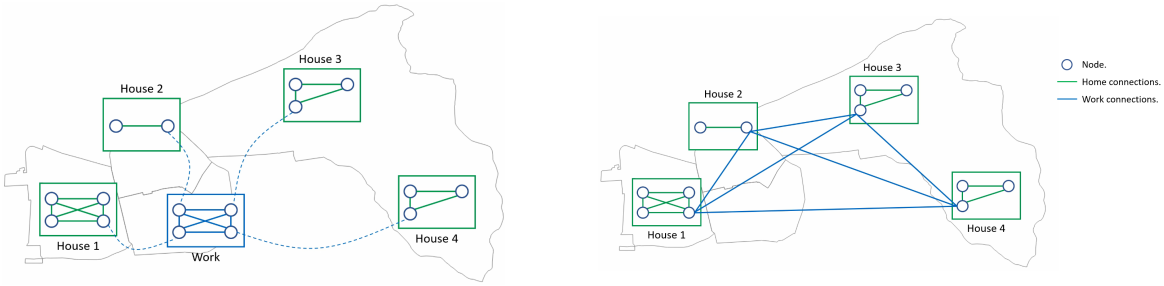



Figura 1.2: An example of six households in which their occupants are connected to their neighbors by randomly generated edges. This allows connectivity among households over short distances.

A fraction of the nodes are assigned to a workplace $w(\mathrm{i})$ with probability $p_{az}$, depending on the age $a$ of the individual and the zone $z$ to which they belong. Each individual i living in zone $z_{h(\mathrm{i})}$ has a probability $Commute^{W}_{z_{h(\mathrm{i})}z}$ of working in zone $z$ given that they live in zone $z_{h(\mathrm{i})}$. On that basis, an individual is assigned to a zone $z$ and uniformly randomly assigned to a workplace $w(\mathrm{i})$ belonging to that zone. There are as many workplaces in each zone as necessary to hold the people working in them. Each workplace has a certain number of workers that follows a size distribution. Workplaces become cliques connecting all individuals assigned to them: two individuals with workplaces $w(\mathrm{i}) = w(j)$ are connected, and the union of these cliques forms the set of edges $E_W$. Figure 1.3 shows how connections work.

Another fraction of nodes are assigned to schools. Each child with a home in zone $z_{h_\mathrm{i}}$ has a probability $Commute^{S}_{z_{h(\mathrm{i})}z}$ of studying in zone $z$ given that they live in zone $z_{h(\mathrm{i})}$. They are assigned to a zone $z$ and then uniformly randomly assigned to a school $s(\mathrm{i})$ in that zone.




(a) (b)

Figura 1.3: figure 1.3a Nodes of different households are assigned to a workplace where they meet one another. The system presented in Figure 1.3b is simplified by eliminating workplaces but maintaining connections among nodes assigned to them.

Schools are cliques connecting students, forming the set of edges $E_S$.

Figure 1.4 shows an example of how different types of cliques with their edges $E$ interact with one another. This allows all nodes to be connected even if they are not in the same zone or are far away.

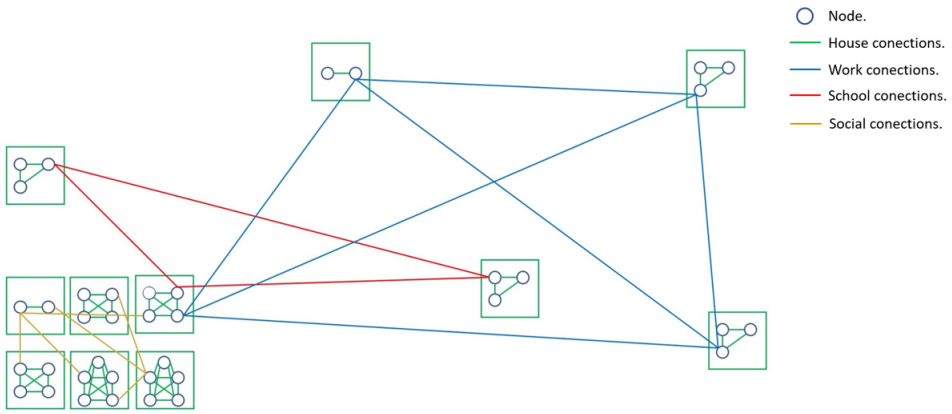



Figura 1.4: The graphic indicates how nodes are connected. Green lines denote home connections, blue lines denote work connections, red lines denote school connections and yellow lines denote social connections.

## 1.2. Node states and infection transmission

Nodes in the network may evolve through nine states described in Table 1.1 and illustrated in Figure 1.5. All nodes are initialized in the *susceptible* state $NI$. Infected cases in the contagious stage transmit the infection to susceptible individuals, who can move to an *asymptomatic* $I_a$ state, with probability $1-r$ or to a *presymptomatic* $I_p$ state, with probability $r$. Let $T_{\mathrm{i}}^c$ be the time that an individual i is infected.

Let $T_{\mathrm{i}}^C$ be the time at which individual i become infected, $T_{\mathrm{i}}^L$ be a random variable denoting the incubation period of individual i and $T^I$ be the time before symptoms emerge when contagiousness begins (fixed and equal for each). A *presymptomatic* individual will become *contagious presymptomatic* $I_p^c$ at time $T_{\mathrm{i}}^C + T_{\mathrm{i}}^L - T^I$. In this state, individuals can infect other people even if they do not present symptoms. Thereafter, the node may become *symptomatic* ($I_s^c$), with probability $p_{symp,a}$, *hospitalized* ($I_h^c$), with probability $(1-p_{symp,a})p_{hosp,a}$, or *critical* ($I_{cr}^c$), with probability $(1-p_{symp,a})(1-p_{hosp,a})$ at time $T_{\mathrm{i}}^C + T_{\mathrm{i}}^L$. Symptomatic, hospitalized and critical patients will remain in this state for $T^S$, $T^H$ and $T^P$ days, respectively, until they recover, becoming *postsymptomatic* $P$.

On the other hand, an *asymptomatic* individual has a virtual incubation period, which is important because they become *contagiously asymptomatic* at time $T_{\mathrm{i}}^c + T_{\mathrm{i}}^L - T^I$. *contagious asymptomatic* nodes will remain in this state for $T^A$ days since symptoms would have emerged. Finally, they recover, becoming *postsymptomatic* $P$.

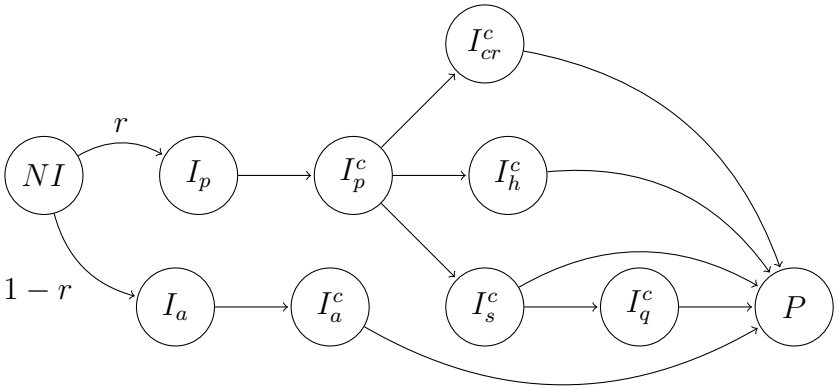



Figura 1.5: Dynamics of the virus spread. An individual starts as *susceptible* $NI$ until he or she is infected by a contagious source. They can become *presymptomatic* or *asymptomatic*. After the latency period (virtual in the case of asymptomatic individuals), they become *contagious presymptomatic* $I_p^c$ and *contagious asymptomatic* $I_a^c$, respectively, not yet presenting symptoms. The presymptomatic, after a fixed amount of time, can become either *symptomatic* $I_s^c$, *hospitalized* $I_h^c$ or *critical* $I_{cr}^c$. When an individual is *hospitalized*, *critical* or *asymptomatic*, after an amount of time different for each, they will be *postsymptomatic* $P$. In the case of the *symptomatic*, they can either self-isolate, become *quarantined*, or become *postsymptomatic*.

