



HAL
open science

Modeling and Control of Epidemics through Testing Policies

Muhammad Umar B. Niazi, Alain Kibangou, Carlos Canudas de Wit, Denis Nikitin, Liudmila Tumash, Pierre-Alexandre Bliman

► **To cite this version:**

Muhammad Umar B. Niazi, Alain Kibangou, Carlos Canudas de Wit, Denis Nikitin, Liudmila Tumash, et al.. Modeling and Control of Epidemics through Testing Policies. *Annual Reviews in Control*, 2021, 52, pp.554-572. 10.1016/j.arcontrol.2021.09.004 . hal-02986566v2

HAL Id: hal-02986566

<https://hal.science/hal-02986566v2>

Submitted on 26 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling and Control of Epidemics through Testing Policies^{*}

Muhammad Umar B. Niazi^{a,*}, Alain Kibangou^a, Carlos Canudas-de-Wit^a,
Denis Nikitin^a, Liudmila Tumash^a, Pierre-Alexandre Bliman^b

^a*Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France.*

^b*Sorbonne Université, Université Paris-Diderot SPC, Inria, CNRS, Laboratoire Jacques-Louis Lions, équipe Mamba, 75005 Paris, France.*

Abstract

Testing is a crucial control mechanism in the beginning phase of an epidemic when the vaccines are not yet available. It enables the public health authority to detect and isolate the infected cases from the population, thereby limiting the disease transmission to susceptible people. However, despite the significance of testing in epidemic control, the recent literature on the subject lacks a control-theoretic perspective. In this paper, an epidemic model is proposed that incorporates the testing rate as a control input and differentiates the undetected infected from the detected infected cases, who are assumed to be removed from the disease spreading process in the population. After estimating the model on the data corresponding to the beginning phase of COVID-19 in France, two testing policies are proposed: the so-called best-effort strategy for testing (BEST) and constant optimal strategy for testing (COST). The BEST policy is a suppression strategy that provides a minimum testing rate that stops the growth of the epidemic when implemented. The COST policy, on the other hand, is a mitigation strategy that provides an optimal value of testing rate minimizing the peak value of the infected population when the total stockpile of tests is limited. Both testing policies are evaluated by their impact on the number of active intensive care unit (ICU) cases and the cumulative number of deaths for the COVID-19 case of France.

Keywords: Epidemic modeling, Testing, Model estimation, Control policies.

^{*}This work is supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERCAdG no. 694209, Scale-FreeBack, website: <http://scale-freeback.eu/>) and by Inria, France, in the framework of the Inria's Mission COVID-19.

^{*}Corresponding author

Email addresses: muhammad-umar-b.niazi@inria.fr (Muhammad Umar B. Niazi),
alain.kibangou@univ-grenoble-alpes.fr (Alain Kibangou),
carlos.canudas-de-wit@gipsa-lab.fr (Carlos Canudas-de-Wit),
denis.nikitin@gipsa-lab.fr (Denis Nikitin), liudmila.tumash@gipsa-lab.fr (Liudmila Tumash), pierre-alexandre.bliman@inria.fr (Pierre-Alexandre Bliman)

1. Introduction

The history of humanity is enameled with various pandemics whose consequences have had a durable impact on our societies. In almost all cases, vaccination is presented as a panacea, however, before vaccination becomes possible, an initial response, as effective as possible, must be provided. This is what the COVID-19 epidemic, started in Wuhan, China, at the end of 2019, has taught us. COVID-19 was declared to be a pandemic by the World Health Organization (WHO) on March 11, 2020. The most common symptoms of the disease include fever, cough, fatigue, shortness of breath, and loss of the senses of smell and taste, where complications may include pneumonia and respiratory distress known as a severe acute respiratory syndrome (SARS). For more than a year, the primary mode of treatment had been symptomatic and supportive therapy [1, 2], and no approved vaccine or specific antiviral treatment was available.

The pandemic shook the economy of the whole world with a significant reduction of exports, a decline in tourism, mass unemployment, and business closures [3]. Governments and health authorities worldwide responded by implementing non-pharmaceutical intervention (NPI) policies such as travel restrictions, lockdown measures, social distancing, workplace hazard controls, closure of schools and workplaces, curfew strategies, and cancellation of public events. Many countries also upgraded existing infrastructure and personnel to increase testing capabilities and facilities for focused isolation. The public was instructed to wash hands several times a day, cover mouth and nose when coughing or sneezing, maintain a certain physical distance from other people, wear a face mask in public places/gatherings, and to monitor and self-isolate if the disease symptoms appear. The extent to which such policies and measures have been implemented is called the stringency index of a country's response to the epidemic [4, 5]. Each country responded in its capacity to find a suitable balance between saving lives and saving livelihoods, which [6] termed as a problem of health versus wealth. Livelihoods can be saved through the implementation of suitable relief and recovery measures for people and small businesses. On the other hand, lives can be saved through the implementation of testing policies and NPIs. In other words, there is a direct relation between the stringency index of the government and 'saving lives'.

All the above strategies and policies are considered to be the control mechanisms for the epidemic. Such strategies fall under two categories: mitigation and suppression [7, 8]. The mitigation strategies slow down the rate of transmission of disease or 'flatten the curve'. However, they do not necessarily stop the spreading of the disease, which is the goal of suppression strategies. Given the required objectives (for e.g., minimizing the number of deaths caused by the epidemic) and constraints (for e.g., socio-economic costs), the problem of finding optimal strategies for epidemic control has been recently studied under the framework of optimal control theory.

1.1. Literature review

To understand, predict, and control the evolution of the COVID-19 epidemic, a huge effort has been devoted by the researchers to design models as accurate and as effective as possible. Each model, by and large, is a variant and/or an extension of SIR (susceptible, infected, recovered) and SEIR (susceptible, exposed, infected, recovered) models, which describe the flow of population through three or four mutually exclusive stages of infection, respectively (see [9] and [10] for a comprehensive review). These basic models have few parameters that are easy to identify [11], and are considered as population models that view the epidemic from the macroscopic perspective. This is in contrast with the approaches that capture heterogeneity of population structure such as network epidemic models [12, 13, 14] or metapopulation epidemic models [15, 16, 17], that view the epidemic from the microscopic perspective. In what follows, however, we study the epidemic through the macroscopic perspective of population models.

Following the outbreak of COVID-19, there has been an effort to produce comprehensive population models with a focus on different facets of the epidemic. Such models are more complex than simple SIR and SEIR, and include several intermediate stages that accurately portray the dynamics of the epidemic. For instance, [18] develops an extension of the SEIR model that incorporates the governmental actions (e.g., preventive measures and restrictions) and the individual behavioral reactions, whereas [19] develops an extension of the SIR model that incorporates the number of deaths due to the epidemic. Another quite interesting model is the one developed in [20] that considers an eight-compartment model called SIDARTHE, which includes eight stages of infection: susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H), and extinct (E). A distinguishing feature of this model is that it differentiates between the infected individuals based on the severity of their symptoms and whether they are diagnosed by a health authority. It is crucial, as also emphasized in [21, 22], to differentiate between diagnosed and undiagnosed individuals because the former are typically isolated and are less likely to spread the infection. Similar models have been adopted and extended to study optimal control policies for the epidemic such as the implementation of social distancing measures [23, 24, 25], lockdown strategies [26, 27, 28], and heterogeneous policy responses based on age-groups [29, 30].

In addition to the above NPI strategies, testing and isolating the infected population from the susceptible population is one of the most important strategies to control the epidemic spread. For instance, it has been reported that COVID-19 was eliminated from the Italian village Vo'Euganeo through testing both symptomatic and asymptomatic cases [31, 32]. Moreover, on his media briefing¹ of March 16, 2020, Dr. Tedros Adhanom Ghebreyesus, the Director-General of WHO, urged the following:

“Social distancing measures can help to reduce transmission and en-

¹Website: WHO Director-General Speech of March 16, 2020. (Accessed 04/06/2020)

able health systems to cope. Hand-washing and coughing into your elbow can reduce the risk for yourself and others. But on their own, they are not enough to extinguish this pandemic. It's the combination that makes the difference. As I keep saying, all countries must take a comprehensive approach. But the most effective way to prevent infections and save lives is breaking the chains of transmission. And to do that, you must test and isolate. You cannot fight a fire blindfolded. And we cannot stop this pandemic if we don't know who is infected. We have a simple message for all countries: TEST, TEST, TEST."

COVID-19 can be detected through two types of tests known as type-1 (RT-PCR) and type-2 (serology). In the type-1 test, a swab is inserted into the subject's nose to qualitatively detect nucleic acid from SARS-CoV-2 in the upper and lower respiratory specimens [33], which enables one to detect whether the subject is currently infected with COVID-19. Type-2 test, on the other hand, is a serum test that detects relevant antibodies enabling one to know whether the subject was infected in the past with COVID-19. Both types of tests are important in the control of an epidemic. Type-1 tests help to limit the disease spread by the identification of infected individuals and their contact tracing [34]. Type-2 tests, on the other hand, are useful in reducing the size of the testable population for type-1 tests [35] that helps to increase the testing specificity. However, the type-1 test was considered to be the only recommended method for the identification and laboratory confirmation of COVID-19 cases according to the WHO [36]. Moreover, only type-1 tests can provide information in real-time related to describe the outburst of the epidemic, which is the reason that the datasets related to testing only include type-1 tests².

Following the recommendation of the WHO director, with different levels of setups, many countries increased their testing capacities, while others feared the economic burden of intensive testing policy. However, [37, 38] show that such a burden is only short-term and, on the contrary, intensive testing reduces the 'overall' cost of the epidemic in the long run because it enables the government to gain rapid control of the epidemic and revive the economy of a country. Testing enables the health authority to identify and isolate the infected people from the susceptible population, which limits the transmission of the disease. Therefore, testing is considered to be a crucial control mechanism for the epidemic [39]. However, few attempts have been dedicated to study the testing policy for an epidemic from a control-theoretic perspective.

In somewhat similar to a resource allocation problem [40, 41] in epidemic control, [42] poses the optimal test allocation as a well-known sensor selection problem in control theory, whereas [43] poses it as a welfare maximization problem by considering specificity and sensitivity of tests. The main assumption in these papers, however, is the availability of information portfolios of all individ-

²Website: Our World in Data: Coronavirus (COVID-19) Testing. (Accessed 30/09/2020)

uals in a society, which enables the decision makers to compute the infection probability of individuals and utility loss for each individual in case of decision errors. On the other hand, [44, 45] study the problem of testing policy from an economic perspective, where the goal is to find an optimal testing policy that minimizes the total number of quarantined people to incur minimal cost on the economic activity of a country while also mitigating the epidemic spread. Without such a policy, health authorities usually resort to indiscriminate quarantining of people that burdens the economy of a country without any reason. Therefore, testing allows to identify and isolate the positive cases for case-dependent quarantining. Another aspect of testing policy is studied in [46], which computes an optimal trade-off between testing effort and lockdown intervention under the constraint of limited Intensive Care Units (ICU).

1.2. *Our contribution*

At the onset of an epidemic, the effective treatments or vaccines are usually not available. In such a case, it is important to devise effective testing policies for epidemic control. We introduce a modified SIR model named SIDUR — susceptible (S), undiagnosed infected (I), diagnosed infected (D), unidentified recovered (U), and identified removed (R) — to study epidemic control through testing. Similar to [20, 21, 22], we differentiate between the undiagnosed and diagnosed infected population. We assume that the diagnosed infected population are either quarantined and/or hospitalised and only the undiagnosed infected population is responsible for the disease transmission to the susceptible population. The identified removed population consists of people who recover or die after being diagnosed and the unidentified recovered population consists of people who recover without getting diagnosed. The control input in the SIDUR model is the testing rate defined as the number of tests performed per day, where the influence of the control is directly linked with the testing specificity. The testing specificity determines the probability of detecting an undiagnosed infected person through a test, which, for instance, can be increased through efficient contact tracing.

First, we estimate the parameters of SIDUR model using the COVID-19 data of France. Then, we propose two testing policies for epidemic control: 1) Best effort strategy for testing (BEST) and 2) Constant optimal strategy for testing (COST).

The best-effort strategy for testing (BEST) is a suppression strategy for an epidemic that provides the minimum number of tests to be performed per day in order to stop the epidemic spread. Thus, BEST is meaningful only during the spreading phase of the disease. We provide an algorithm to compute the number of tests required by BEST policy. Since BEST is a suppression strategy that stops the epidemic growth immediately, it usually requires a large number of tests to be performed per day. However, if implemented earlier, BEST requires feasible number of tests. We illustrate this for the COVID-19 case of France by plotting the number of tests required by BEST with respect to time.

The constant optimal strategy for testing (COST) is a mitigation strategy when the total stockpile of tests is limited. It provides the optimal number of

tests per day that must be allocated in a certain time interval from the onset of the epidemic in order to minimize the peak of infected population. When the stockpile finishes at the terminal time, then the tests are not performed anymore. The main idea of COST is the consideration of two peaks of the epidemic; one that occurs before the stockpile of tests finishes and one that occurs after the stockpile finishes. If the stockpile of tests is allocated constantly such that it reduces the first peak, it will however result in the increase of second peak, and vice versa. Thus, the optimal allocation of tests per day is the one that minimizes both peaks of the infected population, which occurs when both peaks are equal. Both BEST and COST policies are compared with the actual COVID-19 testing scenario of France through the prediction of the number of active ICU cases and the cumulative number of deaths.

1.3. Paper organization

In Section 2, we describe the model, inputs, and outputs, and compute the basic and effective reproduction numbers of the model. Section 3 illustrates the French COVID-19 data and provides the data imputations to infer the missing data from the raw data. Then, Section 4 provides the estimation and fitting of model for the French COVID-19 case. Finally, in Section 5, we propose two testing policies BEST and COST, and evaluate them by comparing the predicted number of active ICU cases and the predicted cumulative number of deaths with the actual data.

2. SIDUR model with testing policy

We consider a five-compartment model with the purpose of evaluating and devising testing policies. We assume that testing allows for diagnosing and isolating the infected people from the population to prevent the transmission of the disease to the susceptible population. The proposed model is named SIDUR, which corresponds to the five compartments: susceptible (S), undiagnosed infected (I), diagnosed infected (D), unidentified recovered (U), and identified removed (R). The model is characterized by four parameters and one input, which is the testing rate.

2.1. Model design

SIDUR is a compartmental model depicted in Figure 1. At time $t \in \mathbb{R}_{\geq 0}$, each compartment is characterized by a single state:

- $x_S(t)$: Number of susceptible people who are prone to the disease.
- $x_I(t)$: Number of infected people who are undetected by the public health authority.
- $x_D(t)$: Number of diagnosed people who are infected and detected by a test.

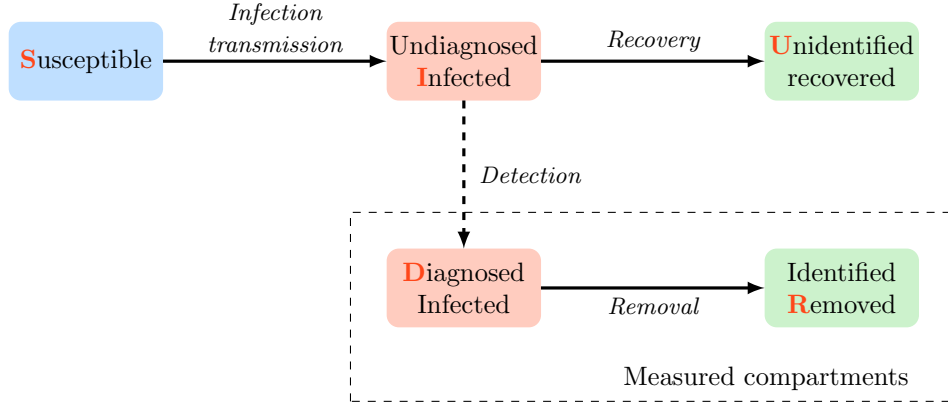


Figure 1: Block diagram of SIDUR model.

- $x_U(t)$: Number of unidentified recovered people who recover without getting diagnosed.
- $x_R(t)$: Number of removed people who either recover or die after being diagnosed.