Transmission rates depend on the demographics, state and contact network of each infected node. Each node $k$ that enters the infectious period can infect its connected nodes in the susceptible state. We use the index $p \in \{H, W, S, C\}$ to denote a place where infections may occur, which is associated with a set of edges connecting nodes, $E_p$. In every period $t$, the

| State | Notation | Description |
|---|---|---|
| Susceptible | $NI$ | Nodes that have not had the disease and are susceptible to be infected |
| Asymptomatic | $I_a$ | Nodes that are infected, will never present symptoms and do not spread the virus. |
| Contagious asymptomatic | $I_a^c$ | Identical to the asymptomatic except that they are capable of spreading the virus. |
| Presymptomatic | $I_p$ | Individuals who are infected and will present symptoms but are not yet symptomatic |
| Contagious presymptomatic | $I_p^c$ | Identical to the presymptomatic except that they are able to spread the virus |
| Symptomatic | $I_s^c$ | Individuals who are infected, present symptoms and spread the virus. |
| Quarantined | $I_q^c$ | Identical to the symptomatic except that they are self-isolated, reducing contact with other people. |
| Hospitalized | $I_h^c$ | Individuals who require special care in hospitals. Their contact is considerably reduced, and they remain sick longer. |
| Critical | $I_{cr}^c$ | Individuals who require critical care in hospitals. Their contact is considerably reduced, and they remain sick longer than the hospitalized. |
| Postcontagious | $P$ | Individuals who have recovered or died after contracting the disease. They do not affect nor are affected by the diseased. |

Tabla 1.1: States of the individuals through the infectious process.

probability that infected node $k$ transmits the infection to a susceptible node i connected through an edge in $E_p$ is given by:

$$\Pr(k \to_p \text{i}) = 1 - \exp(-q_k^p(t)T_p) \tag{1.1}$$

where $q_k^p(t)$ is the *infection rate* and $T_P$ is the exposure interval of the contact, both of which depend on the type of contact $p = \{H, W, S, C\}$.

For the home, work and school contact networks ($\{E_p\}_{p\in\{H,W,S\}}$), the infection rate is specified as:

$$q_k^p(t) = \frac{\beta_p \phi(t-\tau_k)\rho_k \left[1 + C_k\left(\omega - 1\right)\right]}{n_p(k)^{\alpha_p}} \cdot \eta_{\{t\geq \bar{t}\}} \tag{1.2}$$

where:

- $\beta_p$ is the parameter capturing the force of infection associated with the contact network $E_p$.
- $\tau_k$ is the time of the onset of symptoms of the index case $k$.
- $\phi(t-\tau_k)$ corresponds to the infectiousness decay function, which depends on the elapsed time in which symptoms start.
- $\rho_k$ represents the heterogeneity of the force of infection across individuals.

- $C_k$ is 1 if individual $k$ is a critical infection and 0 otherwise.
- $\omega$ is the infectiousness of a critical infection relative to a mild one.
- $n_p(k)$ is the number of connected individuals in the clique of the contact network $E_p$ associated with the infected case.
- $\alpha_p$ is an external parameter to scale $n_p(k)$, fixed as $\alpha_H = 0{,}8$ for households and $\alpha_W = \alpha_S = 1$ for work and school.
- $\eta$ is a factor that scales the force of contagion, representing the effect of full lockdown. This begins to have consequences after time $\bar{t}$, for example, $\eta_{\{t \geq \bar{t}\}} = 1 - (1 - \eta) \cdot \mathbb{1}_{\{t \geq \bar{t}\}}$
- $\bar{t}$ is the time at which full lockdown in *Región Metropolitana* started.

For community contacts, the infection rate depends on the Euclidean distance $\mathrm{d}_{k\mathrm{i}}$ between the index case ($k$) and the contact (i) and is specified as:

$$q^C_{k\mathrm{i}}(t) = \beta_C \varsigma(a_k) \phi(t - \tau_k) \rho_k \left[1 + C_k \left(\omega - 1\right)\right] \frac{f_{pow}(\mathrm{d}_{k\mathrm{i}})}{\sum_{(k,j) \in E_C, \mathrm{d} \leq \bar{\mathrm{d}}} f_{pow}(\mathrm{d}_{kj})} \eta_{\{t \geq \bar{t}\}} \tag{1.3}$$

where:

- $\varsigma(a_k)$ is a relative contact rate that depends on the age $a_k$ of individual $k$. The values of this function are specified in Table 1.2.
- $f_{pow}(\mathrm{d}_{k\mathrm{i}}) = \frac{1}{1 + \left(\frac{\mathrm{d}_{k\mathrm{i}}}{a}\right)^b}$, with $a = 4km$ and $b = 3{,}8$ [13].
- $\mathrm{d}_{k\mathrm{i}}$ corresponds to the Euclidean distance between case $k$ and contact i
- $\bar{\mathrm{d}}$ is the maximum distance at which a case can be linked to a contact, with $\bar{\mathrm{d}} = 0{,}5km$

| **Age $a_k$** | **Relative contact rate $\varsigma(a_k)$** |
|---|---|
| 0 to 5 years | 10 % |
| 5 to 10 years | 25 % |
| 10 to 15 years | 50 % |
| 15 to 20 years | 75 % |
| 20 to 65 years | 100 % |
| 65 to 70 years | 75 % |
| 70 to 75 years | 50 % |
| 75 or more years | 10 % |

Tabla 1.2: Relative contact rate depending on the age of each individual. These factors simulate how people with different ages interact with the community and were taken from Ferguson (2020) [1]

# Capítulo 2

# Data collection

To precisely characterize the way the virus spreads, data were used to represent the main qualities of COVID-19 and the qualities of the population exposed. The population has demographic attributes and certain social interaction patterns, which influence the progress of the outbreak.
The main objective of identifying population characteristics is to make a graph $G = (V, E)$ where the vertices $V$ and the edges $E$ model the position of each person and his or her connections, respectively. On the other hand, the properties of the virus help us to model how fast the virus spreads and for how long an individual can spread it.

## 2.1. Virus characteristics and assumptions

SARS-CoV-2, the causative virus of COVID-19, has some characteristics that directly affect the way the disease spreads. It has been shown that its reproductive number was between 2.2 and 2.5 in Wuhan [3] [4] and 2.6 in the Republic of Korea and Italy [14].

The main assumptions on COVID-19 used in this model are listed below. The mean durations of states of the infection are shown in Table 2.1.

1. The incubation period is log-normally distributed with a mean of 5.2 days and the 95th percentile of the distribution at 12.5 days [12].
2. Infectiousness starts 2.3 days before symptom onset and peaks at 0.7 days before symptom onset, with an estimated proportion of presymptomatic transmission (area under the curve) of 44 %. This is modeled with a gamma density distribution [12].
3. According to Ferretti et al.[15], the relative infectiousness of asymptomatic with respect to symptomatic patients is 0.1.
4. Infection and infectiousness will last for 10 days after the onset of symptoms for symptomatic patients and 10 days after the virtual onset of symptoms for asymptomatic patients. This is based on He et al. [12], who state that infectiousness declines significantly 8 days after the onset of symptoms, which is consistent with our infectiousness function. For standard hospitalized and critical hospitalized patients, their infection will last for 20 and 29 days, respectively [16].

5. A total of 45 % of the cases will be asymptomatic. We also assumed an asymptomatic proportion of agents of 45 % [17].

| **Time period** | **Symbol** | **Mean (std.dev.)** | **Distribution** |
|---|---|---|---|
| Time at which contagion occurs | $T_{\mathrm{i}}^{C}$ | – | – |
| From contagion to onset of symptoms | $T_{\mathrm{i}}^{L}$ | 5,2 (3,91) | Log-normal |
| Time before onset of symptoms in which contagiousness starts | $T^{I}$ | 2,3 | Deterministic |
| From onset of symptoms to self-isolation | $T_{\mathrm{i}}^{SI}$ | 2,0 (1,15) | Log-normal |
| From virtual onset of symptoms to recovered for asymptomatic | $T^{A}$ | 10 | Deterministic |
| From onset of symptoms to recovered for symptomatic | $T^{S}$ | 10 | Deterministic |
| From onset of symptoms to recovered for hospitalized patients | $T^{H}$ | 20 | Deterministic |
| From onset of symptoms to recovered for critical patients | $T^{P}$ | 29 | Deterministic |

Tabla 2.1: Distributions of time periods of node states, their means and their standard deviations (if the time period is a random variable).

## 2.2. Case incidence data adjustments

During the pandemic, there were two sources of data that suggested that there were some problems with the case incidence data. The first is that positivity was highest between May and June of 2020 (see Figure 2.1). This demonstrates that the probability of finding a new case among people who took a test is high, suggesting that there could be more infected people than detected. The second piece of evidence is the probability that the results of a *PCR* test being ready before 3 days since the onset of symptoms was the lowest (see Figure 2.2) during the same period. This is evidence that there was an important delay in reviewing the tests and the data reported were not the most accurate.

To correct the positivity issue, we applied the methods used in Russel (2020) [18]. We assumed that the base case fatality ratio (CFR) is 1.4 %, based on the reported average in China and Korea (1.38 % and 1.4 %, respectively) [17] [19].

Typically, a naive CFR is calculated by:

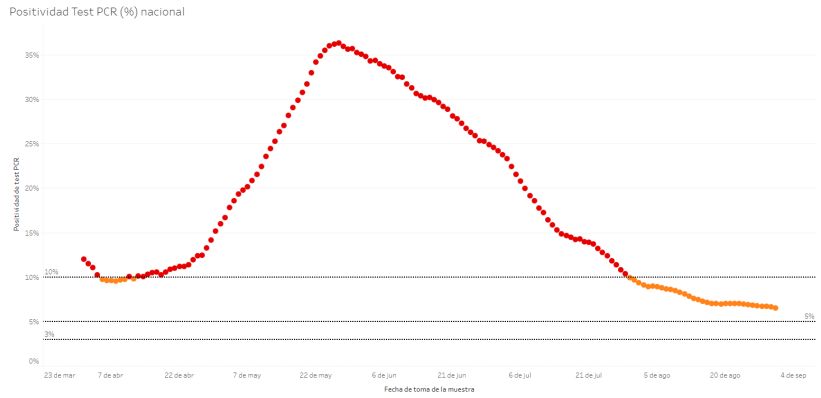



Figura 2.1: Source: icovidchile.cl. Positivity of tests conducted during the pandemic between April 1 and August 31.

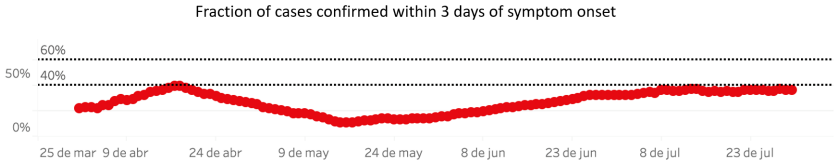



Figura 2.2: Probability that a test will be analyzed and delivered in less than 3 days

$$nCFR_t = \frac{\sum_{\mathrm{i}=1}^{t} \mathrm{d}_t}{\sum_{\mathrm{i}=1}^{t} c_{\mathrm{i}}}$$

where $\mathrm{d}_t$ and $c_t$ denote cumulative deaths and cumulative case incidence until date $t$, respectively. Instead, we used delayed CFR to correct for the time within which a person who became sick would have died. This is given by:

$$\mathrm{d}CFR_t = \frac{\sum_{\mathrm{i}=1}^{t} \mathrm{d}_t}{\sum_{\mathrm{i}=1}^{t} \sum_{j=1}^{\infty} c_{\mathrm{i}-j} g_j}$$

where $g_j$ is the distribution of the delays from the onset of symptoms to death. Having this indicator, we can define

$$u_t = \frac{1{,}4\,\%}{\mathrm{d}CFR_t}$$

, which shows the underreporting rate, assuming that the actual CFR is equal in all countries and that the CFR estimated in China and South Korea is taken into account in the actual case incidence number. With these factors, we can correct the Chilean case incidence series for underreporting, taking $\hat{c}_t = \frac{c_t}{u_t}$. We used data reported for the ICOVID Chile team [20] as input. They used statistical models to estimate the number of daily infections based on the onset of symptoms. Figure 2.3 shows the change in the new daily infection rate according to the curve for the day of the onset of symptoms after adjusting for underreporting.

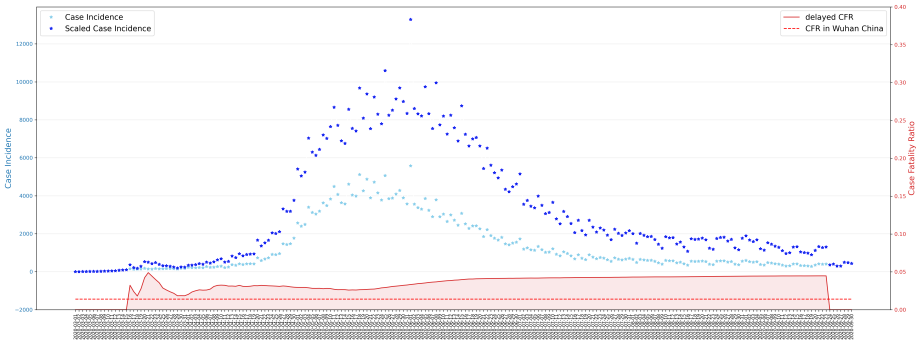



Figura 2.3: Correction of the case incidence series for underreporting.

To correct the test lag issue, we used the number of infected people according to the day of infection instead of the day of the onset of symptoms. This allowed us to avoid the lag between the time of infection and reporting. This also provides flexibility in the model calibration process.

To this end, we used data reported by the ICOVID Chile team [20]. They used statistical models to estimate the number of daily infections based on the onset of symptoms. To shift to contagion according to infection day, we defined the following relation:

$$T_{\mathrm{i}}^{L} = T_{\mathrm{i}}^{C} + T_{\mathrm{i}}^{L} + U_{\mathrm{i}} = T_{\mathrm{i}}^{C} + T_{\mathrm{i}}^{U}$$

where $U_{\mathrm{i}}$ is a canonical uniform random variable representing the day when the contagion occurred and $T_{\mathrm{i}}^{U}$ is the convolution between incubation time and $U_{\mathrm{i}}$. Thus, we construct the contagion curve. Figure 2.4 shows the changes in the series to obtain the one that will be used for calibration.

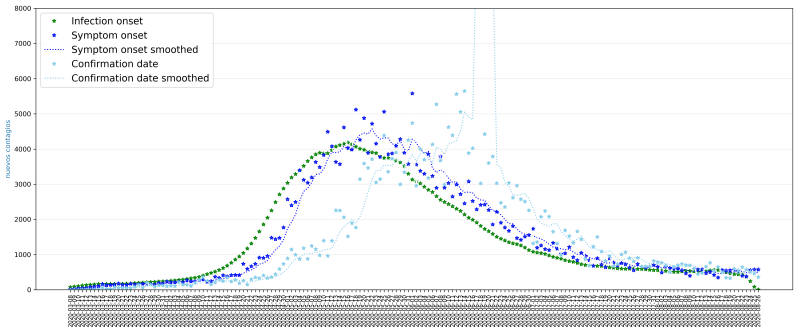



Figura 2.4: Change in series through transformations. Cyan series are daily new cases according to confirmation day. Blue series correspond to daily new cases according to the onset of symptoms. Green series show daily new cases according to contagion day.

## 2.3. Population characteristics

The agents in the model were built to recreate real characteristics and commuting patterns of the population from *Región Metropolitana* in Chile. To do so, data from the 2017 census [21] were used to describe the population density for each commune, age and gender of each agent (see Figure 2.6a) and number of residents in each household (see Figure 2.6b). Due to data availability, only 34 of the 52 communes were used, corresponding to the communes where public transport is available and resulting in $6,386,006$ individuals simulated. An example of how people are located and connected to one another is shown in Figure 2.5. Each color denotes one of the states explained in Section 1.2.

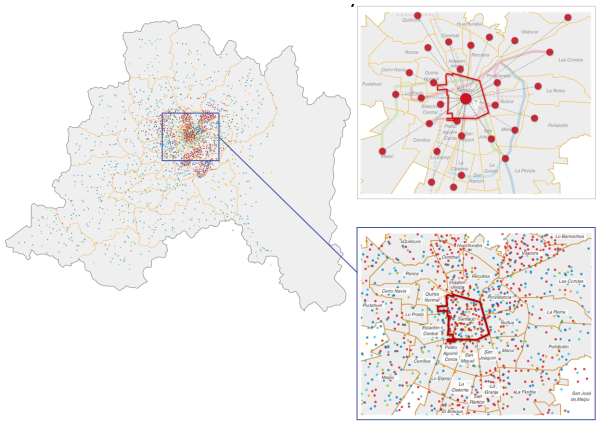
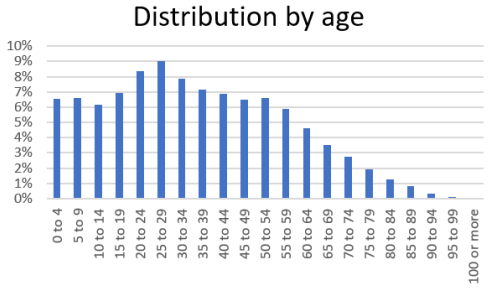
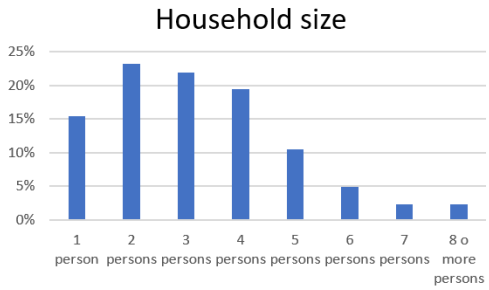


Figura 2.5: People are distributed among the communes of *Región Metropolitana* and connected to one another. Each color shows that an agent can be in different states over time, being susceptible first and postcontagious last.

Households were created with a minimum of 1 person and a maximum of 8, following the distribution of household size of each commune. Figure 2.6b shows the distribution for the whole region. Thus, people are assigned randomly to a household that has coordinates inside the region and its number of inhabitants, allowing people who live in the same household to be connected.