The development of the model is based on the following three assumptions:

Assumption 1. The population remains constant during the evolution of epidemic:

$$x_S(t) + x_I(t) + x_D(t) + x_U(t) + x_R(t) = N$$

where N stands for the total population. ┘

During the evolution of the epidemic, the change in population due to births, deaths due to causes exclusive to the epidemic, and inflow/outflow of travelers from/to other countries is assumed to be negligible with respect to the total population. In other words, during the time horizon under consideration, the birth rate is approximately equal to the death rate and the rate of inflow of travelers is approximately equal to the rate of outflow.

Assumption 2. Only the undiagnosed infected population $x_I(t)$ is responsible for the disease transmission to the susceptible population $x_S(t)$. ┘

The diagnosed infected people $x_D(t)$ are isolated from the population in the form of quarantine or hospitalization. This means that their contact with susceptible people is restricted or, for hospitalized cases, under strict sanitary measures. Therefore, disease transmission due to a diagnosed infected person is assumed to be unlikely. It is possible, however, that a diagnosed infected person could have transmitted the disease to other people before getting diagnosed.

Assumption 3. All the deaths from COVID-19 are identified and reported; they are included in the removed population $x_{\text{R}}(t)$ along with the people who recover after being diagnosed. \lrcorner

The non-surviving cases of the disease usually have severe symptoms. Therefore, they are assumed to be diagnosed and hospitalized before their death.

Assumption 4. The acquired immunity of recovered population is sustainable enough. That is, the unidentified recovered population $x_{\text{U}}(t)$ and the removed population $x_{\text{R}}(t)$ cannot get infected again during the time interval considered in an epidemic. \lrcorner

Once the infected cases recover, they develop antibodies that prevent them from getting infected again by the same variant of the disease. However, if the epidemic endures for an extended period of time, then other variants of the disease emerge causing the recovered population to become susceptible again. Nonetheless, we consider a beginning phase of the epidemic during which we assume that disease variants do not emerge and the acquired immunity of recovered population is sustainable enough.

Based on the above assumptions, the model is given as

$$\dot{x}_{\text{S}}(t) = -\beta x_{\text{S}}(t) \frac{x_{\text{I}}(t)}{N} \quad (1\text{a})$$

$$\dot{x}_{\text{I}}(t) = \beta x_{\text{S}}(t) \frac{x_{\text{I}}(t)}{N} - u(t) \frac{x_{\text{I}}(t)}{x_{\text{T}}(t)} - \gamma x_{\text{I}}(t) \quad (1\text{b})$$

$$\dot{x}_{\text{D}}(t) = u(t) \frac{x_{\text{I}}(t)}{x_{\text{T}}(t)} - \rho x_{\text{D}}(t) \quad (1\text{c})$$

$$\dot{x}_{\text{U}}(t) = \gamma x_{\text{I}}(t) \quad (1\text{d})$$

$$\dot{x}_{\text{R}}(t) = \rho x_{\text{D}}(t) \quad (1\text{e})$$

where β , γ , and ρ are the infection, recovery, and removal rates, respectively, $u(t)$ is the testing rate,

$$\begin{aligned} x_{\text{T}}(t) &= x_{\text{I}}(t) + (1 - \theta)(x_{\text{S}}(t) + x_{\text{U}}(t)) \\ &= \theta x_{\text{I}}(t) + (1 - \theta)(N - x_{\text{D}}(t) - x_{\text{R}}(t)) \end{aligned} \quad (2)$$

is the testable population, and θ is the testing specificity parameter which takes values in the interval $[0, 1]$.

The recovery rate γ is the inverse of the average recovery time $1/\gamma$ after which an undiagnosed infected person recovers, and the removal rate ρ is the inverse of the average removal time $1/\rho$ after which a diagnosed infected person recovers or dies. The average recovery time is expected to be shorter than the average removal time, i.e., $\gamma \geq \rho$, because the undiagnosed infected population that comprises the undetected asymptomatic cases and cases with mild symptoms recover faster than the diagnosed population that comprises mostly the cases with severe symptoms.

The infection rate β is the product of the frequency of contacts among the susceptible and infected populations and the probability of disease transmission after a contact has been made. Thus, the parameters β , γ , and ρ are related to the disease biology. However, the value of β can also be partially impacted by non-pharmaceutical interventions (NPI) such as social distancing, lockdown, confinement, travel restrictions, and preventive policies (i.e., to maintain a certain distance from other people, to wear a face mask in public spheres, to wash/sanitize hands more often, etc.). The value of β is expected to be smaller when NPI's are implemented than the value of β when no NPI is implemented. Depending on the time periods during which different NPI policies are implemented, we assume the infection rate β to be piecewise constant.

The testing specificity parameter θ , on the other hand, is solely dependent on the testing policy implemented by the public health authority. Given that the testing rate is constant, the value of θ will be larger when the tests are allocated efficiently through contact tracing than the value of θ when the tests are performed randomly. However, there are other factors that can also influence θ , for example, if only the people with severe symptoms are tested, then the probability $x_{\text{T}}/x_{\text{T}}$ of detecting an infected person from the testable population is equal to one, i.e., the testing specificity parameter $\theta = 1$. This is to indicate that the larger value of θ doesn't necessarily imply the efficiency of testing policy, rather it only signifies the specificity of tests. Depending on the time periods during which different testing policies are implemented, we assume the testing specificity parameter θ to be piecewise constant.

2.2. Control input and testable population

We consider the testing rate $u(t)$ to be the model input which depends on three factors: the daily testing capacity $c(t)$, the remaining stockpile of tests $r(t)$, and the testable population $x_{\text{T}}(t)$. We consider $c(t)$ to be time-varying in order to take into account the fact that the capacity of testing can change on a daily basis. Moreover, on a given day t , one cannot do more tests than the daily testing capacity $c(t)$, the remaining stockpile of tests $r(t)$, or the testable population $x_{\text{T}}(t)$. Therefore, we have

$$u(t) := \min(c(t), r(t), x_{\text{T}}(t)). \quad (3)$$

In case the total stockpile of tests r_{max} is limited, the remaining stockpile of tests at time t is given by

$$r(t) := r_{\text{max}} - \int_0^t u(\eta) d\eta.$$

In case new tests can be produced and supplied easily, (3) is simply given by

$$u(t) := \min(c(t), x_{\text{T}}(t)).$$

Usually, the testable population $x_{\text{T}}(t)$ is much larger than the daily testing capacity, thus

$$u(t) = \begin{cases} c(t), & \text{if } r_{\text{max}} \text{ is unlimited} \\ \min(c(t), r(t)), & \text{if } r_{\text{max}} \text{ is limited.} \end{cases} \quad (4)$$

In order to diagnose the infected people at time t , the tests are allocated to a proportion of the testable population $x_{\text{T}}(t)$, which is a sample from the total population N . From (2), it is obvious that the infected population $x_{\text{I}}(t) \leq x_{\text{T}}(t)$ at any given time t . Thus, given the testing specificity parameter $\theta \in [0, 1]$, the probability of detecting an infected person per test in a homogeneous population structure is given by $x_{\text{I}}(t)/x_{\text{T}}(t)$.

The testing specificity parameter θ allows for the adjustment of the testable population to accommodate for the detection rate of tests. In most countries, at the beginning of an epidemic outbreak, the number of available tests are limited. Thus, the available tests are usually utilized to confirm the symptomatic infected cases or to diagnose certain people such as medical care agents, politicians, athletes, etc. In such a case, the testable population is close to the infected population and the value of θ increases to approximately one. Once the capacity of testing is increased, the size of the testable population is also increased that can include, for example, contacts of diagnosed people, the whole population of a city where a cluster is identified, travellers, etc. As a consequence, the value of θ decreases.

2.3. Outflows from the model compartments

The SIDUR model is described by the one-way transfer of population between compartments, where an outflow from one compartment is the inflow to the other compartments. Thus, it suffices to describe only the outflows from the compartments to describe the dynamics of the model.

Infection transmission. In the beginning of the epidemic, most of the population is in the susceptible compartment (S) with the exclusion of those who are initially infected and/or diagnosed. Some of the susceptible people in S may get infected and leave this compartment when they come in contact with an infected person. The rate of the outflow from this compartment is according to the infection transmission rate, which depends on the product of the number of susceptible and infected populations, and is given as

$$\beta x_{\text{S}}(t) \frac{x_{\text{I}}(t)}{N}$$

where β is the infection rate. The term $x_{\text{I}}(t)/N$ is the proportion of undetected infected population at any time t in a homogeneous population structure. Note that in light of Assumption 2, diagnosed population $x_{\text{D}}(t)$ does not participate in the infection transmission because they are either quarantined and/or hospitalized, i.e., they are temporarily removed from the population. Finally, by Assumption 4, there is no inflow to the susceptible compartment.

Detection. The outflow from the infected compartment (I) is either due to detection (i.e., transfer to the diagnosed compartment (D)) or recovery without detection (i.e., transfer to the unidentified recovered compartment (U)). The first outflow is due to the testing rate $u(t)$, i.e., the number of tests performed

per day. Since the probability of detecting an infected person from a testable population by a single test is $x_I(t)/x_T(t)$, therefore we have

$$u(t) \frac{x_I(t)}{x_T(t)}$$

the rate of diagnosing the infected population in I compartment.

Recovery. The second outflow from the I compartment consists of those people who are not diagnosed and recover naturally with an average recovery period of $1/\gamma$. The unidentified recovered compartment (U) accumulates the infected people who recover naturally without being detected with a recovery rate γ .

Removal. The diagnosed compartment (D) admits $u(t)x_I(t)/x_T(t)$ as an inflow, whereas the outflow is $\rho x_D(t)$ with ρ being the removal rate. That is, $1/\rho$ is the average time period after which a typical diagnosed person either recovers or dies. The removed compartment (R) accumulates the diagnosed people who die or recover with a removal rate ρ .

2.4. Output signals from the model

The outputs signals $y_i(t)$, $t \in \mathbb{R}_{\geq 0}$, $i = 1, 2, \dots, m$, from the model (or model outputs) correspond to the sampled output measurements $\bar{y}_i(k)$, $k \in \mathbb{Z}_{\geq 0}$, from the data (or data outputs), which can be approximated by a continuous signals $\bar{y}_i(t)$, respectively. The data outputs and model outputs are related as follows:

$$\bar{y}_i(t) = y_i(t) + w_i(t)$$

where $w_i(t)$ is the measurement noise/error.

The model outputs are functions of the states of SIDUR model, which are of two types: (i) outputs whose functions are known and (ii) outputs whose functions are not known. First, we define three model outputs whose function is known.

- Cumulative number of diagnosed people

$$y_1(t) = x_D(t) + x_R(t). \tag{5}$$

- Cumulative number of removed people

$$y_2(t) = x_R(t). \tag{6}$$

- Number of positively tested people (or positive test results) per day

$$y_3(t) = u(t) \frac{x_I(t)}{x_T(t)}. \tag{7}$$

These model outputs are fitted with the data outputs in order to estimate the model parameters $\beta, \theta, \gamma, \rho$ in Section 4. Note that these model outputs are related to each other. Since the number of diagnosed infected people at any time t can be obtained as $x_D(t) = y_1(t) - y_2(t)$, which is also known as the number of active diagnosed cases, we obtain the following relation between $y_1(t)$ and $y_2(t)$ from (1e)

$$\dot{y}_2(t) = \rho(y_1(t) - y_2(t)). \quad (8)$$

On the other hand, the number of positive test results per day $y_3(t)$ is related to the cumulative number of diagnosed cases $y_1(t)$ by the following relation

$$y_3(t) = \dot{x}_D(t) + \dot{x}_R(t) = \dot{y}_1(t). \quad (9)$$

The cumulative number of diagnosed people $y_1(t)$ can be obtained by integrating the daily number of positive test results as

$$y_1(t) - y_1(0) = \int_0^t y_3(\eta) d\eta. \quad (10)$$

These output relations (8), (9), (10) are used to infer the missing data from the available data in Section 3.

Second, one can note that the number of active ICU patients, denoted $B(t)$, (or ICU beds occupied) and the cumulative number of deaths, denoted $E(t)$, (or extinct cases) are positively correlated with the total number of active infected cases $A(t) = x_I(t) + x_D(t)$ and the cumulative number of infected cases $I(t) = N - x_S(t)$, respectively. Therefore, we define two additional model outputs:

- Number of active Intensive Care Unit (ICU) cases (or the number of ICU beds currently occupied by COVID-19 patients):

$$y_4(t) := B(t) = g(A(t - \psi^{-1})) \quad (11)$$

where $A(t) = x_I(t) + x_D(t)$ is the number of active infected cases, ψ^{-1} is the average time period a typical COVID-19 ICU case takes from getting infected to being admitted to ICU, and the function g is to be defined.

- Cumulative number of deaths due to COVID-19 (or extinct cases):

$$y_5(t) := E(t) = h(I(t - \phi^{-1})) \quad (12)$$

where $I = N - x_S(t)$ is the cumulative number of infected cases, ϕ^{-1} is the average time period a typical COVID-19 extinct case takes from getting infected to death, and the function h is to be defined.

The functions g and h will be estimated from available data on the number of active ICU cases and the cumulative number of deaths, respectively, in Section 4. Then, the model outputs $B(t)$ and $E(t)$ will be used as performance outputs to evaluate the testing strategies proposed in Section 5.

2.5. Basic and effective reproduction numbers

An important quantity to assess the epidemic potential of a disease is the *basic reproduction number* R_0 , which is defined as the expected number of secondary infected cases produced by a single infected person in a completely susceptible population [10]. If $R_0 > 1$, then each generation of infected cases produces more secondary cases in the next generation and the disease has a potential of becoming an epidemic. If $R_0 < 1$, then each generation of infected cases produces less secondary cases in the next generation and the disease will eventually die out. It is worth noticing, however, that the definition of R_0 assumes that the people around a primary infected case are all susceptible. This suggests that determining R_0 is important only at the onset of an epidemic. However, in the later stages, more people get infected and not all people around an infected person are necessarily susceptible. As more people get infected, the conditions favoring the disease to propagate change and the number of susceptible people that an infected person infects is actually less than that what R_0 predicts. Thus, a more suitable quantity during the later stages of the epidemic is the *effective reproduction number* R_t , which takes into account the proportion of susceptible people in the total population [47].

For the SIDUR model, we define the effective reproduction number R_t to be the ratio of the inflow and the outflow of the undiagnosed infected compartment (I). That is, R_t is the ratio of the number of newly infected people and the number of newly diagnosed and recovered people at time t . If $R_t < 1$, this means that more people are being diagnosed and recovered than the people being infected at time t , which implies that $x_I(t)$ will decrease. If $R_t > 1$, this means that more people are being infected than the people being diagnosed and recovered at time t , which implies that $x_I(t)$ will increase. Notice that the definition of the basic reproduction number R_0 is same as R_t for $t = 0$.

To derive the expression of R_t , we consider the model equation (1b), where the undiagnosed infected population $x_I(t)$ satisfies

$$\dot{x}_I(t) = \beta x_S(t) \frac{x_I(t)}{N} - u(t) \frac{x_I(t)}{x_T(t)} - \gamma x_I(t).$$

The positive rate (inflow) $\beta x_S(t)x_I(t)/N$ tells how many new infections will be generated in the next moment, and the negative rates (outflows) $u(t)x_I(t)/x_T(t)$ and $\gamma x_I(t)$ tell how many infected people will be diagnosed or recovered in the next moment, respectively. Therefore, the effective reproduction number is the following ratio

$$R_t = \frac{\beta x_S(t) \frac{x_I(t)}{N}}{u(t) \frac{x_I(t)}{x_T(t)} + \gamma x_I(t)} = \frac{\beta}{\frac{u(t)}{x_T(t)} + \gamma} \frac{x_S(t)}{N}. \quad (13)$$

To derive the expression of R_0 , we consider the expression of R_t at $t = 0$, which corresponds to the onset of the epidemic. For $t = 0$, we can assume few infected cases, which implies that $x_S(0) \approx N$ and $x_T(0) \approx (1 - \theta)N$. Under these

approximations, we have

$$R_0 = \frac{\beta}{\frac{u(0)}{(1-\theta)N} + \gamma}. \quad (14)$$

This expression can also be obtained by following the methodology of [48]. Notice that the reproduction number R_0 depends on the initial testing policy $u(0)$. This indicates that it is possible to suppress the epidemic in the beginning by having an intensive testing policy, which can be seen, for example, in the case of South Korea [49]. In general, however, we have $u(0) \approx 0$, which gives $R_0 \approx \beta/\gamma$.