(a) Distribution of ages within the population of the region




(b) Distribution of the size of the households located in the region

Although every person belongs to a household, he or she can either go to a workplace, university, school or none of these. A node has probability $Occup^p_{a_i}$ of being associated with a place $p = \{H, W, S, C\}$, which depends on the age $a_i$ of the agent. First, a node is assigned

to a workplace with probability $Occup^W_{a_i}$. If not, the node can be assigned to a university with probability $\left(1 - Occup^p_{a_i}\right) \cdot Occup^u_{a_i}$. If not, the node can be assigned to a school with probability $\left(1 - Occup^w_{a_i}\right) \cdot \left(1 - Occup^u_{a_i}\right) \cdot Occup^s_{a_i}$. If not, the node remains unemployed with probability $\left(1 - Occup^w_{a_i}\right) \cdot \left(1 - Occup^u_{a_i}\right) \cdot \left(1 - Occup^s_{a_i}\right)$. The probabilities of what people do depend on age. Census and Education Ministry data [22] [23] indicate the proportion of people working for each age from 15 to 100. The same is true for universities for people between 15 and 40 years of age and for schools for people aged between 6 and 18 years.
To simplify the model, people who attend universities are considered workers, which means that they are assigned to a workplace as any other worker.

Workplace sizes were built from the taxes office survey of Chile [24], and school size was taken from Education Ministry data [22]. In both cases, a probability distribution over the sizes was used to simulate the number of people who work or study in each establishment.

Households, workplaces and schools were placed across *Región Metropolitana* and uniformly distributed to match the number of persons that work and study in each commune. Work commuting is explained in Section 2.4

## 2.4. Commuting

*Región Metropolitana* has a particular behavior in terms of work commuting. It is normal to have persons who work at places far from their home, even if they have to travel 2 hours to reach their destinations. To model this, we separate public from private transportation. In the first case, we used travel patterns to describe the probability of a person who works in $j$ given that they live in i ($TRAVEL^{w,public}_{i,j}$) estimated from smart card data [25], considering public buses and *metro*. For the second, we used an origin-destination survey [26] to estimate private travel patterns $TRAVEL^{w,private}_{i,j}$, considering cars, motorcycles, taxis, bicycles and walking transportation. We also estimate the proportion of people who use public transport $PUBLIC$ and private transport $1 - PUBLIC$ from this survey, through which we were able to estimate a weighted sum of the proportion $TRAVEL^w_{i,j} = PUBLIC * TRAVEL^{w,public}_{i,j} + (1 - PUBLIC) * TRAVEL^{w,private}_{i,j}$
The patterns estimated allowed us to recreate how many people work in each commune and thus generate as many workplaces as necessary for people to have a place to work.

As in the case of work, school commuting is characterized by considerable movement. It is normal to have children who study in communes different from those where they live. We used data from the school choice system [27], which provides the exact coordinates of every public school and every child that studies in it. This allowed us to estimate a distribution over where a child studies given where he or she lives.

### 2.4.1. Mobility information during the pandemic

Mandatory (announced by the government) and voluntary social distancing changed people's commuting since the first case of COVID-19 was reported in *Región Metropolitana* on March 4, 2020. These changes can be seen in people who ceased going to work or school, peo-

ple who stopped visiting their relatives or even people who stopped leaving their households.

This phenomenon cannot be captured by simply turning lockdowns on and off and adding a certain effect to them because in *Región Metropolitana*, a system of partial quarantines by commune was implemented. This can produce a person who lives in a commune in lockdown who works in another without a lockdown continuing to go to work. Additionally, a person who is not in a commune in lockdown but a work commune in lockdown could stop going to work. This depends on the industry the person is associated with and his or her socioeconomic level [28].

To identify the uncertainty of lockdown compliance, we used cellphone data, which record where a person is each time they use any type of internet connection. This reveals the actual commuting of people despite their workplaces being either in quarantine or not. Figure 2.7, taken from Carranza (2020) [2], shows that mobility could be much lower in communes with a lower percentage of the low-income segment. Thus, controlling for changes in mobility only by setting or removing quarantines would introduce bias.

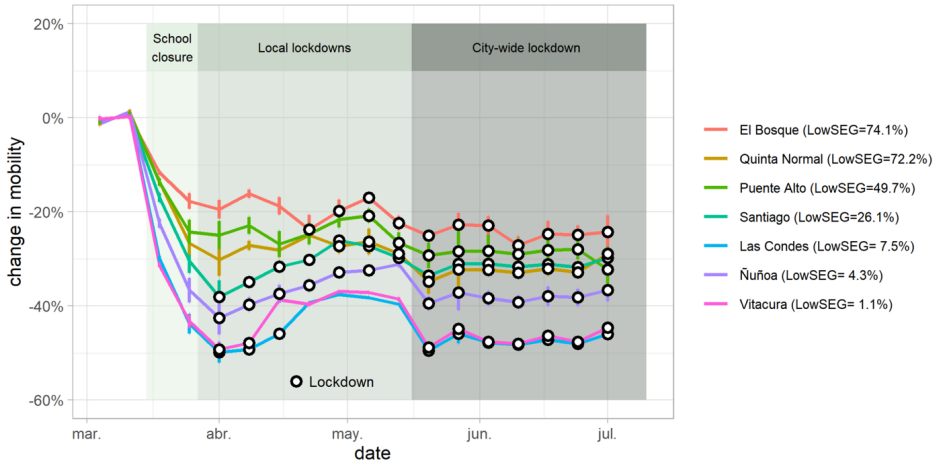



Figura 2.7: Source: Carranza (2020) [2]. Level of compliance of the people living in different communes of *Región Metropolitana*. Large circles represent when a commune was in lockdown. The acronym *lowSEG* in the legend represents the percentage of individuals who belong to a low-income segment.

Cellphone data can be understood as indicating exact coordinates and time and were used to estimate where a person lives and works. Data on overnight stays represent where a person lives, and data from the afternoon (between 13:00 and 17:00) corresponds to where the person works.

Using the previous information, the sum of all people who live in census zone i and work in census zone $j$ in week $t$ can be represented as $R_{ij}^{t}$, providing us with a matrix for every origin and destination for every week since the first week of 2020. Henceforth, we will use

$\hat{R}_{ij} = \frac{R^9_{ij} + R^{10}_{ij}}{2}$ as a tool to normalize the change in mobility, since weeks 9 and 10 correspond to the 2 weeks before mobility changed drastically.

### 2.4.2. Change in work mobility

To characterize the change in work mobility $M_c^t$ of commune $c$ in week $t$, we summed the departures from a commune and normalized them with respect to the average of departures in weeks 9 and 10.

Let $Z_c$ be the set of census zones that belong to commune $c$. The departures are given by:

$$S_c^t = \sum_{i \in Z_c} \sum_{j \neq i} R^t_{ij}$$

and change in mobility is given by:

$$M_c^t = \frac{S_c^t}{\hat{S}_c}$$

with $\hat{S}_c = \sum_{i \in Z_c} \sum_{j \neq i} \hat{R}_{ij}$

Thus, $M_i^t$ represents the proportion of people who continue going to their workplaces, while $1 - M_i^t$ remain at home, equivalent to being unemployed or working from home while the lockdown is in place. For the purposes of the model, this translates into taking a proportion $M_i^t$ of the working edges $E_W$.

### 2.4.3. Change in social commuting

For changes in social commuting, we used a different formula. Our aim here is to estimate the change in the number of people a person will meet if he or she is in a certain commune with respect to the control weeks. To achieve this, let:

- $N_c$ be the population of commune $c$
- $Pop_i$ be the population of census zone i

Thus, the proportion of the total flux that goes from census zone i to census zone $j$ is given by:

$$f^t_{ij} = \frac{R^t_{ij}}{\sum_k R^t_{ik}}$$

The probability of going to $j$ given that one lives in commune $c$ is:

$$P_{cj}^{t} = \frac{\sum_{\mathrm{i} \in Z_c \setminus \{j\}} f_{\mathrm{ij}}^{t} \cdot Pop_{\mathrm{i}}}{N_c}$$

The number of people who go to $j$ from any zone i:

$$D_j^t = \sum_{\mathrm{i}} f_{\mathrm{ij}}^{t} \cdot Pop_{\mathrm{i}}$$

The weighted sum of people who visit each commune $c$:

$$ED_c^t = \sum_j P_{cj} \cdot D_j^t$$

$ED_c^t$ represents how many people one would meet if one were to go to or remain in commune $c$.

As with work mobility, we normalized with weeks 9 and 10 to obtain the change in social commuting across weeks using $\hat{ED}_c = \frac{ED_c^9 + ED_c^{10}}{2}$.

Finally, we obtain the change in social commuting:

$$D_c^t = \frac{ED_c^t}{\hat{ED}_c}$$

Unlike changes in work mobility $M_c^t$, changes in social commuting $D_c^t$ were used to weight the contagiousness of an infected agent belonging to the corresponding commune. This is equivalent to slowing contacts of the infected agent with uninfected agents, reducing the probability of contagion.