3. Data acquisition and imputation

The data related to COVID-19 in France is collected from the French government’s platform for publicly available data³ for the time period of January 24 to July 01, 2020. In particular, we use datasets provided by the French Ministry of Social Affairs and Health (Ministère des Solidarités et de la Santé (MSS)) and the French Public Health Agency (Santé Publique France (SPF)). From MSS, we obtain the data about different categories of people affected by COVID-19, i.e., diagnosed, hospitalized, recovered from hospitals, and dead. From SPF, we obtain the data for the number of PCR tests performed and positive test results obtained per day.

The data obtained from both sources is incomplete in several aspects. For instance, the data for the number of recovered people does not record those who recover from their homes after being diagnosed. These people do not show severe symptoms of the disease and, therefore, are not hospitalized, but are quarantined in their homes for some days. Only those who are hospitalized after being diagnosed are recorded as recovered when they are discharged from the hospital. On the other hand, the data for COVID-19 PCR tests is also incomplete. To illustrate this, we consider three intervals of time: (1) January 24 to March 09, 2020, (2) March 10 to May 12, 2020, and (3) May 13 to July 01, 2020. There is no data available for the tests during the first interval. During the second interval, the testing data is collected only from the medical laboratories and not from the hospitals. However, we have reliable data only during the third interval which is collected both from the medical laboratories and the hospitals. Therefore, the data obtained from the above sources can be considered as a raw data which needs to be imputed.

3.1. Raw data

This subsection illustrates the data obtained from MSS and SPF without any modification.

³Website: Open platform for French public data. (Accessed 17/10/2020)

Cumulative number of diagnosed cases. We denote the data for the cumulative number of diagnosed cases by \bar{y}_1 , which is illustrated in Figure 2 and corresponds to the model output $y_1(t)$ in (5). It is also known as the total “confirmed” cases. This is a cumulative data for all the cases diagnosed with the disease through RT-PCR tests⁴. Thus, it includes both the active cases (those who are either admitted to the hospitals and/or quarantined) and the inactive cases (those who either recovered or died after being diagnosed). That is, $\bar{y}_1(k)$ corresponds to the sum of people in the diagnosed (D) and removed (R) compartments of the SIDUR model (Figure 1) on a given day k , as given in (5).

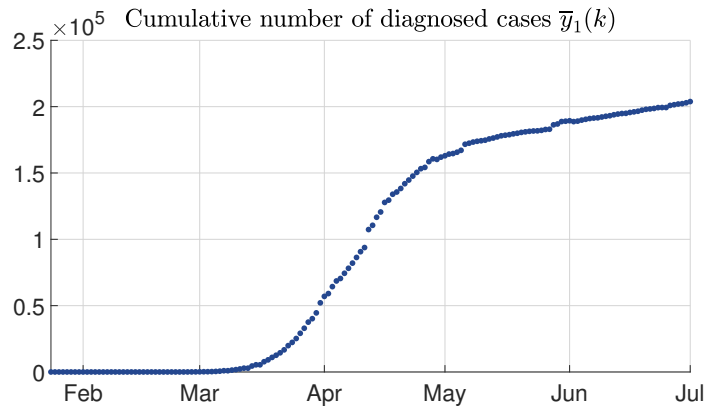


Figure 2: Cumulative number of diagnosed cases $\bar{y}_1(k)$ from January 24 to July 01, 2020. Source: MSS.

There is also an additional data for the diagnosed cases from French retirement homes (EHPAD). However, the French government database⁵ and several other international databases^{6 7} do not add the diagnosed cases from EHPAD to the cumulative number of diagnosed (confirmed) cases. That is, the data for the cumulative number of diagnosed cases is considered to be inclusive of the diagnosed cases from EHPAD. However, in all the above databases, the data on cumulative number of deaths is collected separately from both the hospitals and EHPAD.

Number of active hospitalized and ICU cases. The data on the number of active hospitalized cases is denoted as \bar{H} and is illustrated in Figure 3 along with the number of active ICU cases \bar{B} . This data corresponds to the number of people who are admitted to the hospitals and/or ICU on a given day. That is, it is not a cumulative data. Moreover, it doesn’t include those who were diagnosed

⁴Website: Definition of a COVID-19 confirmed case by SPF. (Accessed 14/07/2020)

⁵Website: COVID-19 - France. (Accessed 01/10/2020)

⁶Website: European Centre for Disease Prevention and Control. (Accessed 01/10/2020)

⁷Website: Worldometers/coronavirus/France. (Accessed 01/10/2020)

but not hospitalized. That is, this data corresponds to a certain proportion of people in the diagnosed compartment (D) of the SIDUR model. This data is available from March 17, 2020, onward.

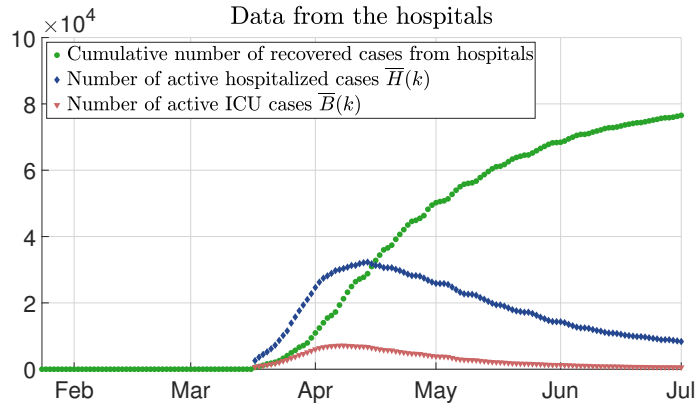


Figure 3: Total number of recovered cases who returned home after hospitalization from January 24 to July 01, 2020. The number of active COVID-19 hospitalized cases $\bar{H}(k)$ and ICU cases $\bar{B}(k)$ from March 17 to July 01, 2020. Source: MSS.

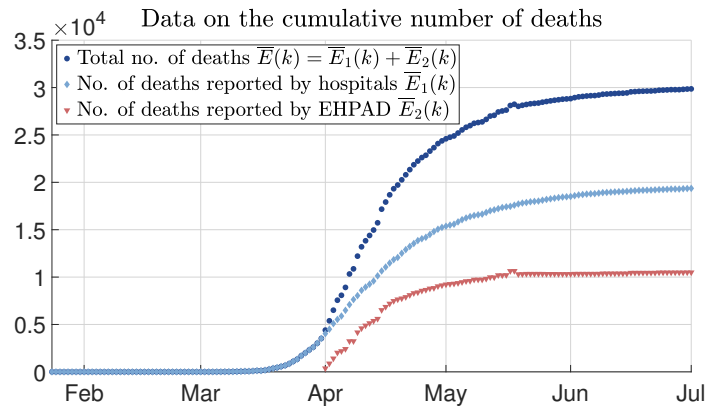


Figure 4: Total number of deaths from COVID-19 reported (a) by hospitals from January 24 to July 01, 2020, and (b) by retirement homes (EHPAD) from April 01 to July 01, 2020. Source: MSS.

Cumulative number of recovered cases from hospitals. This data is illustrated in Figure 3. It corresponds to people who, after recovering from the disease, were discharged from the hospitals. Obviously, prior to recovering, they were diagnosed with the disease and hospitalized due to their severe symptoms.

Cumulative number of deaths. The data on the cumulative number of diagnosed cases is considered to be inclusive of the diagnosed cases from the French retirement homes (EHPAD). However, the case for data on the cumulative number of deaths is different. Those who died at the hospitals and those who died in the retirement homes (EHPAD) are considered to be distinct. Thus, the cumulative number of deaths is the sum of both data, which are illustrated in Figure 4.

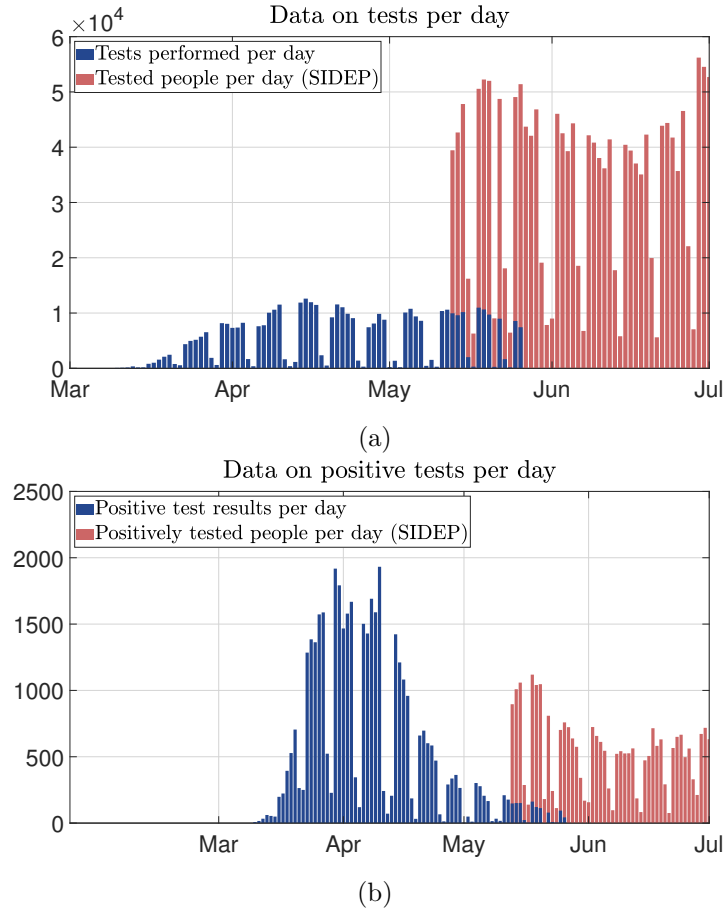


Figure 5: Data on the PCR tests: (a) The number of tests performed per day from March 10 to May 26 and the number of tested people per day from May 13 to July 01; (b) The number of positive test results per day from March 10 to May 26 and the number of positively tested people per day from May 13 to July 01. Source: SPF.

Number of tests and positive tests per day. We have two types of data related to COVID-19 PCR tests. The first type of data is collected by SPF on the number of tests performed and positive test results per day from March 10 to May 26, 2020. However, this data is collected only from the central sampling laboratories: Eurofins Biomnis and Cerba. Figure 5(a) (blue) illustrates the number of tests

performed per day and Figure 5(b) (blue) illustrates the number of positive test results per day.

The second type of data was made available after the deployment of a new information screening system (SI-DEP) by the SPF. This data is available from May 13, 2020, onward. It is collected from both the laboratories and the hospitals. However, the data reported by SI-DEP is the number of ‘tested people’ per day instead of the number of ‘tests performed’ per day. SI-DEP guarantees that only one test is counted per person. In the case of, for instance, multiple negative test results for a certain person, SI-DEP considers only the first date on which the PCR test was performed. Later, if that person gets a positive test result, then only this new result is reported in the data and the previous data is erased. Figure 5(a) (red) illustrates the number of tested people per day and Figure 5(b) (red) illustrates the number of positively tested people per day.

3.2. Imputed data

In the raw data, we only have the data for those who recover or die in the hospitals after being diagnosed with the disease. However, the removed compartment of the SIDUR model also comprises the diagnosed cases who were not hospitalized but were quarantined in their homes. There is no data that records the recovery of these people. Moreover, the data on PCR tests is also incomplete; there is no data on PCR tests from January 24 to March 09, 2020, and the data from March 10 to May 12, 2020, doesn’t include the tests performed in the hospitals. Therefore, in order to infer the missing data, we impute the raw data by making reasonable assumptions.

3.2.1. Cumulative number of removed cases

From the data on the total number of recovered people from hospitals shown in Figure 3, we see that 76,540 people have recovered from the hospitals as of July 01, 2020. If we subtract this number and the total number of deaths (Figure 4), i.e., 29,860, from the total number of diagnosed cases (Figure 2), i.e., 165,700, we obtain $165,700 - 76,540 - 29,860 = 59,300$ people. Further subtracting the currently hospitalized cases (Figure 3), i.e., 8336 as of July 01, we obtain $59,300 - 8336 = 50,964$ people, who might still be infected or have recovered. These people were diagnosed but were not hospitalized; they were quarantined in their homes. However, there is no data that provides a correct answer for how many people have recovered and how many of them are still infected. Therefore, using the relevant raw data, we infer the cumulative number of removed cases \bar{y}_2 by estimating the number of diagnosed cases who recovered from home.

We use the following notations for simplicity and brevity:

\bar{y}'_1	Total diagnosed and hospitalized
\bar{y}''_1	Total diagnosed but not hospitalized
\bar{y}'_2	Total recovered/died in a hospital
\bar{y}''_2	Total recovered from home after diagnosis

By definition, we have

$$\begin{aligned}\bar{y}_1(k) &= \bar{y}'_1(k) + \bar{y}''_1(k) \\ \bar{y}_2(k) &= \bar{y}'_2(k) + \bar{y}''_2(k)\end{aligned}\tag{15}$$

where k is from January 24 to July 01, 2020. Note that $\bar{y}_1(k)$ is illustrated in Figure 2 and

$$\bar{y}'_1(k) = \bar{y}'_2(k) + \bar{H}(k),$$

where $\bar{y}'_2(k)$ is the sum of the total number of recovered cases from hospitals (Figure 3) and the total number of deaths (Figure 4), and \bar{H} is the number of active hospitalized cases (Figure 3). Thus, we can compute the total diagnosed cases who were not hospitalized as

$$\bar{y}''_1(k) = \bar{y}_1(k) - \bar{y}'_1(k).$$

Since there is no data for the diagnosed people who recovered from home, therefore $\bar{y}''_2(k)$ is unknown. Thus, we assume the following:

$$\frac{\bar{y}''_2(k)}{\bar{y}''_1(k)} = \frac{\bar{y}'_2(k)}{\bar{y}'_1(k)}.\tag{16}$$

That is, the ratio of the diagnosed cases who recovered in homes to the total diagnosed cases who were quarantined at homes is equal to the ratio of the diagnosed cases who recovered or died in hospitals to the total diagnosed cases who were hospitalized. In other words, we assume that the removal rate of people who were not hospitalized is equal to the removal rate of people who were hospitalized. Thus, from (15) and (16), we obtain

$$\bar{y}_2(k) = \bar{y}'_2(k) \left(1 + \frac{\bar{y}''_1(k)}{\bar{y}'_1(k)} \right) = \bar{y}'_2(k) \frac{\bar{y}_1(k)}{\bar{y}'_1(k)},$$

which corresponds to the model output $y_2(t)$ in (6).

3.2.2. Combining two types of testing data

From Figure 5, we see that the first type of data, which is available from March 10 to May 26, 2020, considers the number of tests performed and positive test results per day. On the other hand, the second type of data, which is available from May 13, 2020, onward, considers the number of tested people and positively tested people per day. However, no person is usually tested more than once per day. Therefore, we assume that the number of tested people per day is same as the number of tests performed per day. Similarly, the number of positively tested people per day is same as the number of positive test results per day. Note that if a person is tested more than once but on different days, then this assumption is not violated.

We consider three time intervals: (i) January 24–March 09, when there is no data on PCR tests; (ii) March 10–May 12, when there is incomplete data; (iii) May 13–July 01, when there is complete data. Let \bar{u} and \bar{y}_3 denote the

number of tests performed and the number of positive test results per day, respectively, for the entire interval January 24 to July 01, 2020. Let \bar{u}' , \bar{u}'' , \bar{u}''' and \bar{y}'_3 , \bar{y}''_3 , \bar{y}'''_3 be the number of tests performed and the number of positive test results obtained for the first, second, and third time intervals, respectively. Since the data in the third interval is reliable, we do not make any imputations for \bar{u}''' and \bar{y}'''_3 . For the other two intervals, we make reasonable assumptions to complete the data.