### 2.4.4. Changes in behavior due to the presence of symptoms

Health officials mandate that individuals consult with a physician and isolate experiencing symptoms, changing individuals' behavior when they find that they are sick. To recreate this, we assume that individuals stay at home at random, depending on the severity of the illness and the importance they accord to self-care and caring for others. Some individuals cease contact because of the severity of their symptoms, while others do not change their behavior, even if they present symptoms. To achieve this, we estimated a distribution over self-isolation time, the time after symptom onset during which a person stops having contact with others (e.g., leaving home). Recognizing that self-isolation has an effect over serial intervals (the period between symptom onset of successive cases), we simulate different self-isolation times for individuals during the spread of the virus. Our objective was to maintain the same serial interval estimated in Xi et al. [12], where we find that the best approach was a self-isolation

time represented by a log-normal distribution with a mean at 2 days and mode at 1,3 days.

In summary, the changes in mobility are shown in Table 2.2. The default is 1, and changes are represented by a decrease factor belonging to $[0, 1)$. Each factor represents the weight that is multiplied by the *infection rate* (see equation 1.2). The self-isolated individuals are represented as *quarantined*.

| **State** | Normal | Lockdown |
|---|---|---|
| **Susceptible $NI$**<br>**Asymptomatic $I_a$**<br>**Contagious asymptomatic $I_a^c$**<br>**Presymptomatic $I_p$**<br>**Contagious presymptomatic $I_p^c$**<br>**Symptomatic $I_s^c$** | $h_t = 1$<br>$w_t = M_t$<br>$c_t = D_t$ | $h_t = 1$<br>$w_t = \alpha \cdot M_t$<br>$c_t = \alpha \cdot D_t$ |
| **Quarantined $I_q^c$** | $h_t = 0{,}25$<br>$w_t = 0$<br>$c_t = 0$ | $h_t = 0{,}25$<br>$w_t = 0$<br>$c_t = 0$ |
| **Hospitalized $I_h^c$**<br>**Critical $I_{cr}^c$** | $h_t = 0{,}01$<br>$w_t = 0$<br>$c_t = 0$ | $h_t = 0{,}01$<br>$w_t = 0$<br>$c_t = 0$ |
| **Postinfectious $P$** | $h_t = 0$<br>$w_t = 0$<br>$c_t = 0$ | $h_t = 0$<br>$w_t = 0$<br>$c_t = 0$ |

Tabla 2.2: Decreasing contact rates for each state of individuals, for when they are in quarantine and when they are not, for a certain week $t$. $h_t$ corresponds to contact within households, $w_t$ to contact within workplaces and $c_t$ to community contact.

# Capítulo 3

# Model estimation

To ensure that the model performs the simulations in a way that replicates reality, it is necessary to set certain conditions and calibrate some parameters.

## 3.1. Parameters to estimate

First, we defined the calibration period $\Sigma$ between March 08, 2020 and July 31, 2020. To ensure a starting point similar to reality, we randomly infect 139 nodes located in the same communes as the actual infected available on March 17.

Second, we need to calibrate different types of parameters $\theta$ belonging to the family of all parameters $\Theta$ that determine how the virus spreads in the microsimulation. Each of them has its own characteristics:

- The contagion force $\beta_p$ increase the probability of contagion within place $p = \{H, W, S, C\}$, as shown in equations 1.2 and 1.3.
- The proportion of the initial number of asymptomatic with respect to symptomatic patients $x_a$.
- The decrease in the contagion force $\eta$ from the day the total quarantine began.

Once the model is set with a group of arbitrary parameters $\theta \in \Theta$ to begin the calibration process, it iterates over every infected individual and his or her contacts to implement the infection transmission mechanism explained in Section 1.2. This simulates contagion for a certain day $t$ and repeats it $L$ times, saving these data as $y_t = (y_{t1}, y_{t2}, \dots y_{tl}, \dots y_{tL})$ and retaining the simulation that minimizes the distance between the real and simulated data. When the simulation of the calibration period is complete, we have a matrix $Y = \{y_t\}_{t \in \Sigma}$ as an output, which will be used for the selection of the best set of parameters $\theta$. This process and calibration process are developed in the next section.

## 3.2. Estimation via Maximum Likelihood

The data are represented by a vector time series $y_t = (y_{t1}, y_{t2}, \dots y_{tl}, \dots y_{tL)}$, $t = 1 \dots T$, where $l$ represents a single simulation for day $t$. We seek to build a likelihood function for these data. To facilitate exposition, consider for now $y_t$ as a random variable (not vector) representing the number of accumulated infections.

The simulation model is specified as a function:

$$h_\theta(s_0, \xi, t^*) = (s_1, \dots, s_{t^*}),$$

where $s_0$ is the initial condition, $s_t$ represents the state at time $t$ (where $t^*$ is the simulation horizon) and $\xi$ is the seed of the simulation. Starting from the same state $s_0$, multiple simulation runs $r$ can be generated with different seeds $\xi^{(r)}$, generating states $\{s_t^{(r)}\}_{t=1}^{t^*}$. The simulation model is also determined by the vector of parameters $\theta$. The function $\phi(s_t^{(r)})$ maps each simulated state to a simulated value $\hat{y}_t^{(r)}$.

The data generating process is defined by the density function $f_\theta(y_1, y_2, \dots y_T)$, where $\theta$ is the parameter vector to be estimated. The log-likelihood can be written as:

$$logL(\theta | y_1 \dots y_T) = \sum_{t=1}^{T} \log f_\theta(y_{t+1} | y_t, \dots y_1).$$

For a given value of $\theta$, the conditional density function is generated via simulation using the following algorithm.

- Initialize: $t = 1$, and build an initial set of *origin* states $\mathcal{S}_1 = \{s_1^o : \phi(s_1^o) \approx y_1\}$.
- From each origin state $s_t^o \in \mathcal{S}_t$, generate many $(R)$ one-step simulations $h_\theta(s_t^o, \xi^{(r)}, 1) = s_{t+1}^{(r)}$.
- Using all the simulated states, construct the sample $\hat{y}_{t+1}^{(r)} = \phi\left(s_{t+1}^{(r)}\right)$.
- Based on the simulated sample $\{\hat{y}_{t+1}^{(r)}\}$, use nonparametric methods (e.g., kernel) to construct an empirical density function evaluated at $y_{t+1}$ (the observed value in the data): $\hat{f}_\theta(y_{t+1} | y_t)$.
- Choose the next set of origin states $\mathcal{S}_{t+1} = \{s_{t+1}^{(r)} : \phi\left(s_{t+1}^{(r)}\right) \approx y_{t+1}\}$.
- Set $t \leftarrow t + 1$ and repeat until $t = T$.

This process is repeated for different values of $\theta$, forming a set of possible parameters $\Theta$. The MLE estimate is defined as:

$$\hat{\theta} = arg \max_{\theta \in \Theta} \sum_{t=1}^{T} \log \hat{f}_\theta(y_{t+1} | y_t)$$

Note that the empirical density is only conditional on $y_t$ and not the whole history. However, the simulation used to construct $\hat{f}$ is based on a path of states $s_1 \dots s_t$ that attained y-values at all the observed values $y_1 \dots y_t$, thereby implicitly conditioning on the whole vector.

## 3.3. Estimation results

In the first instance, we calibrated the model using data from March 15 until April 6. We used cumulative infection according to confirmation day, which was the only information available to that date. For this process, the estimated $R_0$ was approximately 1,6 with schools closed and social distancing, while $R_0$ was approximately 2,75 with no restrictions. This calibration was used later to predict the future spread of the virus under different scenarios (see Section 4.2).

After all the changes made to the contagion curve, to avoid underreporting biases and to work with the contagion date, a more sophisticated calibration process was implemented, as shown in Section 3.2.

For every day, we simulated $L = 1000$ states, given the last day's state. Thus, we calculated medians to compare with the real data and calculated the likelihood contribution for each day. We ran the model with several sets of parameters to achieve the best fit. As we can see in Table 3.1, the model is very sensitive to varying any of the parameters, not following a particular pattern.

Figure 3.1 shows the performance of the best parameters, and we can see that the model fits quite well except between May 05 and May 26, matching the period when the positivity of the test and reporting delay were the highest (see Section 2.2). The existing break at the peak of the curve is due to the influence of a city-wide lockdown, represented by $\nu$. At this point, infections began to decrease quickly, which is overillustrated when we take the medians.