- (i) *January 24–March 09*: This interval corresponds to the beginning of the epidemic in France and the data for tests performed and positive test results per day for this interval is \bar{u}' and \bar{y}'_3 , respectively. During this interval, only those people were tested who showed symptoms. Moreover, recall the output relation (9). Then, we compute the data as follows: $\bar{u}'(k) \approx \bar{y}'_3(k)$ and $\bar{y}'_3(k) = \bar{y}_1(k+1) - \bar{y}_1(k)$, where k is from January 24 to March 09. In other words, during the first interval, the number of tests performed per day is assumed to be approximately equal to the number of positive test results obtained per day. Moreover, the number of positive test results obtained per day is equal to the number of diagnosed cases that day.
- (ii) *March 10–May 12*: In the second interval, we have the data on PCR tests that is reported only by the laboratories and not by the hospitals. During this interval, we compute the data as follows: \bar{u}'' is same as the data (Figure 5) and $\bar{y}''_3(k) = \bar{y}_1(k+1) - \bar{y}_1(k)$, where k is from March 10 to May 12.

Based on the above data imputations, we obtain the number of tests performed per day $\bar{u} = [\bar{u}' \quad \bar{u}'' \quad \bar{u}''']$, which corresponds to the control input $u(t)$, and the number of positive tests obtained per day $\bar{y}_3 = [\bar{y}'_3 \quad \bar{y}''_3 \quad \bar{y}'''_3]$, which corresponds to the model output $y_3(t)$ in (7), for the complete time interval January 24 to July 01, 2020.

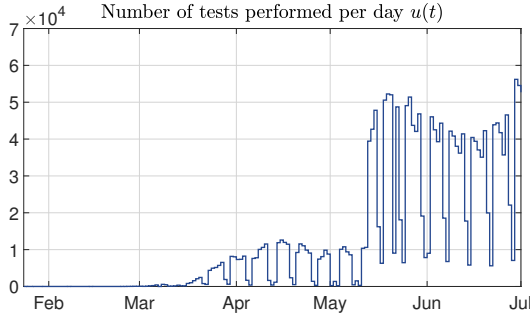


Figure 6: Input signal from the data.

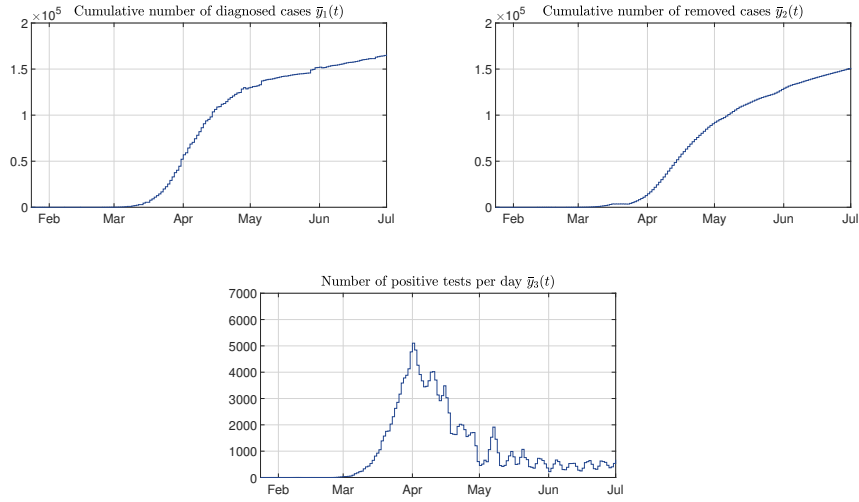


Figure 7: Output signals from the data.

3.3. Input and output signals from the data

The input $u(t)$ to the model corresponds to the number of tests performed per day \bar{u} . Let $\{1, 2, \dots, \tau\}$ be the index set of the 160 days from January 24 to July 01, 2020, where $\tau = 160$. Then, for $k = 1, 2, \dots, \tau$, we define the input signal as

$$u(t) = \bar{u}(k), \quad \text{for } [t] \leq k < [t+1]$$

which is illustrated in Figure 6.

We denote by $\bar{y}_i(t)$, for $i = 1, 2, 3$, the outputs obtained from the data. That is, the output signals from the data and the model are related as follows

$$\begin{aligned} \bar{y}_1(t) &= y_1(t) + w_1(t) \\ \bar{y}_2(t) &= y_2(t) + w_2(t) \\ \bar{y}_3(t) &= y_3(t) + w_3(t) \end{aligned}$$

where $w_i(k)$, for $i = 1, 2, 3$, represents the measurement noise. The outputs $\bar{y}_1(t)$, $\bar{y}_2(t)$, and $\bar{y}_3(t)$ correspond to the cumulative number of diagnosed cases, the cumulative number of removed cases, and the number of positive test results per day, respectively. Similar to the input $u(t)$, we define the output signals from the data as

$$\begin{cases} \bar{y}_1(t) = \bar{y}_1(k), & \text{for } [t] \leq k < [t+1] \\ \bar{y}_2(t) = \bar{y}_2(k), & \text{for } [t] \leq k < [t+1] \\ \bar{y}_3(t) = \bar{y}_3(k), & \text{for } [t] \leq k < [t+1] \end{cases}$$

which are illustrated in Figure 7.

4. Estimation of model parameters

In this section, we validate the SIDUR model by estimating the model parameters ρ , β , θ , and γ for the case of COVID-19 in France.

4.1. Estimation of ρ

The removal rate ρ can be directly estimated from the data outputs \bar{y}_1 and \bar{y}_2 . Consider a daily sampling of the model equation (1e), which leads to

$$\Delta x_{\text{R}}(k) \approx \rho x_{\text{D}}(k)$$

where Δ stands for the forward difference operator, i.e., $\Delta x_{\text{R}}(k) = x_{\text{R}}(k+1) - x_{\text{R}}(k)$ for some nonnegative integer k . Therefore, from the relation between y_1 and y_2 in (8), we obtain

$$\Delta \bar{y}_2(k) = \rho \bar{y}_{12}(k) + e(k) \tag{17}$$

where $\bar{y}_{12}(k) = \bar{y}_1(k) - \bar{y}_2(k)$ and $e(k)$ is the error term due to measurement noise. Then, the problem of estimating ρ can be formulated as follows: Find ρ^* such that

$$\rho^* = \arg \min_{\rho \in [0,1]} \sum_{k=1}^{\tau} \|\Delta \bar{y}_2(k) - \rho \bar{y}_{12}(k)\|^2.$$

Notice that the solution of this problem can be obtained through least-square estimation [50, Chapter 7].

4.2. Estimation of β, θ, γ

We formulate a problem of fitting the model outputs $y_1(t), y_2(t), y_3(t)$ to the data outputs $\bar{y}_1(k), \bar{y}_2(k), \bar{y}_3(k)$, where $k = 1, 2, \dots, \tau$ with τ being the final time. The model fitting is done by optimizing the parameters β, θ, γ for the time interval $[0, \tau]$ under the assumption that γ is constant whereas β and θ are piecewise constants.

To limit the rate of spread of COVID-19, the French government announced to place a lockdown all over France from March 17 to May 10, 2020, which included restricted human mobility, strict social distancing measures, and closure of schools, offices, and marketplaces. However, the essential services and public establishments were authorized to remain open under strict preventive measures. People were allowed to leave their homes with face masks only for necessary groceries, brief exercise within a certain radius of their homes, or for urgent medical reasons. Such an intervention from the public authority is necessary to mitigate the rate of spread of the disease and to reduce the value of infection rate β . Therefore, in relation to the case of France, we divide the time into three intervals: (i) Before lockdown (January 24 to March 16), (ii) During lockdown (March 17 to May 10), and (iii) After lockdown (May 11 to July 01). We consider a different value of the infection rate β during each of these intervals, i.e.,

$$\beta(k) = \begin{cases} \beta_1, & \text{for } k = \text{January 24 to March 16} \\ \beta_2, & \text{for } k = \text{March 17 to May 10} \\ \beta_3, & \text{for } k = \text{May 11 to July 01} \end{cases}$$

where $\beta_1, \beta_2, \beta_3$ are positive real numbers. For the testing specificity parameter θ , we divide the time into two intervals: (i) Before May 11 and (ii) After May

11, where May 11 corresponds to the change in testing policy in France [5] (also see the website of Our-World-in-Data⁸). Thus, we have

$$\theta(k) = \begin{cases} \theta_1, & \text{for } k = \text{January 24 to May 10} \\ \theta_2, & \text{for } k = \text{May 11 to July 01} \end{cases}$$

where θ_1 and θ_2 are real numbers in the interval $[0, 1]$.

Let $p = [\beta_1 \ \beta_2 \ \beta_3 \ \theta_1 \ \theta_2 \ \gamma]^T$ be the parameter vector. Then, the goal is to find p^* such that

$$p^* = \arg \min_p \mathcal{J}(p) \quad (18)$$

where the cost function is given by

$$\mathcal{J}(p) = \sum_{k=1}^{\tau} \left[(y_1(k, p) - \bar{y}_1(k))^2 + (y_2(k, p) - \bar{y}_2(k))^2 + (y_3(k, p) - \bar{y}_3(k))^2 \right] \quad (19)$$

with the model outputs y_i , $i = 1, 2, 3$, depending on the parameter vector p . Note that we consider the data from January 24 to July 01, 2020, therefore we have $\tau = 160$ days.

To solve this problem, one can also pose it as a least-square estimation, as we did for the removal rate ρ , by defining relations between the data outputs $\bar{y}_1, \bar{y}_2, \bar{y}_3$. However, such relations include the difference operator Δ applied twice to the data outputs, which is usually not recommended when the data is noisy because it amplifies the measurement noise. Moreover, the gradient-based estimation algorithms [51, Chapter 4] are also not suitable due to the difficulty of computing the gradient of the cost function \mathcal{J} online with respect to the parameter vector p . This is because the model outputs y_1, y_2, y_3 do not depend directly on the parameters but through the solution trajectories of the SIDUR model. For simplicity, therefore, we choose the particle swarm optimization (PSO) [52] described in Appendix A, which is a ‘derivative-free’ algorithm, to estimate the parameter vector p .

Infection rate	$\beta_1 = 0.3708$ $\beta_2 = 0.0707$ $\beta_3 = 0.3717$
Testing specificity	$\theta_1 = 0.9948$ $\theta_2 = 0.9967$
Recovery rate	$\gamma = 0.1589$
Removal rate	$\rho = 0.0499$

Table 1: Estimated parameter values.

The estimated parameter values are provided in Table 1. The estimated recovery rate γ and removal rate ρ show that an undiagnosed person recovers in an average period of about 6.3 days and a diagnosed person recovers or dies in an average period of about 20 days. The testing specificity parameter changes

⁸Website: Our World in Data: COVID-19 Testing Policies. (Accessed 30/09/2020)

slightly from $\theta_1 = 0.9948$ to $\theta_2 = 0.9967$, which can have significant impact on the positive test results because it multiplies with the sum of the susceptible and unidentified recovered population in (2) that is in the order of 10^7 in the case of France.

The infection rate β changes its value twice. First, it drops from $\beta_1 = 0.3708$ to $\beta_2 = 0.0707$ when the lockdown is implemented in France on March 17, which significantly decreased the rate of the epidemic spread. Then, it rises from $\beta_2 = 0.0707$ to $\beta_3 = 0.3717$ when the lockdown is lifted on May 10. Many restrictions like social distancing and wearing of face masks were still in place after May 10 in order to prevent the spread of COVID-19 in France. However, the increase in the value of β can be explained by the summer vacations when people were allowed to travel everywhere across France and Europe⁹. This made the places with tourist attractions very crowded and resulted in a higher infection rate.

4.3. Model fitting

Using the estimated values of the model parameters in Table 1, we run the model from January 24 to July 01, 2020. The model fits the output signals data as shown in Figure 8.

The basic reproduction number R_0 at the outbreak of the COVID-19 epidemic in France is computed using (14) and the average value of effective reproduction number R_t during three phases (before, during, and after lockdown) are computed using (13). For the ‘after lockdown’ phase, we chose July 01, 2020, to compute R_t because it is the date up to which our data is considered. These computed values are shown in Table 2 with a comparison to the ones reported by the government^{10 11}.

Epidemic phases	Computed from model	Reported by French government
Outbreak	$R_0 = 2.33$	$R_0 = 2.7$
Before lockdown	$R_t = 2.3$	$R_t = 2.7$
During lockdown	$R_t = 0.33$	$R_t = 0.7$
After lockdown	$R_t = 1$	$R_t = 1$

Table 2: The basic reproduction number R_0 at the outbreak and the values of the effective reproduction number R_t at the end of each phase of the COVID-19 epidemic in France. The values computed by our model are quite close to the ones reported by the French government.

The change in the value of R_t also influences the evolution of diagnosed population $x_D(t)$. This is because larger value of R_t results in a larger infected population $x_I(t)$ and smaller value of R_t results in a smaller infected population $x_I(t)$, which respectively increases and decreases the probability of detecting an infected person $x_I(t)/x_T(t)$ by a single test. Keeping the number of tests

⁹Website: Sortir à Paris: Summer holidays in France and Europe. (Accessed 30/09/2020)

¹⁰Website: COVID-19 France: Epidemiological update. (Accessed 30/09/2020)

¹¹Website: COVID-19 France: Monitoring of the epidemic. (Accessed 30/09/2020)

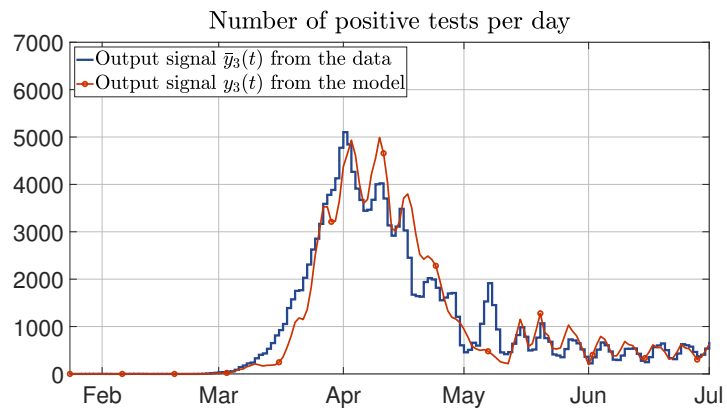
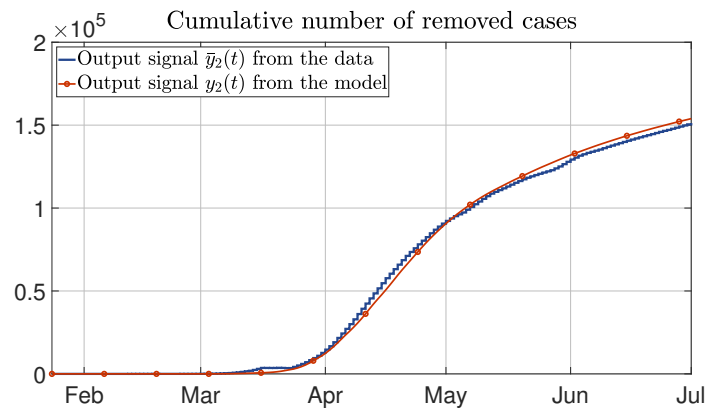
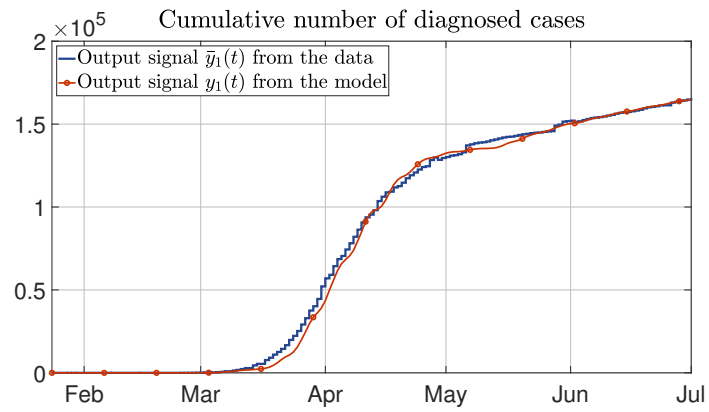


Figure 8: Model fitting.

performed per day same, the larger probability of detection $x_I(t)/x_T(t)$ results in a larger diagnosed population $x_D(t)$.

In Table 2, we see that the placement and lifting of lockdown on March 17 and May 11, respectively, had a significant impact on the value of R_t . Such an effect on R_t impacted the evolution of the diagnosed population $x_D(t) = y_1(t) - y_2(t)$, which can be interpreted as the number of active confirmed cases and is illustrated in Figure 9. The placement of lockdown reduced the value of R_t and resulted in less number of active confirmed cases as compared to the scenario in Figure 9 where the lockdown was not placed on March 17. In this scenario, as shown in Figure 9, the number of active confirmed cases would have increased to a point that could have challenged the available medical facilities such as hospital beds, ventilators, and ICUs. On the other hand, the lifting of lockdown increased the value of R_t and resulted in more number of active confirmed cases as compared to the scenario where the lockdown was not lifted on May 11.

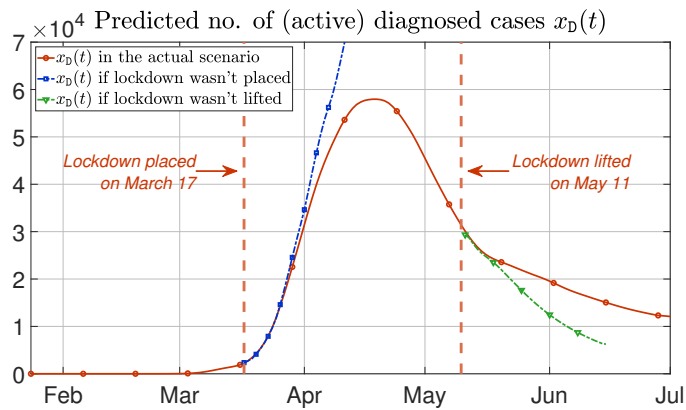


Figure 9: The comparison between the number of active diagnosed cases in the actual scenario vs. two scenarios: if the lockdown was not placed on March 17 and if the lockdown was not lifted on May 11.

4.4. Number of active ICU patients and deaths

The number of active ICU patients $B(t)$ is a function of the number of active infected people $A(t)$. Since an infected person starts to show symptoms after the average incubation period of approximately 5 days, [53], and a person takes on average 12 days from being diagnosed to being admitted to ICU, [54], we assume $\psi^{-1} = 5 + 12 = 17$ days to be the average time delay from getting infected to being admitted to ICU for a typical COVID-19, critically ill case. Thus, we model the number of active ICU patients $B(t)$ as a function of $A(t - \psi^{-1})$, which is approximated by:

$$B(t) = b_1 A(t - \psi^{-1}) + b_2 \sqrt{A(t - \psi^{-1})} \quad (20)$$

where b_1 and b_2 are the parameters (Appendix B, Table B.4) that are determined via the least-square solution to fit (20) to the data on the number of ICU patients. This is illustrated in Figure 10.

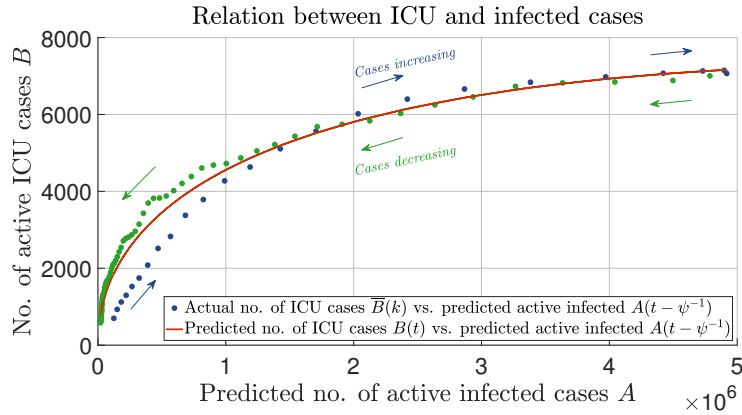


Figure 10: Number of active ICU patients $B(t)$ with respect to the number of active infected cases $A(t)$.

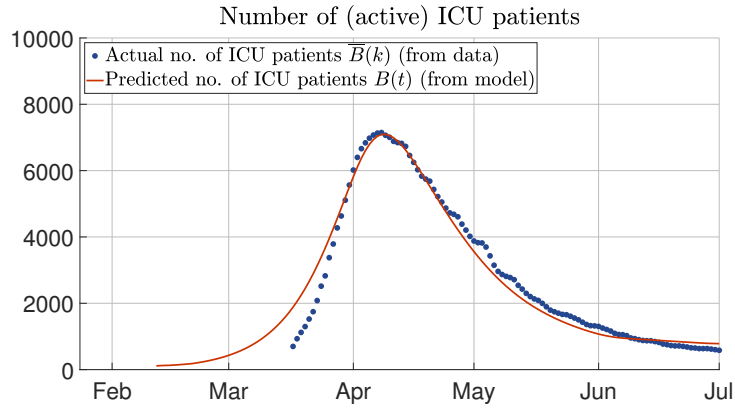


Figure 11: Model fit of the data on the number of active COVID-19 ICU cases $B(t)$ in France using the relation (20).

Similar to the case of the number of active ICU patients, a typical non-surviving case has an average incubation period of 5 days and, in addition to that, an average removal period of $\rho^{-1} \approx 20$ days, where ρ is the removal rate, Table 1. Thus, assuming $\phi^{-1} = 5 + \rho^{-1} = 25$ days to be the average time delay from getting infected to death of a typical non-surviving COVID-19 case, we model the number of deaths $E(t)$ as a function of $I(t - \phi^{-1})$, which is

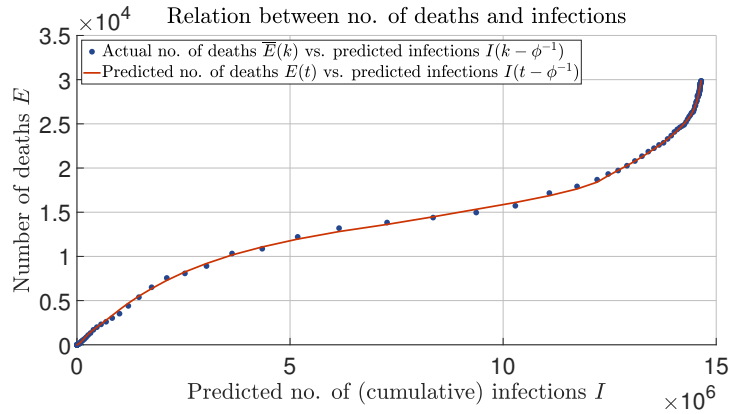


Figure 12: Cumulative number of deaths $E(t)$ with respect to the cumulative number of infected cases $I(t)$.

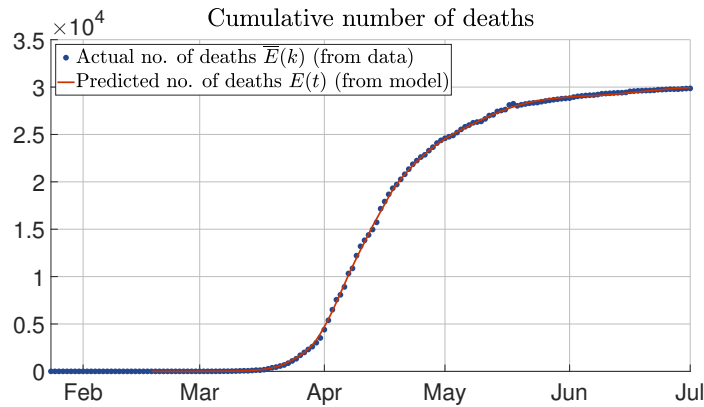


Figure 13: Model fit of the data on the number of COVID-19 deaths $E(t)$ in France using the relation (21).

approximated by the following polynomial:

$$E(t) = \sum_{i=1}^{10} e_i I^i(t - \phi^{-1}) \quad (21)$$

where e_i , for $i = 1, \dots, 10$, are the parameters (Appendix B, Table B.4) that are determined via the least-square solution to fit (21) to the data on the number of deaths. This is illustrated in Figure 12.

Using the relations (20) and (21), we illustrate the model fit of the number of active ICU cases and the cumulative number of deaths with the data in Figure 11 and 13.

5. Study and design of testing policies

In this section, we use the model validated with the COVID-19 data of France to design two types of testing policies. The first one assumes that the total stockpile of tests is unlimited, but the testing capacity per day is limited. This results in the so called *Best Effort Strategy for Testing (BEST)*, which gives the minimum number of tests needed to be performed per day in order to stop the epidemic from growing. In other words, if the BEST is applied, then the number of new infections stop to grow with respect to time.

The second strategy, named COST (*Constant Optimal Strategy for Testing*) assumes that the total stockpile of tests in a country is limited. In such a case, we investigate the optimal testing policy per day that results in a minimum epidemic peak. In contrast with BEST, there is an optimal value of tests to be performed per day which is smaller than the maximum testing capacity per day.

5.1. Best effort strategy for testing

Assume that the total stockpile of tests is unlimited during the whole epidemic period. This assumption is valid for a country that can manufacture or buy tests continuously during the time of the epidemic. Based on this, we provide a testing policy recommendation on the daily testing capacity starting from a certain time t^* in order to change the course of the epidemic in a sense that is defined below. For simplicity, we further assume that the number of tests performed per day is considered to be the daily testing capacity. In other words, the daily testing capacity is utilized completely each day, i.e., $u(t) = c(t)$.

We say that, at any time t , an epidemic is *spreading* if the number of undiagnosed infected population $x_I(t)$ is increasing, i.e., the effective reproduction number $R_t > 1$. On the other hand, an epidemic is *non-spreading* if $x_I(t)$ is not increasing, i.e., the effective reproduction number $R_t \leq 1$.

Definition 1 (BEST). The *best effort strategy for testing (BEST)* at a given time t^* is the minimum number of tests to be performed per day from time t^* onward such that the epidemic switches from spreading to non-spreading at t^* .

In other words, BEST provides the smallest lower bound on the number of tests performed per day, sufficient to change at given time t^* the course of the epidemic from spreading to non-spreading. In order to compute the BEST at a given time t^* , we first define the following function:

$$c^*(t) = x_I(t) \left| \frac{\beta(t)}{N} x_S(t) - \gamma \right|_+ \quad (22)$$

where, by definition, for any scalar z , $|z|_+ = z$ if $z > 0$ and $|z|_+ = 0$ otherwise.

Proposition 1. Assume that the infection rate β is non-increasing while the testing specificity parameter θ is non-decreasing on a time interval $[t^*, t_1)$, for some $t^* < t_1$. Then, the best effort strategy for testing (BEST) at time t^* is given by

$$u(t) = c^*(t^*) = x_I(t^*) \left| \frac{\beta(t^*)}{N} x_S(t^*) - \gamma \right|_+, \quad \forall t \in [t^*, t_1).$$

Proof. In order to prove that $c^*(t^*)$ for $t \in [t^*, t_1]$ is the BEST at t^* , we show the following:

- (i) If $u(t) > c^*(t)$ (resp., $u(t) \geq c^*(t)$) for any $t \in [t^*, t_1]$, then x_I is decreasing (resp., non-increasing) on $[t^*, t_1]$.
- (ii) If $u(t) > c^*(t^*)$ (resp., $u(t) \geq c^*(t^*)$) for any $t \in [t^*, t_1]$, then x_I is decreasing (resp., non-increasing) on $[t^*, t_1]$.

Assume that $u(t) > c^*(t)$ on $[t^*, t_1]$. Then, $\Phi(t) := \beta(t) \frac{x_S(t)}{N} - \frac{u(t)}{x_I(t)} - \gamma < 0$ which implies that x_I is decreasing since $\dot{x}_I(t) = \Phi(t)x_I(t)$ almost everywhere. If only the weaker assumption $u(t) \geq c^*(t)$ on $[t^*, t_1]$ is fulfilled, then, by using the continuity of the solutions of ODE with respect to perturbations of the right-hand side, one gets that x_I is non-increasing.

Assume now that $u(t) > c^*(t^*)$ on $[t^*, t_1]$, where $c^*(t^*)$ is constant. Then, by continuity, $u(t) > c^*(t)$ on a certain interval $[t^*, t_2]$, for some $t_2 \in (t^*, t_1]$. As a consequence of the result (i) shown previously, x_I decreases on $[t^*, t_2]$. Moreover, assume that t_2 is the maximal point in $(t^*, t_1]$ having this property. In order to show that $t_2 = t_1$, it is sufficient to show that $u(t_2) > c^*(t_2)$, otherwise one may consider a larger value for t_2 which will lead to a contradiction with the fact that it is maximal. Since x_I decreases on $[t^*, t_2]$ and θ is non decreasing, from Lemma 6 (see Appendix C), we can conclude that x_I also decreases on this interval. On the other hand, since x_S is always decreasing and β is non increasing, one can conclude that $c^*(t)$ also decreases on $[t^*, t_2]$. This is obtained by upper bounding $c^*(t)$. Thus, one has $c^*(t^*) > c^*(t)$, which implies that $u(t_2) > c^*(t_2)$. Therefore, as $t_2 = t_1$, we have established that x_I decreases on the whole interval $[t^*, t_1]$. For the case where $u(t) \geq c^*(t^*)$, we can use the same argument of continuity of the trajectories.

From the previous results, one deduces that the BEST is given by $c^*(t^*)$ and the testing rate $u(t) \geq c^*(t^*)$ for $t \in [t^*, t_1]$. If $u(t) < c^*(t^*)$, for $t \in [t^*, t_1]$, then one can show easily that the epidemic goes on spreading in the interval $[t^*, t_1]$. Hence, $u(t) = c^*(t^*)$ is the BEST policy at t^* . \square

Algorithm 1: Computation of the BEST policy at time t^* .

1. Inputs: $N, \beta, \gamma, \theta, t^*, x_S(t^*)$ and $x_I(t^*)$.
 2. Compute the BEST policy $c^*(t^*)$ using (22).
 3. Set $u(t) = c^*(t^*)$, for all $t \geq t^*$.
 4. Return to step 2 if β increases or θ decreases.
-

Proposition 1 states that the peak of $x_I(t)$ is uniquely determined by the BEST policy $c^*(t^*)$, where the peak is achieved at time t^* . Therefore, Algorithm 1 can be used to set the peak time t^* once parameters β, γ , and θ are learned from the data.

Remark 1. Requiring that β must not increase and θ must not decrease in the interval (t^*, t_1) for some $t_1 > t^*$ is necessary for the BEST policy. It is thus important to keep the external conditions that determine the values of β and θ either constant or such that β decreases (e.g., through the implementation of lockdown) and/or θ increases (e.g., through efficient contact tracing).

Remark 2. The case where β decreases and/or θ increases at some time $t_1 > t^*$ has the effect of speeding up the suppression of the epidemic under BEST policy.

Remark 3. From (1b), we can note that if $x_S(t)/N < \gamma/\beta$, the epidemic naturally decreases. In this case, doing no testing $u(t) = 0$ is the BEST policy, which, by definition, gives a minimum number of tests to be performed in order to stop the growth of the infected population x_I . However, if testing is resumed in this case, i.e., $u(t) > 0$, it will further speed up the decrease of the infected population.