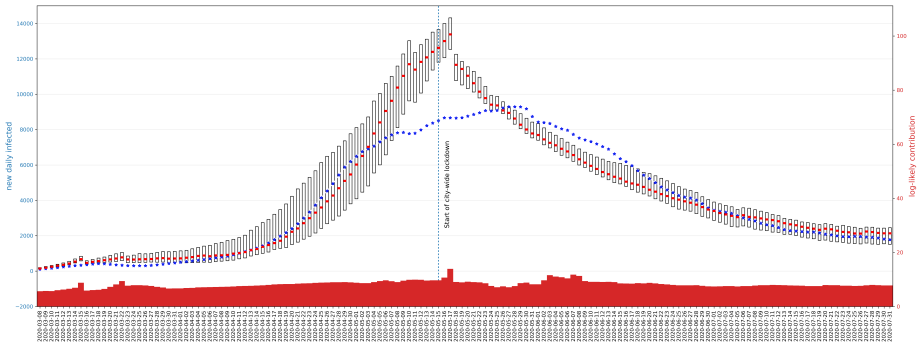



Figura 3.1: Prediction of the daily contacts simulated with the microsimulation compared with the contagion curve adjusted for underreporting. Each box represents 25 and 75 percentiles of data simulated for the corresponding day. Red squares are the median of the samples for each day, and blue stars are daily infections according to contagion day corrected for underreporting. The parameters selected were $\beta_H = 11{,}75$, $\beta_W = 14{,}1$, and $\beta_S = 0{,}94$ , $\beta_C = 3{,}375$, $\eta = 0{,}9$, $x_a = 0{,}45$.

| $x_a$ | $\beta_c$ | $\beta_h$ | $\beta_w$ | $\nu$ | $LL$ |
|---|---|---|---|---|---|
| 0.45 | 45.0 | 25.0 | 30.0 | 0.9 | -1195.97 |
| 0.45 | 45.0 | 20.0 | 30.0 | 0.9 | -1208.46 |
| 0.45 | 40.0 | 20.0 | 30.0 | 1.0 | -1216.9 |
| 0.45 | 35.0 | 15.0 | 35.0 | 1.0 | -1219.58 |
| 0.45 | 40.0 | 25.0 | 30.0 | 0.9 | -1238.02 |
| 0.3 | 40.0 | 10.0 | 35.0 | 1.0 | -1254.56 |
| 0.45 | 30.0 | 10.0 | 40.0 | 1.0 | -1267.25 |
| 0.45 | 35.0 | 20.0 | 35.0 | 0.9 | -1281.72 |
| 0.3 | 30.0 | 10.0 | 40.0 | 1.0 | -1284.71 |
| 0.3 | 30.0 | 10.0 | 40.0 | 1.0 | -1284.71 |
| 0.45 | 45.0 | 20.0 | 30.0 | 1.0 | -1295.76 |
| 0.3 | 30.0 | 10.0 | 35.0 | 1.0 | -1302.06 |
| 0.45 | 10.0 | 30.0 | 40.0 | 1.0 | -1326.65 |
| 0.05 | 30.0 | 20.0 | 50.0 | 1.0 | -1334.33 |
| 0.05 | 30.0 | 20.0 | 50.0 | 1.0 | -1334.33 |
| 0.45 | 10.0 | 10.0 | 50.0 | 1.0 | -1348.75 |
| 0.05 | 15.0 | 15.0 | 55.0 | 1.0 | -1352.7 |
| 0.05 | 30.0 | 40.0 | 40.0 | 1.0 | -1353.97 |
| 0.05 | 30.0 | 40.0 | 40.0 | 1.0 | -1353.97 |
| 0.45 | 35.0 | 20.0 | 35.0 | 1.0 | -1358.98 |
| 0.3 | 30.0 | 20.0 | 35.0 | 1.0 | -1368.3 |
| 0.1 | 20.0 | 10.0 | 50.0 | 1.0 | -1374.23 |
| 0.05 | 10.0 | 15.0 | 55.0 | 1.0 | -1377.66 |
| 0.05 | 40.0 | 20.0 | 40.0 | 1.0 | -1377.85 |
| 0.05 | 40.0 | 20.0 | 40.0 | 1.0 | -1377.85 |
| 0.05 | 60.0 | 20.0 | 40.0 | 1.0 | -1377.88 |
| 0.05 | 60.0 | 20.0 | 40.0 | 1.0 | -1377.88 |
| 0.05 | 20.0 | 30.0 | 50.0 | 1.0 | -1378.23 |
| 0.05 | 20.0 | 30.0 | 50.0 | 1.0 | -1378.23 |
| 0.05 | 30.0 | 30.0 | 50.0 | 1.0 | -1380.86 |
| 0.05 | 30.0 | 30.0 | 50.0 | 1.0 | -1380.86 |
| 0.45 | 30.0 | 40.0 | 30.0 | 1.0 | -1385.85 |
| 0.05 | 20.0 | 20.0 | 50.0 | 1.0 | -1388.74 |
| 0.05 | 20.0 | 20.0 | 50.0 | 1.0 | -1388.74 |
| 0.45 | 10.0 | 20.0 | 50.0 | 1.0 | -1390.86 |
| 0.05 | 60.0 | 30.0 | 40.0 | 1.0 | -1391.02 |
| 0.05 | 60.0 | 30.0 | 40.0 | 1.0 | -1391.02 |
| 0.05 | 20.0 | 50.0 | 50.0 | 1.0 | -1391.06 |
| 0.05 | 20.0 | 50.0 | 50.0 | 1.0 | -1391.06 |
| 0.05 | 30.0 | 40.0 | 50.0 | 1.0 | -1393.75 |
| 0.05 | 30.0 | 40.0 | 50.0 | 1.0 | -1393.75 |
| 0.45 | 10.0 | 40.0 | 40.0 | 1.0 | -1395.81 |
| 0.05 | 30.0 | 30.0 | 40.0 | 1.0 | -1398.93 |
| 0.05 | 30.0 | 30.0 | 40.0 | 1.0 | -1398.93 |
| 0.05 | 15.0 | 10.0 | 55.0 | 1.0 | -1402.14 |
| 0.05 | 50.0 | 20.0 | 30.0 | 1.0 | -1404.6 |
| 0.05 | 50.0 | 20.0 | 30.0 | 1.0 | -1404.6 |
| 0.05 | 60.0 | 50.0 | 40.0 | 1.0 | -1413.11 |
| 0.05 | 60.0 | 50.0 | 40.0 | 1.0 | -1413.11 |
| 0.05 | 30.0 | 20.0 | 40.0 | 1.0 | -1415.8 |

Tabla 3.1: List of the best 50 likelihood-scored parameters evaluated by the model.

# Capítulo 4

# Evaluating mitigation strategies

Having calibrated the model, we were able to predict how the future of the pandemic would change given certain nonpharmaceutical mitigation strategies. According to the *Centers of Disease Control and Prevention*, nonpharmaceutical interventions (NPIs) are actions, apart from being vaccinated and taking medicine, that people and communities can take to help slow the spread of illnesses such as COVID-19 [29]. The NPI applied by default was social distancing (according to the approach established in Ferguson (2020) [1]) and schools closed, given that those were suggested by the government in the early stage of the pandemic. The NPIs evaluated in this work are full lockdown, lockdowns with a threshold policy and contact tracing and selective case isolation.

## 4.1. Full lockdown strategy

At the beginning of the pandemic, there was resistance to establishing quarantines in *Región Metropolitana*. Therefore, a question to answer was: what would be the impact of imposing a total quarantine? According to our first calibration and its predictions (see Figure 4.1, the virus would get out of control if the city had maintained the status quo, with more than 10,000 new daily cases in June. On the other hand, under a full-lockdown policy, it would have been possible to combat the spread of the virus by June. This was achieved by reducing contact at the same rate as done in Ferguson (2020) when people followed voluntary home quarantine.

## 4.2. Lockdowns with threshold policy

As the government was not willing to apply a total quarantine strategy due to the negative impact it could have had on the economy, we evaluated a less strict measure that consisted of a system of intermittent quarantines. The *Región Metropolitana* has 6 health services, which consist of the hospitals belonging to the area of each service. This characteristic offers the potential of working in a coordinated manner in terms of capacity and medical personnel. Thus, we aggregate the communes belonging to a certain health service to use a threshold policy. Once the communes belonging to a health service exceed 50 daily infections per 100,000

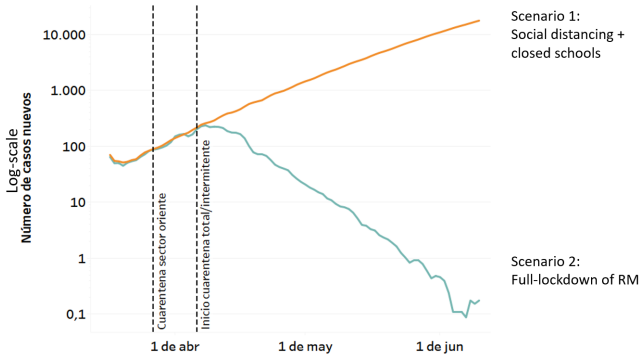



Figura 4.1: Evolution of the new daily cases according to reporting day for the first calibration. The yellow line represents social distancing assumptions and schools closed. The blue line represents full lockdown of the region.

inhabitants, they will go into quarantine. To ensure the efficacy of the policy, it would be necessary to have an approximate number of 2.5 million people quarantined at the same time across the communes in lockdown.