5.1.1. Evaluation of the BEST policy

Giving data from France, we first compute $c^*(t^*)$ for different values of t^* from January 24 to March 13. Figure 14 shows the number of tests per day required by the BEST policy if it is implemented on day k and the corresponding value of peak of infected cases $x_I(k^*)$. One can note that the later BEST is applied the higher is the required number of tests. An exponential increase can even be observed from February 28 which corresponds to an acceleration of the infection. Now, we consider a scenario where BEST is implemented on

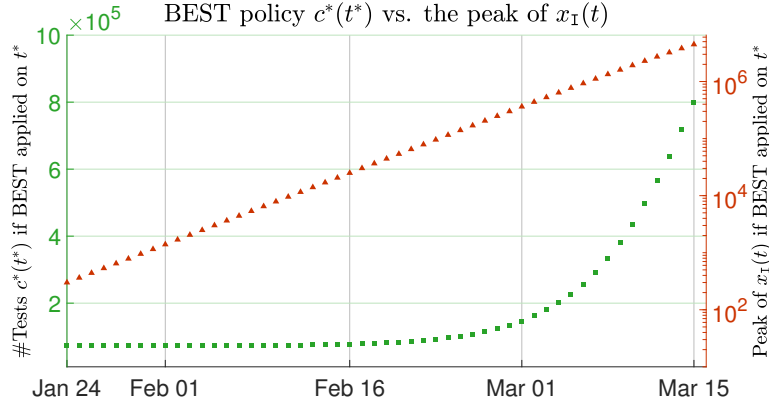


Figure 14: Number of tests per day required by the BEST policy (left y-axis, green) vs. peak of infection (right y-axis, red, in logscale) for an implementation day t^* .

March 01. Figure 15 depict the number of active cases when $u(t)$ is the actual testing scenario (see Figure 6) and when $u(t)$ is given by BEST. To evaluate BEST, we use $u(t)$ as given by recorded data before March 01 then use $u(t)$ given by BEST from March 01. In the first case, the peak of the infected population $x_I(t)$, which are the active undiagnosed cases, is about 6 million. In

the second case, the peak of infected population in this case is 363,169. The required number of tests per day to be performed for the implementation of BEST on March 01 is $c^* \approx 147,000$.

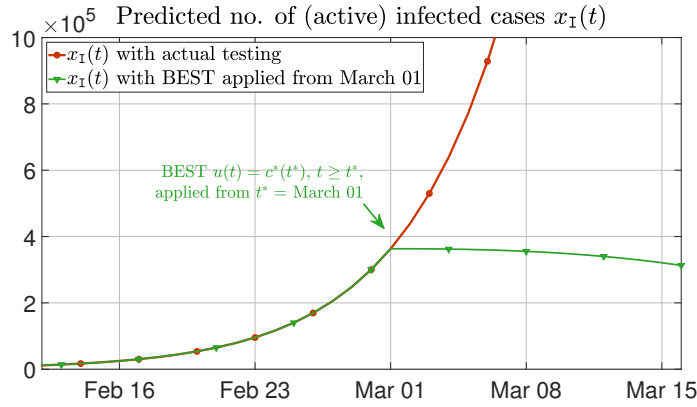


Figure 15: Predicted number of infected cases $x_I(t)$: actual testing scenario vs. BEST

The impact in terms of ICU occupation and number of deaths is now evaluated using the equations (20) and (21) respectively. The results are illustrated in Figure 16 and 17.

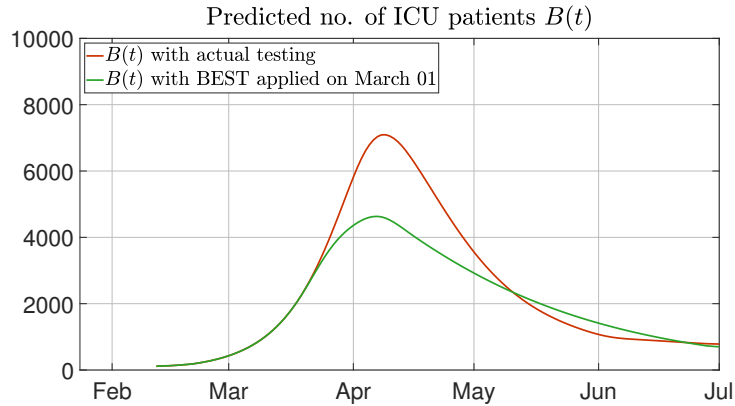


Figure 16: The prediction of the number of active ICU cases $B(t)$: actual scenario vs. BEST policy.

We observe that the peak of the number of active ICU patients could have been reduced by 34.71% and the number of deaths could have been reduced by 74.45% if the BEST policy was applied from March 01, 2020.

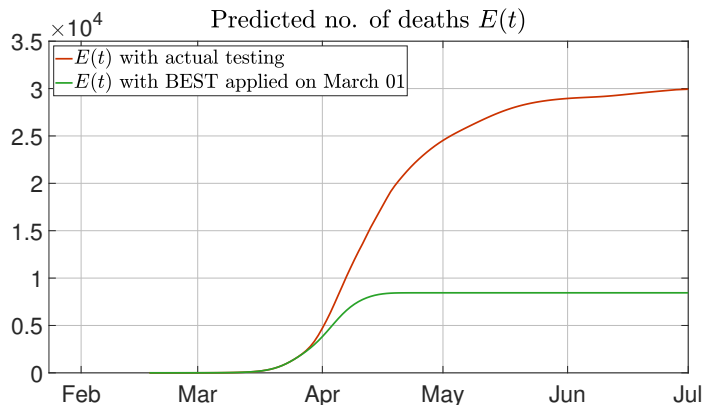


Figure 17: The prediction of the cumulative number of deaths $E(t)$: actual scenario vs. BEST policy.

5.2. Constant optimal strategy for testing

Consider now that the total stockpile of tests is limited and given by r_{\max} . In this case, the testing rate should be chosen carefully so as to not finish the stockpile of tests too early by performing too many tests per day, or to be unable to control the epidemic spread by performing too few tests per day. In the former case, when the stockpile of tests finishes too early due to intensive testing, the infections, even if significantly reduced in the beginning, will start to spread again and result in a second wave of the epidemic with an infection peak much higher than before. On the other hand, if the number of tests performed per day is too little, this would result in the infected population reaching a very high peak during the first wave of the epidemic, which could challenge the available medical facilities of a country. Thus, in this section, we determine a constant optimal allocation of a limited stockpile of tests r_{\max} such that both peaks of the infected population are minimized.

Given the total stockpile of tests r_{\max} , we assume that the number of tests performed per day is given by

$$u(t) = \begin{cases} C, & \text{if } 0 \leq t \leq T \\ 0, & \text{if } t > T \end{cases} \quad (23)$$

where the time period $T := r_{\max}/C$ represent the duration of the testing policy once C is determined.

Definition 2 (COST). The constant optimal strategy for testing (COST) with the total stockpile of tests r_{\max} is the policy of class (23) that minimizes the peak of the infected population $x_I(t)$.

In other words, the COST policy allocates the limited stockpile of tests r_{\max} as C tests per day for the time interval $[0, T]$, where T is measured in days, such that the maximum value of the infected population $x_I(t)$ is minimized.

In the following, we first study the SIDUR model in a new coordinate of ‘infection’ time ξ with the aim of finding analytic solutions. Then, we use those solutions to compute the two peak values of $x_{\text{I}}(t)$, where the first peak arrives at $t \leq T$ and the second at $t > T$. Finally, we compute the optimal value of C that minimizes those peak values by equating them, which is depicted in Figure 18.

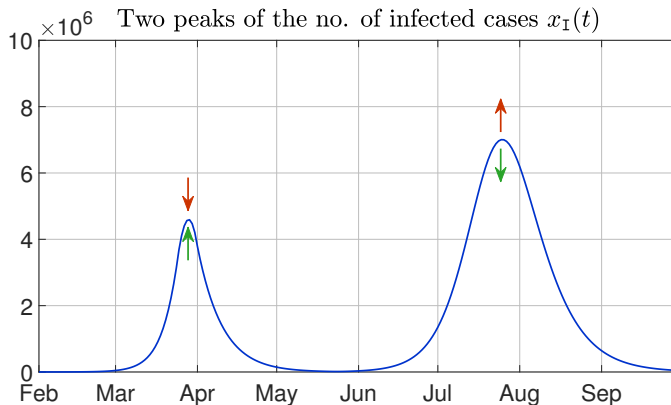


Figure 18: Under the limited stockpile of tests r_{\max} , the number of infected cases $x_{\text{I}}(t)$ will have two peaks. The first peak happens before the stockpile of tests is finished and the second peak happens after the stockpile finishes. The second peak is the result of the second wave of the epidemic during which no testing is done. If the testing policy is chosen such that the first (respectively, second) peak is reduced, then the second (respectively, first) peak will increase. The minimum of both peaks is achieved when the peak values of both waves of the epidemic are equal. This is ensured by COST policy C .

5.2.1. Partial solution to the SIDUR model

The SIDUR model has an interesting property: most of the terms on the right-hand side of the model (1) depend linearly on $x_{\text{I}}(t)$. This gives an intuition that the epidemic process goes faster when the number of infected people x_{I} is high. Indeed, this idea was used very early to analyze the basic SIR model, see [9]. It is possible to define a new time variable ξ , which we call *infection time*, as an integral of undetected infected people up to time t , i.e.,

$$d\xi = x_{\text{I}}(t)dt.$$

This means exactly that the speed of the processes in the new infection time is proportional to x_{I} . The ‘real’ time can be reconstructed from the infection time by:

$$t(\xi) = \int_0^{\xi} \frac{d\xi'}{x_{\text{I}}(\xi')}. \quad (24)$$

Note that the infection timescale preserves the peak value of epidemic, which is of interest to us. That is,

$$\frac{dx_{\text{I}}}{dt} = \frac{dx_{\text{I}}}{d\xi} \frac{d\xi}{dt} = \frac{dx_{\text{I}}}{d\xi} x_{\text{I}}$$

and $dx_{\text{I}}/dt = 0$ implies $dx_{\text{I}}/d\xi = 0$ and vice versa (as long as $x_{\text{I}} \neq 0$). Therefore, one can perform a model analysis based on the infection timescale ξ . Moreover, by writing the equations in ξ , the partial solution of SIDUR model can be obtained analytically.

Proposition 2. *Evolution of the unidentified recovered population x_{U} is affine with respect to the infection time ξ , whereas the susceptible population x_{S} is exponential with respect to ξ . That is,*

$$x_{\text{U}}(\xi) = x_{\text{U}}(0) + \gamma\xi = \gamma\xi \quad (25)$$

$$x_{\text{S}}(\xi) = x_{\text{S}}(0)e^{-\frac{\beta}{N}\xi}. \quad (26)$$

Proof. It is straightforward to see that the solution of (1d) is given by (25). On the other hand, (1a) yields

$$\frac{dx_{\text{S}}(t)}{d\xi} = -\frac{\beta}{N}x_{\text{S}}(t)$$

whose solution is given by (26). \square

It is not so trivial to find analytic solutions for x_{I} , x_{D} , and x_{R} . However, under certain hypothesis on the testable population, it is possible to get a reasonable approximation of the solutions.

Assumption 5. The testable population x_{T} is approximated as

$$x_{\text{T}}(t) \approx (1 - \theta)N. \quad (27)$$

This makes it possible to find an analytic solution for the infected population $x_{\text{I}}(t)$.

Proposition 3. *Evolution of the infected population x_{I} under Assumption 5 is given by*

$$x_{\text{I}}(\xi) = x_{\text{I}}(0) + x_{\text{S}}(0) \left(1 - e^{-\frac{\beta}{N}\xi}\right) - \frac{1}{(1 - \theta)N} \int_0^{\xi} u(s)ds - \gamma\xi. \quad (28)$$

Proof. Equation (28) can be directly obtained by integrating (1b), if one first transforms it using the infection time and then substitutes the definition for x_{T} from Assumption 5 and the solution for x_{S} from (26). \square

Using (23), we can now write a complete evolution of the infected population:

$$x_I(\xi) = \begin{cases} x_I(0) + x_S(0) \left(1 - e^{-\frac{\beta}{N}\xi}\right) - \frac{C\xi}{(1-\theta)N} - \gamma\xi, & \xi < \xi^* \\ x_I(0) + x_S(0) \left(1 - e^{-\frac{\beta}{N}\xi}\right) - \frac{C\xi^*}{(1-\theta)N} - \gamma\xi, & \xi \geq \xi^* \end{cases} \quad (29)$$

where ξ^* denotes the moment of stopping the testing in the infection time coordinates, which by definition of the infection time is given by an implicit formula

$$\int_0^{\xi^*} \frac{d\xi}{x_I(0) + x_S(0) \left(1 - e^{-\frac{\beta}{N}\xi}\right) - \frac{C\xi}{(1-\theta)N} - \gamma\xi} = T = \frac{r_{\max}}{C} \quad (30)$$

derived from the relation

$$\int_0^{\xi^*} \frac{d\xi}{x_I(\xi)} = T$$

where we substitute the solution for $x_I(\xi)$ given by (28), and the fact that $u(t) = C$ during the testing interval $[0, T]$.

5.2.2. Analysis of the evolution of infection peaks

From the definition of the effective reproductive number (13), for convenience, we can introduce two reproduction numbers at the origin: R_C with the COST policy C and R_W without testing:

$$R_C = \frac{x_S(0)\beta}{\frac{C}{1-\theta} + \gamma N}, \quad R_W = \frac{x_S(0)\beta}{\gamma N}, \quad R_W > R_C. \quad (31)$$

In general the evolution of the infected population given by (29) can have at most two peaks, one during testing and one after stopping. To have guarantees that the peaks exist, we further state the following assumption:

Assumption 6. The COST policy C does not suppress the epidemic at the origin, i.e., $R_C > 1$.

To explain this assumption, suppose in contrary that the COST policy C suppresses the epidemic at the origin. However, such a policy will not develop the herd immunity, and, in the case of the limited stockpile of tests r_{\max} , the second peak will arrive once the testing is stopped. Therefore, it is reasonable to develop at least a partial herd immunity using the controlled number of tests C in the first wave of epidemic.

When the two peaks of the infected population exist, their values can be obtained by setting the derivatives of (29) with respect to ξ to zero:

$$\begin{aligned} \frac{dx_{I,peak1}}{d\xi} &= x_S(0) \frac{\beta}{N} e^{-\frac{\beta}{N}\xi_{peak1}} - \frac{C}{(1-\theta)N} - \gamma = 0 \\ \frac{dx_{I,peak2}}{d\xi} &= x_S(0) \frac{\beta}{N} e^{-\frac{\beta}{N}\xi_{peak2}} - \gamma = 0 \end{aligned} \quad (32)$$

where the peaks positions in the infection time ξ are given as:

$$\xi_{peak1} = \frac{N}{\beta} \ln \frac{x_S(0)\beta}{\frac{C}{1-\theta} + \gamma N}, \quad \xi_{peak2} = \frac{N}{\beta} \ln \frac{x_S(0)\beta}{\gamma N}. \quad (33)$$

Writing them in terms of reproduction numbers R_C and R_W , we obtain

$$\xi_{peak1} = \frac{N}{\beta} \ln R_C, \quad \xi_{peak2} = \frac{N}{\beta} \ln R_W. \quad (34)$$

Since $R_W > R_C > 1$, by Assumption 6, we get $\xi_{peak2} > \xi_{peak1} > 0$. Further one should notice that the peak ξ_{peak1} occurs only if $\xi_{peak1} < \xi^*$, otherwise it is ill-defined. Similarly, $\xi_{peak2} > \xi^*$.

The peaks values themselves are given by

$$\begin{aligned} x_{I,peak1} &= x_I(0) + x_S(0) \left(1 - \frac{1}{R_C}\right) - x_S(0)R_C \ln R_C, \\ x_{I,peak2} &= x_I(0) + x_S(0) \left(1 - \frac{1}{R_W}\right) - x_S(0)R_W \ln R_W - \frac{C\xi^*}{(1-\theta)N}, \end{aligned} \quad (35)$$

It is obvious that $x_{I,peak1}$ decreases as C increases, because $x_{I,peak1}$ represents a peak under still performed testing.