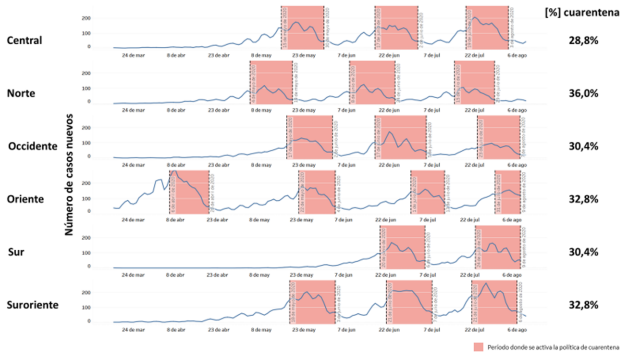



Figura 4.2: Daily infections in *Región Metropolitana* grouped by health services. The pink area represents the period of time in which the population associated with a health service is in lockdown.

Figure 4.2 shows daily infections over time grouped by health service and the percentage of time that they would be in quarantine following this strategy. With this strategy, the pandemic was under control, approximating a reproduction number of $R_t = 1$. We also can see that, on average, *Región Metropolitana* would be in lockdown approximately 30 % of the time, avoiding health systems collapsing and maintaining a low impact on the economy.

## 4.3. Contact tracing and selective case isolation

In addition to the strategies evaluated above, as suggested by Ferretti (2020) [15], a key measure to curb the spread of the virus is contact tracing.
*How does Chilean contact tracing work?*
Once a person is declared infected, by confirmation of a *PCR* test, the contact tracing process is triggered. Every contact of the individual is notified to take a test. This lasts an average of

2 days until the results are ready. If a test is positive, that contact will remain quarantined until he or she is healthy, while in parallel, the process starts again, but now the focus of the monitoring is the contact detected. We implement this mechanism in the model from August 1, following the behavior of the Chilean government. Despite contact tracing beginning earlier, an interview with Health Ministry personnel revealed that a robust strategy of contact tracing was developed in mid-July[30].

To provide flexibility to the model, there are two parameters to control the effectiveness of contact tracing:

1. $\alpha$-tracing: the proportion of contacts of an infected individual who are traced
2. $\beta$-tracing: the proportion of infected presenting symptoms who decide to take a *PCR* test and, therefore, trigger the contact tracing process.

First, we can see in Figure 4.3 that if we set $\alpha$-tracing $= 0$ and $\beta$-tracing $= 0$, the pandemic cannot be controlled. This is equivalent to performing no contact tracing.

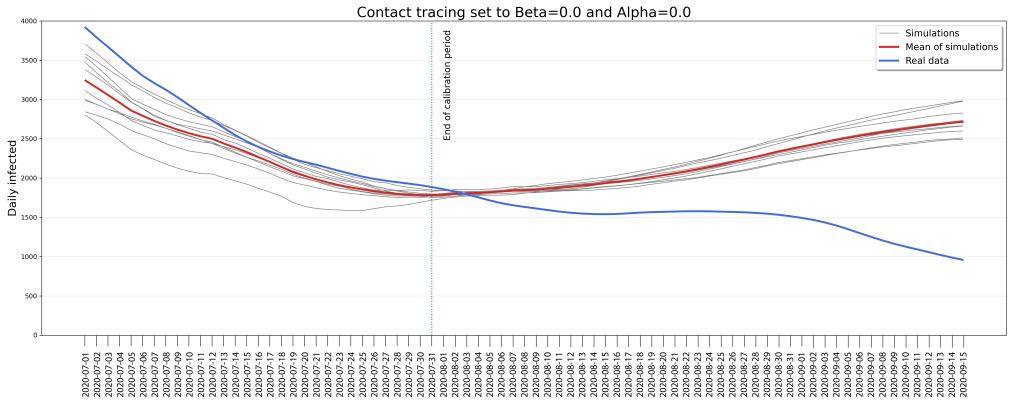



Figura 4.3: Daily infection curve obtained by not applying contact tracing.

We varied $\alpha$-tracing and $\beta$-tracing $\in \{0{,}1, 0{,}2, 0{,}3, 0{,}4, 0{,}5\}$ to estimate different outcomes. We can see in Figure 4.4 that real data can be fitted by setting ($\alpha$-tracing, $\beta$-tracing) $=$ $(0{,}1, 0{,}2)$ or ($\alpha$-tracing, $\beta$-tracing) $= (0{,}2, 0{,}1)$. This is because only 20 % (or 10 %) of infected people trigger the process, and only 10 % (or 20 %) of the infected contacts are traced. Those are low levels of contact tracing, presenting considerable room for improvement. Furthermore, higher values of $\alpha$ and $\beta$ lead to substantial reductions in the new cases.

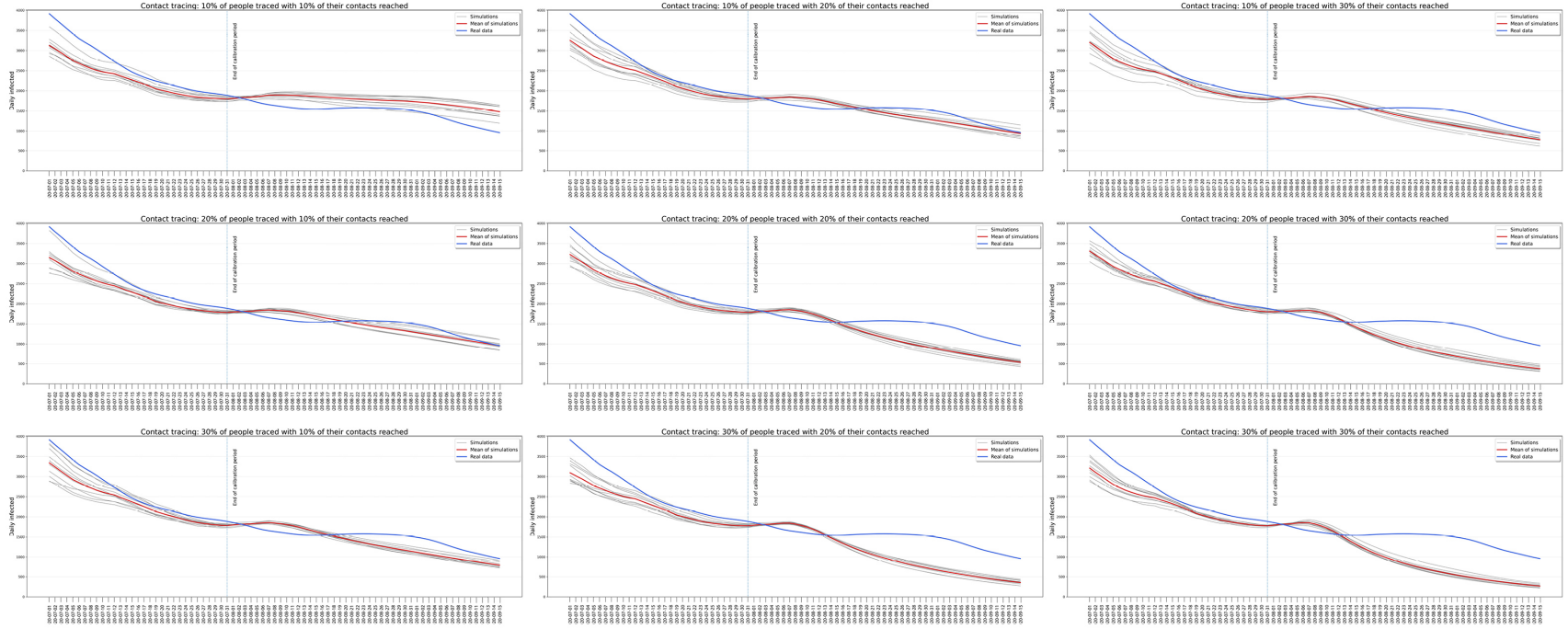



Figura 4.4: Course of the pandemic when applying different levels of contact tracing. $\beta$ denotes the proportion of people who trigger the tracing process once they self-isolate (equivalent to the percentages of people who decide to take a $PCR$ test) and $\alpha$ the proportion of contacts from the infected source who will be reached.

# Conclusión

First, because new infection data were biased by reporting delays and high positivity levels, it was necessary to adjust the data by using the contagion day and correcting for underreporting. This is relevant because using biased data would have led to incorrect conclusions, such as underestimating the contagion force or delaying its effect. As long as governments do not achieve reliable data tracking, supported by indicators such as low positivity and an absence of lags in reporting, there is a possibility of underreporting in the new daily infection curve. Here, it was key to take into account that in Wuhan, China, during the first days of the pandemic, there was a good record of the virus spread, which allows us to use it as a baseline for the actual CFR and compare it to ours.

On the other hand, it was also important to include mobility data from cellphones to measure changes in the behavior of the population. The lockdown policy implemented in *Región Metropolitana* hindered the measurement of its effects, since there were many people living in communes that were in quarantine but working in communes that were not (and vice versa). That is why mobility data present a good alternative, providing individual information on the change in commuting of every citizen and, thus, enabling the measurement of the effects of lockdowns throughout the pandemic.