Further, it is possible to prove that $C\xi^*$ decreases as C increases for all sufficiently large C . Indeed, (30) can be rewritten as

$$C \int_0^{\xi^*} \frac{d\xi}{x_I(\xi)} = r_{\max}.$$

Taking the derivative with respect to ξ^* and substituting again (30) in place of the integral, we see that

$$\frac{dC}{d\xi^*} \frac{r_{\max}}{C} + C \frac{1}{x_I(\xi^*)} = 0,$$

which is the same as

$$\frac{d\xi^*}{dC} = -\frac{r_{\max}x_I(\xi^*)}{C^2}.$$

Therefore,

$$\frac{d(C\xi^*)}{dC} = \xi^* - \frac{r_{\max}x_I(\xi^*)}{C} = \int_0^T [x_I(\tau) - x_I(T)] d\tau,$$

where ξ^* is rewritten by the definition of the infection time (24), while r_{\max}/C is represented by the integral of 1 over the time $T = r_{\max}/C$. The result is negative unless $x_I(T)$ is smaller than an average number of infected people all the time before, which means that T is long after the peak has come. Thus

for all reasonable sufficiently large C such that T is not too large C , $x_{I,peak1}$ decreases, while $x_{I,peak2}$ increases.

This property can be used to optimize a maximum between two peaks. Indeed, a maximum between a decreasing and an increasing function is minimized when they are equal each other. Thus, we equate two peak values in (35) and solve it with respect to ξ^* , obtaining an optimality condition as

$$\xi^* = \frac{N}{\beta} \left(1 + \ln R_C - \frac{R_C}{R_W - R_C} \ln \frac{R_W}{R_C} \right). \quad (36)$$

Equation (36) provides an optimality condition which, being combined with the connection between ξ^* and C given by (30), constitutes a complete system determining the COST C .

It is possible to show that the optimal strategy provides switching in between of two peaks ξ_{peak1} and ξ_{peak2} , thus both peaks exist. Indeed, first we prove a simple lemma:

Lemma 4. *For any $x \geq 1$*

$$x - 1 \geq \ln x \geq 1 - \frac{1}{x},$$

and the equality is possible only if $x = 1$.

Proof. First of all, it is obvious that for $x = 1$ all three parts of the inequality are equal to zero. Further, let us take the derivative of this inequality:

$$1 \geq \frac{1}{x} \geq \frac{1}{x^2},$$

which is always true for all $x \geq 1$, and it holds strictly for all $x > 1$. This concludes the proof. \square

Now, consider the difference $\xi^* - \xi_{peak1}$:

$$\begin{aligned} \frac{\beta}{N} (\xi^* - \xi_{peak1}) &= 1 - \frac{R_C}{R_W - R_C} \ln \frac{R_W}{R_C} \\ &= 1 - \frac{1}{x-1} \ln x \\ &= \frac{1}{x-1} (x-1 - \ln x) \geq 0, \end{aligned}$$

where we denote $x = R_W/R_C$, and the last inequality comes from Lemma 4. Thus the first peak is well-defined. Further, writing the difference $\xi_{peak2} - \xi^*$ in the same manner and using the same definition for x , we see that

$$\begin{aligned} \frac{\beta}{N} (\xi_{peak2} - \xi^*) &= \left(\frac{R_C}{R_W - R_C} + 1 \right) \ln \frac{R_W}{R_C} - 1 \\ &= \frac{x}{x-1} \ln x - 1 \\ &= \frac{x}{x-1} \left(\ln x - 1 + \frac{1}{x} \right) \geq 0. \end{aligned}$$

Thus we have proven that $0 < \xi_{peak1} \leq \xi^* \leq \xi_{peak2}$, which means the stopping of testing happens in between of the two peaks.

5.2.3. Computation of the COST policy

Finally, the COST policy can be obtained by solving (30) with respect to C , where (36) are used as an upper limit of the integral. This can be formally stated as follows:

Proposition 5. *Consider the testing rate defined by (23). Then, under Assumptions 5 and 6, the COST policy C is given by solving (30)–(36), i.e.,*

$$0 = \frac{r_{\max}}{C} - \int_0^{\xi^*} \frac{d\xi}{x_I(0) + x_S(0) \left(1 - e^{-\frac{\beta}{N}\xi}\right) - \frac{C\xi}{(1-\theta)N} - \gamma\xi} \quad (37)$$

$$\xi^* = \frac{N}{\beta} \left(1 + \ln \frac{x_S(0)\beta}{\frac{C}{1-\theta} + \gamma N}\right) - \frac{(1-\theta)\gamma N^2}{C\beta} \ln \frac{\frac{C}{1-\theta} + \gamma N}{\gamma N}. \quad (38)$$

The time duration for positive testing rate (23) is given by $T = r_{\max}/C$. Note that, by construction, a solution to the system of equations (37) and (38) exists and is unique as long as Assumption 6 holds. This system of equations can be solved numerically for C by using the Newton's method (if the initial estimate C_0 is chosen sufficiently close to the real optimal value C). The process is sketched in Algorithm 2.

5.2.4. Evaluation of COST policy

We first validate Assumption 5. This assumption leads to a new set of model parameters given by Table 3. With these parameters, the model fitting is illustrated in Figure 19.

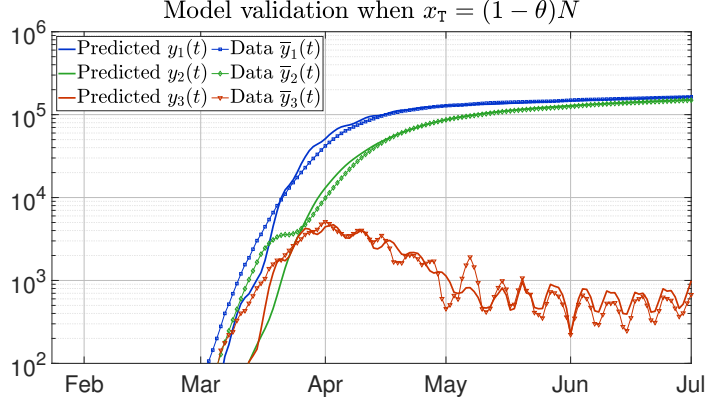


Figure 19: Model fitting under Assumption 5 and parameter values of Table 3.

Similar to BEST, we evaluate the COST through the prediction of the number of active ICU cases $B(t)$ and the cumulative number of deaths $E(t)$ using

Algorithm 2: Computation of the COST policy.

1. Take the initial estimate for COST C_0 .
2. On n -th iteration step, given C_n , find ξ_n^* by (38).
3. Find the value of the right-hand side of (37). This can be done by computing the integral numerically. Denote the result as $f(C_n)$.
4. Find the derivative $f'(C_n)$ of the right-hand side of (37). This can be done analytically:

$$f'(C_n) = -\frac{r_{\max}}{C_n^2} - \frac{1}{x_{\mathbf{I}}(\xi_n^*)} \frac{d\xi_n^*}{dC_n}$$

where $x_{\mathbf{I}}(\xi_n^*)$ is computed from (29) and, from (38),

$$\frac{d\xi_n^*}{dC_n} = -\frac{N}{\beta C_n} + \frac{(1-\theta)\gamma N^2}{C_n^2 \beta} \ln \frac{\frac{C_n}{1-\theta} + \gamma N}{\gamma N}.$$

5. Once $f(C_n)$ and $f'(C_n)$ are computed, update the COST by the Newton's method:

$$C_{n+1} = C_n - f(C_n)/f'(C_n).$$

6. Repeat, from step 2, until the desired accuracy is achieved.
-

Infection rate	$\beta_1 = 0.2643$	$\beta_2 = 0.0006$	$\beta_3 = 0.0642$
Testing specificity	$\theta_1 = 0.9415$	$\theta_2 = 0.7993$	
Recovery rate	$\gamma = 0.0542$		
Removal rate	$\rho = 0.0499$		

Table 3: Estimated parameter values under Assumption 5.

equations (20) and (21), respectively. However, we fit these equations to the data $\bar{B}(k)$ and $\bar{E}(k)$ by using the model parameters of Table 3 that are estimated under the Assumption 5.

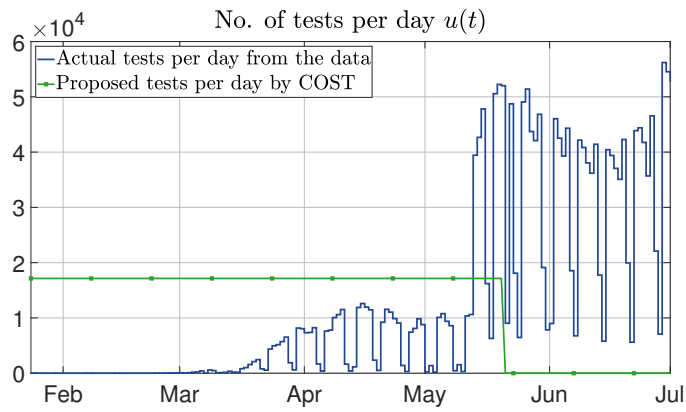


Figure 20: The comparison between the actual control input $u(t)$ versus the proposed control input $u(t)$ required by COST.

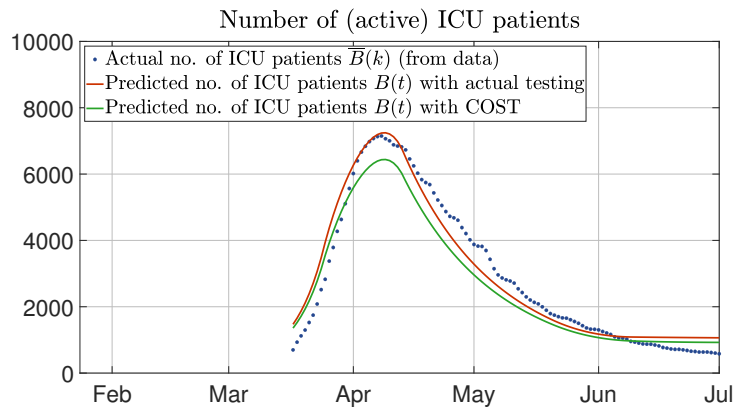


Figure 21: The comparison between the predicted number of active ICU cases $B(t)$ in the actual testing scenario and with COST policy.

Assuming the total stockpile of tests $r_{\max} = 2,038,037$, which is the total

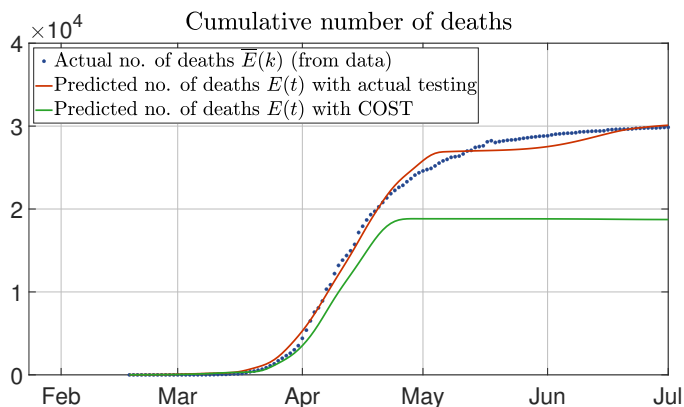


Figure 22: The comparison between the predicted cumulative number of deaths $E(t)$ in the actual testing scenario and with COST policy.

number of tests performed from January 24 to July 01, 2020, in France, we obtain the optimal number of tests to be performed per day as $C = 17,144$ from Algorithm 2, where the terminal time $T = 118$ (i.e., May 20). This is illustrated in Figure 20 along with the actual number of tests per day from the data.

Using the actual number of tests per day, we fit the model outputs (20) and (21) with the data of the number of active ICU cases $\bar{B}(k)$ and the cumulative number of deaths $\bar{E}(k)$, where the parameter values are given in Appendix B, Table B.5. Then, we predict the number of active ICU cases $B(t)$ and the cumulative number of deaths when the COST is applied instead of actual number of tests, as illustrated in Figure 21 and Figure 22, respectively. We observe that COST reduces the peak of $B(t)$ by 11.12% and the total number of cumulative deaths $E(t)$ by 37.52%.

Notice that the total stockpile of tests in the actual testing and COST is the same. However, most of the tests in the actual testing are consumed after May 13, whereas all the tests in COST are allocated equally in the time interval January 24 to May 20. The number of tests performed per day in the actual testing exceeds 50,000 per day after May 13, which is about three times more than what is required by COST, i.e., 17,144. Thus, COST is optimal in a sense that it is practical, it decreases the burden on medical facilities, and it reduces the number of deaths significantly.

6. Concluding remarks

We proposed a SIDUR model for the control of the COVID-19 epidemic through the number of RT-PCT tests performed per day. Such tests enable the government to diagnose and isolate the infected people from the susceptible population. We estimated and validated the model on the French COVID-19 data,

and proposed two testing policies to control the epidemic: 1) best effort strategy for testing (BEST) and 2) constant optimal strategy for testing (COST). BEST provides the minimum number of tests to be performed from a certain day onward in order to make the increasing infected population non-increasing immediately. That is, it changes the course of epidemic from spreading to non-spreading. On the other hand, COST considers a limited stockpile of tests that are optimally allocated in a time interval starting from the beginning of the epidemic in order to minimize the peak of infected population.

The control input in SIDUR model corresponds to the number of RT-PCR tests performed per day. However, another type of test, a serology test, which is not considered in the current model because of the unavailability of its data, is also very important. A serology test determines the relevant antibodies in a subject's serum in order to detect whether he/she was infected in the past. By performing serology tests on the testable population, one can detect the unidentified recovered population and transfer them in the identified removed compartment of the model. This reduces the size of the testable population, which in turn increases the testing specificity of RT-PCR tests. In other words, the serology tests complement the RT-PCR tests [34, 35]. Therefore, as a future prospect, it will be interesting to consider two control inputs corresponding to both types of test in the SIDUR model.

The model is estimated and validated by fitting the model outputs with the available data of France. This allows us to predict the unmeasured states of the model. However, there is no certainty whether the predicted states correspond to the reality. Another prospect is to design an observer for the SIDUR model in order to estimate the true states of the model.

Both BEST and COST policies are easy to compute and implement, however they are static. Thus, their influence on the control of epidemic is limited. In future, it will be interesting to solve a finite/infinite-horizon optimal control problem to minimize the peak and/or cumulative number of the infected population by a dynamic control input.

Acknowledgment

This work is partially supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-AdG no. 694209, Scale-FreeBack, website: <http://scale-freeback.eu/>) and by Inria, France, in the framework of the COVID-19 fast-track research program.

Data and code availability

The raw data used in this paper is available at <https://www.data.gouv.fr/fr/datasets/> with the following pointers: Chiffres-clés concernant l'épidémie de COVID-19 en France, Données relatives aux tests de dépistage de COVID-19 réalisés en laboratoire de ville, and Données relatives aux résultats des tests

virologiques COVID-19 (SI-DEP). The imputed data and codes are available at <http://scale-freeback.eu/>.

Appendix A. Particle swarm optimization

In this appendix, we briefly describe the particle swarm optimization (PSO) algorithm, [52], which is considered to be one of the most powerful algorithms. It considers a foraging swarm of n particles who collectively search for an optimal solution of (18) in the parameter space. At time step $h = 0, 1, 2, \dots$, each particle i visits a position \hat{p}_h^i by moving with velocity v_h^i . Initially, when $h = 0$, the positions \hat{p}_0^i , for all $i \in \{1, \dots, n\}$, are chosen randomly in the parameter space and the velocities $v_0^i = 0$. Each particle i stores its personal best pair $(\hat{p}_h^{i*}, J_h^{i*})$ and the social best pair (s_h^*, J_h^{s*}) in memory, where $J_h^{i*} = \mathcal{J}(\hat{p}_h^{i*})$ and $J_h^{s*} = \mathcal{J}(s_h^*)$ are the costs (19) of personal best position \hat{p}_h^{i*} and social best position $s_h^* = \arg \min_{\hat{p}_h^{i*}, i \in \{1, \dots, n\}} \mathcal{J}(\hat{p}_h^{i*})$, respectively. Notice that $J_h^{s*} \leq J_h^{i*}$ for all $i \in \{1, \dots, n\}$. The personal best pair of a particle corresponds to the best position in the parameter space it has visited so far. The social best pair, on the other hand, corresponds to the best position in the parameter space that anyone in the swarm has visited so far.

At every time step, each particle updates its velocity, position, its personal best pair, and the social best pair. The velocity and position are updated as follows:

$$\begin{aligned} v_{h+1}^i &= wv_h^i + c_1 r_{h,1} (\hat{p}_h^{i*} - \hat{p}_h^i) + c_2 r_{h,2} (s_h^* - \hat{p}_h^i) \\ \hat{p}_{h+1}^i &= \hat{p}_h^i + v_{h+1}^i \end{aligned} \quad (\text{A.1})$$

where w is the inertia weight, c_1, c_2 are the acceleration coefficients, and $r_{h,1}, r_{h,2}$ are uniformly distributed random numbers in $[0, 1]$ generated at each time step h . There are many ways of choosing these parameters [55, 56, 57].

Each particle i computes the cost $J_{h+1}^i = \mathcal{J}(\hat{p}_{h+1}^i)$ at its current position and updates its personal best pair as

$$(\hat{p}_{h+1}^{i*}, J_{h+1}^{i*}) = \begin{cases} (\hat{p}_{h+1}^i, J_{h+1}^i), & \text{if } J_{h+1}^i \leq J_h^{i*} \\ (\hat{p}_h^{i*}, J_h^{i*}), & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

Each particle i then communicates its personal best pair with all the other particles and each of them finds the social best pair for time $h + 1$ as

$$(s_{h+1}^s, J_{h+1}^s) = (\hat{p}_{h+1}^b, J_{h+1}^b)$$

where $b = \arg \min_{j \in \{1, \dots, n\}} J_{h+1}^j$. Finally, the social best pair is updated as

$$(s_{h+1}^*, J_{h+1}^{s*}) = \begin{cases} (s_{h+1}^s, J_{h+1}^s), & \text{if } J_{h+1}^s \leq J_h^{s*} \\ (s_h^*, J_h^{s*}), & \text{otherwise.} \end{cases} \quad (\text{A.3})$$

Parameters of $B(t)$	$b_1 = -0.54 \times 10^4$	$b_2 = 1.25 \times 10^4$
Parameters of $E(t)$	$e_1 = 4.14 \times 10^4$	$e_2 = 7.92 \times 10^5$
	$e_3 = -1.27 \times 10^7$	$e_4 = 9.04 \times 10^7$
	$e_5 = -3.63 \times 10^8$	$e_6 = 8.81 \times 10^8$
	$e_7 = -1.32 \times 10^9$	$e_8 = 1.19 \times 10^9$
	$e_9 = -5.93 \times 10^8$	$e_{10} = 1.25 \times 10^8$

Table B.4: Estimated parameters b_1 and b_2 in (20) and e_i , for $i = 1, \dots, 10$, in (21) when the testable population x_{T} is given by (2).

Parameters of $B(t)$	$b_1 = 8.87 \times 10^{-4}$	$b_2 = 1.62$
Parameters of $E(t)$	$e_1 = 4.03 \times 10^4$	$e_2 = -1.62 \times 10^6$
	$e_3 = 2.76 \times 10^7$	$e_4 = -2.16 \times 10^8$
	$e_5 = 9.27 \times 10^8$	$e_6 = -2.34 \times 10^9$
	$e_7 = 3.59 \times 10^9$	$e_8 = -3.28 \times 10^9$
	$e_9 = 1.63 \times 10^9$	$e_{10} = -3.44 \times 10^8$

Table B.5: Estimated parameters b_1 and b_2 in (20) and e_i , for $i = 1, \dots, 10$, in (21) when the testable population x_{T} is given by (27).

Appendix B. Parameter values for curve fitting of the number of ICU cases and deaths

Appendix C. Decreasing property of the testable population

Lemma 6. *The testable population x_{T} decreases on any interval on which x_{I} is decreasing and θ is non-decreasing.*

Proof. Let us consider an interval (t, t') , $t < t'$ on which x_{I} is decreasing while θ is non-decreasing. First, one can note that

$$\begin{aligned}
x_{\text{T}}(t) &= \theta(t)(x_{\text{I}}(t) + x_{\text{D}}(t) + x_{\text{R}}(t) - N) + (N - x_{\text{D}}(t) - x_{\text{R}}(t)) \\
&\geq \theta(t')(x_{\text{I}}(t) + x_{\text{D}}(t) + x_{\text{R}}(t) - N) + (N - x_{\text{D}}(t) - x_{\text{R}}(t)) \\
&= \theta(t')x_{\text{I}}(t) + (1 - \theta(t'))(N - x_{\text{D}}(t) - x_{\text{R}}(t)), \tag{C.1}
\end{aligned}$$

because $x_{\text{I}}(t) + x_{\text{D}}(t) + x_{\text{R}}(t) - N < 0$ and θ is non-negative and non-decreasing by assumption. Since x_{I} is supposed to be decreasing, then

$$\theta(t')x_{\text{I}}(t) \geq \theta(t')x_{\text{I}}(t').$$

On the other hand

$$\dot{N} - \dot{x}_{\text{D}} - \dot{x}_{\text{R}} = -u \frac{x_{\text{I}}}{x_{\text{T}}} < 0,$$

meaning that $N - x_{\text{D}} - x_{\text{R}}$ is a decreasing function. As $1 - \theta(t') \geq 0$, one gets

$$(1 - \theta(t'))(N - x_{\text{D}}(t) - x_{\text{R}}(t)) \geq (1 - \theta(t'))(N - x_{\text{D}}(t') - x_{\text{R}}(t')).$$

Adding the two inequalities, one deduces from (C.1) that $x_{\mathcal{T}}(t) \geq x_{\mathcal{T}}(t')$ whenever $x_{\mathcal{I}}$ is decreasing on (t, t') . A tighter examination shows that, as both expressions $\theta(t')$ and $1 - \theta(t')$ cannot be zero together, at least one of the two terms of the sum indeed decreases between t and t' . Therefore, $x_{\mathcal{T}}(t) > x_{\mathcal{T}}(t')$. \square

References

- [1] X. Cao, COVID-19: Immunopathology and its implications for therapy, *Nature reviews immunology* 20 (5) (2020) 269–270.
- [2] L. R. Baden, E. J. Rubin, COVID-19 — The search for effective therapy, *New England Journal of Medicine* 382 (19) (2020) 1851–1852.
- [3] N. V. Loayza, S. Pennings, Macroeconomic policy in the time of COVID-19: A primer for developing countries, *World Bank Research & Policy Briefs* (28) (2020).
- [4] T. Hale, N. Angrist, B. Kira, A. Petherick, T. Phillips, S. Webster, Variation in government responses to COVID-19, *Blavatnik School of Government Working Paper BSG-WP-2020/032* (2020).
- [5] T. Hale, S. Webster, A. Petherick, T. Phillips, B. Kira, Oxford COVID-19 government response tracker, *Blavatnik School of Government* (2020).
URL <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>
- [6] A. Glover, J. Heathcote, D. Krueger, J.-V. Ríos-Rull, Health versus wealth: On the distributional effects of controlling a pandemic, *NBER Working Paper No. 27046* (2020).
- [7] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al., Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce covid19 mortality and healthcare demand, *Imperial College London* (16-03-2020) (2020). doi:10.25561/77482.
- [8] P. Walker, C. Whittaker, O. Watson, M. Baguelin, K. Ainslie, S. Bhatia, S. Bhatt, A. Boonyasiri, O. Boyd, L. Cattarino, et al., Report 12: The global impact of COVID-19 and strategies for mitigation and suppression, *Imperial College London* (26-03-2020) (2020). doi:10.25561/77735.
- [9] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London Series A* 115 (772) (1927) 700–721.
- [10] H. Hethcote, The mathematics of infectious diseases, *SIAM Review* 42 (1) (2000) 599–653.

- [11] G. Massonis, J. R. Banga, A. F. Villaverde, Structural identifiability and observability of compartmental models of the COVID-19 pandemic, arXiv:2006.14295 (2020).
- [12] A. Khanafer, T. Başar, B. Gharesifard, Stability of epidemic models over directed graphs: A positive systems approach, *Automatica* 74 (2016) 126–134.
- [13] P. E. Paré, C. L. Beck, A. Nedić, Epidemic processes over time-varying networks, *IEEE Transactions on Control of Network Systems* 5 (3) (2018) 1322–1334.
- [14] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, T. Başar, Analysis, estimation, and validation of discrete-time epidemic processes, *IEEE Transactions on Control Systems Technology* 28 (1) (2018) 79–93.
- [15] V. Colizza, A. Vespignani, Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations, *Journal of theoretical biology* 251 (3) (2008) 450–467.
- [16] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, *Reviews of modern physics* 87 (3) (2015) 925.
- [17] F. Della Rossa, D. Salzano, A. Di Meglio, F. De Lellis, M. Coraggio, C. Calabrese, A. Guarino, R. Cardona-Rivera, P. De Lellis, D. Liuzza, et al., A network model of italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic, *Nature Communications* 11 (1) (2020) 1–9.
- [18] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S. Musa, M. Wang, Y. Cai, W. Wang, L. Yang, D. He, A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in wuhan, china with individual reaction and governmental action, *International Journal of Infectious Diseases* 93 (2020) 211–216.
- [19] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, *PLoS One* 15 (e0230405) (2020).
- [20] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, *Nature medicine* 26 (2020) 855–860.
- [21] Z. Liu, P. Magal, O. Seydi, G. Webb, Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions, *Biology* 9 (3) (2020) 50.

- [22] A. Ducrot, P. Magal, T. Nguyen, G. Webb, Identifying the number of unreported cases in SIR epidemic models, *Mathematical medicine and biology: A journal of the IMA* 37 (2) (2020) 243–261.
- [23] J. Köhler, L. Schwenkel, A. Koch, J. Berberich, P. Pauli, F. Allgöwer, Robust and optimal predictive control of the COVID-19 outbreak, arXiv:2005.03580 (2020).
- [24] M. M. Morato, I. M. Pataro, M. V. da Costa, J. E. Normey-Rico, A parametrized nonlinear predictive control strategy for relaxing COVID-19 social distancing measures in Brazil, arXiv:2007.09686 (2020).
- [25] A. Perkins, G. Espana, Optimal control of the COVID-19 pandemic with non-pharmaceutical interventions, medRxiv 2020.04.22.20076018 (2020).
- [26] F. Casella, Can the COVID-19 epidemic be controlled on the basis of daily test reports?, *IEEE Control Systems Letters* 5 (3) (2020) 1079–1084.
- [27] L. E. Olivier, S. Botha, I. K. Craig, Optimized lockdown strategies for curbing the spread of COVID-19: A South African case study, arXiv:2006.16379 (2020).
- [28] F. E. Alvarez, D. Argente, F. Lippi, A simple planning problem for COVID-19 lockdown, NBER Working Paper No. 26981 (2020).
- [29] D. Acemoglu, V. Chernozhukov, I. Werning, M. D. Whinston, Optimal targeted lockdowns in a multi-group SIR model, NBER Working Paper No. 27102 (2020).
- [30] L. Brotherhood, P. Kircher, C. Santos, M. Tertilt, An economic model of the COVID-19 epidemic: The importance of testing and age-specific policies, CESifo Working Paper No. 8316 (2020).
- [31] P. Romagnani, G. Gnone, F. Guzzi, S. Negrini, A. Guastalla, F. Annunziato, S. Romagnani, R. De Palma, The COVID-19 infection: Lessons from the Italian experience, *Journal of Public Health Policy* (2020) 1–7.
- [32] M. Day, COVID-19: Identifying and isolating asymptomatic people helped eliminate virus in Italian village, *BMJ: British Medical Journal (Online)* 368 (2020).
- [33] Emergency use authorization (EUA) summary: COVID-19 RT-PCR test, Tech. rep., US Food and Drug Administration.
URL <https://www.fda.gov/media/136151/download>
- [34] D. de Walque, J. Friedman, R. Gatti, A. Mattoo, How two tests can help contain COVID-19 and revive the economy, *World Bank Research & Policy Briefs* (29) (2020).
- [35] A. K. Winter, S. T. Hegde, The important role of serology for COVID-19 control, *The Lancet Infectious Diseases* 20 (7) (2020) 758–759.

- [36] Advice on the use of point-of-care immunodiagnostic tests for COVID-19: Scientific brief, 8 april 2020, Technical documents, World Health Organization (2020).
- [37] M. S. Eichenbaum, S. Rebelo, M. Trabandt, The macroeconomics of testing and quarantining, NBER Working Paper No. 27104 (2020).
- [38] M. Salathé, C. L. Althaus, R. Neher, S. Stringhini, E. Hodcroft, J. Fellay, M. Zwahlen, G. Senti, M. Battegay, A. Wilder-Smith, et al., COVID-19 epidemic in Switzerland: On the importance of testing, contact tracing and isolation, *Swiss medical weekly* 150:w20225 (2020).
- [39] G. Chowell, P. Fenimore, M. Castillo-Garsow, C. Castillo-Chavez, SARS outbreaks in Ontario, Hong Kong and Singapore: The role of diagnosis and isolation as a control mechanism, *Journal of Theoretical Biology* 224 (1) (2003) 1–8.
- [40] C. Nowzari, V. M. Preciado, G. J. Pappas, Analysis and control of epidemics: A survey of spreading processes on complex networks, *IEEE Control Systems Magazine* 36 (1) (2016) 26–46.
- [41] C. Nowzari, V. M. Preciado, G. J. Pappas, Optimal resource allocation for control of networked epidemic models, *IEEE Transactions on Control of Network Systems* 4 (2) (2017) 159–169.
- [42] M. Pezzutto, N. B. Rossello, L. Schenato, E. Garone, Smart testing and selective quarantine for the control of epidemics, arXiv:2007.15412 (2020).
- [43] J. Ely, A. Galeotti, J. Steiner, Optimal test allocation, Tech. rep., Mimeo (2020).
- [44] F. Piguillem, L. Shi, Optimal COVID-19 quarantine and testing policies, CEPR Discussion Paper No. DP14613 (2020).
- [45] D. Berger, K. Herkenhoff, S. Mongey, An SEIR infectious disease model with testing and conditional quarantine, NBER Working Paper No. 26901 (2020).
- [46] A. Charpentier, R. Elie, M. Laurière, V. Tran, COVID-19 pandemic control: Balancing detection policy and lockdown intervention under ICU sustainability, arXiv:2005.06526v3 (2020).
- [47] K. J. Rothman, S. Greenland, T. L. Lash, *Modern epidemiology*, 3rd Edition, Lippincott Williams & Wilkins, 2008.
- [48] P. Van den Driessche, J. Watmough, Further notes on the basic reproduction number, in: *Mathematical epidemiology*, Springer, 2008, pp. 159–178.

- [49] J. Oh, J.-K. Lee, D. Schwarz, H. L. Ratcliffe, J. F. Markuns, L. R. Hirschhorn, National response to COVID-19 in the republic of korea and lessons learned for other countries, *Health Systems & Reform* 6 (1) (2020) e1753464.
- [50] L. Ljung, *System Identification: Theory for the user*, 2nd Edition, Prentice Hall PTR, Upper Saddle River, NJ-USA, 1999.
- [51] O. Nelles, *Nonlinear system identification: From classical approaches to neural networks and fuzzy models*, Springer-Verlag, Berlin Heidelberg, 2001.
- [52] J. Kennedy, R. Eberhart, Particle swarm optimization, in: *Proceedings of the International Conference on Neural Networks (ICNN)*, Vol. 4, IEEE, 1995, pp. 1942–1948.
- [53] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application, *Annals of internal medicine* 172 (9) (2020) 577–582.
- [54] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, china: a retrospective cohort study, *The Lancet* (2020).
- [55] M. Clerc, J. Kennedy, The particle swarm-explosion, stability, and convergence in a multidimensional complex space, *IEEE transactions on Evolutionary Computation* 6 (1) (2002) 58–73.
- [56] R. Poli, J. Kennedy, T. Blackwell, Particle swarm optimization: An overview, *Swarm intelligence* 1 (1) (2007) 33–57.
- [57] Z.-H. Zhan, J. Zhang, Y. Li, H. S.-H. Chung, Adaptive particle swarm optimization, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39 (6) (2009) 1362–1381.