A new and novel technique was implemented to measure the goodness of fitting the model. The model runs many simulations per day, providing a large amount of data over time. This led to the idea of estimating via maximum likelihood for each day, and the total contribution is given by the sum of likelihoods. This is different from distance-based approaches (such as the Euclidean norm), and this was our aim. Considering this is a simulation, comparing the mean or median with real data does not account for the dispersion of the simulated data. On the other hand, using the maximum likelihood estimator computes how likely it is that we will obtain the real data in our simulation, which does account for dispersion.

Full lockdown is an effective alternative according to the model. Our estimations show that if a full lockdown regime had been mandatory for all individuals living in *Región Metropolitana* and they complied responsibly, the spread of the virus could have been slowed and controlled over time. As of May 15, a total quarantine regime was established in *Región Metropolitana* but was not as effective as our model suggests. Mobility data used to measure the real effect of quarantine showed that the level of compliance by people was between 20 %

and 50 % depending on the income level of individuals living there. This could have been due to many reasons, but it is useful as evidence that lower levels of mobility were necessary to accomplish the goal of combating COVID-19.

Lockdowns with threshold policies were proven to be an effective alternative to fight SARS-CoV-2 according to the model. In addition, having 70 % of the population not quarantined allows economic activity to not be as affected as in the case of full lockdown. This policy must be applied to at least 2.5 million people to maintain a reproduction number $R_0$ close to 1. This measure is insufficient to eradicate the virus but is an alternative to ensure that health services will not collapse. Unfortunately, this strategy was only partially realized in reality. The government implemented partial quarantines by communes but not grouped by health services and also failed to achieve the quantity of people needed to control the pandemic.

Considering that full lockdown was an overly strong measure for *Región Metropolitana* in terms of people's compliance and that the threshold policy does not eradicate the virus but maintains it at a stable level, contact tracing represents as a feasible option. We could see that low levels of contact tracing (less than 20 % of confirmed infected traced and less than 20 % of their contacts called) emulate reality, and any higher level would curb the spread radically. This measure is subject to the quality of the contact tracing system and the effectiveness on people willing to take a PCR test after being traced. We recommend strengthening the contact tracing system to achieve more people being traced and more of their contacts being reached.

# Bibliografía


[1] N Ferguson, D Laydon, G Nedjati Gilani, N Imai, K Ainslie, M Baguelin, S Bhatia, A Boonyasiri, ZULMA Cucunuba Perez, G Cuomo-Dannenburg, A Dighe, I Dorigatti, H Fu, K Gaythorpe, W Green, A Hamlet, W Hinsley, L Okell, S Van Elsland, H Thompson, R Verity, E Volz, H Wang, Y Wang, P Walker, P Winskill, C Whittaker, C Donnelly, S Riley, and A Ghani. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Technical report, Imperial College London, March 2020.

[2] Aldo Carranza, Marcel Goic, Eduardo Lara, Marcelo Olivares, Gabriel Y. Weintraub, Julio Covarrubia, Cristian Escobedo, Natalia Jara, and Leonardo J. Basso. The Social Divide of Social Distancing: Lockdowns in Santiago during the COVID-19 Pandemic. SSRN Scholarly Paper ID 3691373, Social Science Research Network, Rochester, NY, September 2020.

[3] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy S.M. Leung, Eric H.Y. Lau, Jessica Y. Wong, Xuesen Xing, Nijuan Xiang, Yang Wu, Chao Li, Qi Chen, Dan Li, Tian Liu, Jing Zhao, Man Liu, Wenxiao Tu, Chuding Chen, Lianmei Jin, Rui Yang, Qi Wang, Suhua Zhou, Rui Wang, Hui Liu, Yinbo Luo, Yuan Liu, Ge Shao, Huan Li, Zhongfa Tao, Yang Yang, Zhiqiang Deng, Boxi Liu, Zhitao Ma, Yanping Zhang, Guoqing Shi, Tommy T.Y. Lam, Joseph T. Wu, George F. Gao, Benjamin J. Cowling, Bo Yang, Gabriel M. Leung, and Zijian Feng. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, March 2020. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa2001316.

[4] Joseph T. Wu, Kathy Leung, and Gabriel M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225):689–697, February 2020. Publisher: Elsevier.

[5] Samuel Lalmuanawma, Jamal Hussain, and Lalrinfela Chhakchhuak. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139:110059, October 2020.

[6] Giuseppe C. Calafiore, Carlo Novara, and Corrado Possieri. A Modified SIR Model for the COVID-19 Contagion in Italy. *arXiv:2003.14391 [physics]*, March 2020. arXiv: 2003.14391.

[7] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu. A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons. *IEEE Transactions on Network Science and Engineering*, pages 1–1, 2020. Conference Name: IEEE Transactions on Network Science and Engineering.

[8] S. Rica and G. A. Ruz. Estimating SIR model parameters from data using differential evolution: an application with COVID-19 data. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–6, October 2020.

[9] Sheryl L. Chang, Nathan Harding, Cameron Zachreson, Oliver M. Cliff, and Mikhail Prokopenko. Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature Communications*, 11(1):5710, November 2020. Number: 1 Publisher: Nature Publishing Group.

[10] Mohammad Akbarpour, Cody Cook, Aude Marzuoli, Simon Mongey, Abhishek Nagaraj, Matteo Saccarola, Pietro Tebaldi, Shoshana Vasserman, and Hanbin Yang. Socioeconomic Network Heterogeneity and Pandemic Policy Response. SSRN Scholarly Paper ID 3623111, Social Science Research Network, Rochester, NY, June 2020.

[11] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, May 2020.

[12] Xi He, Eric H. Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yujuan Guan, Xinghua Tan, Xiaoneng Mo, Yanqing Chen, Baolin Liao, Weilie Chen, Fengyu Hu, Qing Zhang, Mingqiu Zhong, Yanrong Wu, Lingzhai Zhao, Fuchun Zhang, Benjamin J. Cowling, Fang Li, and Gabriel M. Leung. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5):672–675, May 2020.

[13] Neil M. Ferguson, Derek A.T. Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S. Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, September 2005.

[14] Zian Zhuang, Shi Zhao, Qianying Lin, Peihua Cao, Yijun Lou, Lin Yang, Shu Yang, Daihai He, and Li Xiao. Preliminary estimating the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020. *medRxiv*, page 2020.03.02.20030312, March 2020. Publisher: Cold Spring Harbor Laboratory Press.

[15] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, May 2020.

[16] Le coronavirus expliqué graphiquement - Infographie de la contagion, March 2020.

[17] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick Walker, Han Fu, Amy Dighe, Jamie Griffin, Anne Cori, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma M Cucunuba, Rich Fitzjohn, Katy A M Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsand, Erik Volz, Haowei Wang, Yuanrong Wang, Xiayoue Xi, Christl Donnelly, Azra Ghani, and Neil Ferguson. Estimates of the severity of COVID-19 disease. preprint, Epidemiology, March 2020.

[18] Timothy W Russell, Nick Golding, Joel Hellewell, Sam Abbott, Lawrence Wright, Carl A B Pearson, Kevin van Zandvoort, Christopher I Jarvis, Hamish Gibbs, Yang Liu, Rosalind M Eggo, John W Edmunds, and Adam J Kucharski. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. preprint, Epidemiology, July 2020.

[19] Eunha Shim, Kenji Mizumoto, Wongyeong Choi, and Gerardo Chowell. Estimating the risk of COVID-19 death during the course of the outbreak in Korea, February- May, 2020. preprint, Epidemiology, April 2020.

[20] ICOVID Chile.

[21] WEB DISEMINACIÓN CENSO 2017 WEB DISEMINACIÓN CENSO 2017.

[22] Descarga Bases de datos de matrícula por estudiante · MINEDUC. Library Catalog: datos.mineduc.cl.

[23] Descarga Bases de Datos Matrícula en Educación Superior · MINEDUC. Library Catalog: datos.mineduc.cl.

[24] SII | Estadísticas de Empresas por Tramo según ventas (13 tramos) y Comuna.

[25] Margarita Amaya, Ramón Cruzat, and Marcela A. Munizaga. Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis. *Journal of Transport Geography*, 66:330–339, January 2018.

[26] Encuesta de Origen y Destino de Viajes Santiago 2012 - Portal de Datos Abiertos.

[27] Jose Correa, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, and Boris Epstein. School Choice in Chile. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 325–343, Phoenix AZ USA, June 2019. ACM.

[28] Movilidad en Región Metropolitana bajó sólo 35 por ciento promedio durante el primer mes de cuarentena masiva - Universidad de Chile.

[29] CDC. CDC Works 24/7, November 2020.

[30] El Mostrador. El imputado Mañalich: exministro declara por investigación judicial por muertes de Covid-19 en medio de acusaciones de manejo de cifras, November 2020. Section: País.