

## Original Paper

**Cite this article:** Solari HG, Natiello MA (2021). Stochastic model for COVID-19 in slums: interaction between biology and public policies. *Epidemiology and Infection* **149**, e206, 1–15. <https://doi.org/10.1017/S0950268821001746>

Received: 6 January 2021  
Revised: 11 April 2021  
Accepted: 13 July 2021

**Key words:**

Markov-jump process; stochastic compartmental model; surveillance protocols; viraemic levels

**Author for correspondence:**

M. A. Natiello,  
E-mail: [mario.natiello@math.lth.se](mailto:mario.natiello@math.lth.se)

<sup>1</sup>Departamento de Física, FCEN-UBA and IFIBA-CONICET, Buenos Aires, Argentina and <sup>2</sup>Centre for Mathematical Sciences, Lund University, Lund, Sweden

**Abstract**

We present a mathematical model for the simulation of the development of an outbreak of coronavirus disease 2019 (COVID-19) in a slum area under different interventions. Instead of representing interventions as modulations of the parameters of a free-running epidemic, we introduce a model structure that accounts for the actions but does not assume the results. The disease is modelled in terms of the progression of viraemia reported in scientific studies. The emergence of symptoms in the model reflects the statistics of a nation-wide highly detailed database consisting of more than 62 000 cases (about a half of them confirmed by reverse transcription-polymerase chain reaction tests) with recorded symptoms in Argentina. The stochastic model displays several of the characteristics of COVID-19 such as a high variability in the evolution of the outbreaks, including long periods in which they run undetected, spontaneous extinction followed by a late outbreak and unimodal as well as bimodal progressions of daily counts of cases (second waves without ad-hoc hypothesis). We show how the relation between undetected cases (including the ‘asymptomatic’ cases) and detected cases changes as a function of the public policies, the efficiency of the implementation and the timing with respect to the development of the outbreak. We show also that the relation between detected cases and total cases strongly depends on the implemented policies and that detected cases cannot be regarded as a measure of the outbreak, being the dependency between total cases and detected cases in general not monotonic as a function of the efficiency in the intervention method. According to the model, it is possible to control an outbreak with interventions based on the detection of symptoms only in the case when the presence of just one symptom prompts isolation and the detection efficiency reaches about 80% of the cases. Requesting two symptoms to trigger intervention can be enough to fail in the goals.

**Introduction**

Ever since the emergence of coronavirus disease 2019 (COVID-19) [1], mathematical models have been proposed to examine, illustrate and forecast the possible evolution of the pandemic, as well as recommending public measures for managing it. Modelling epidemics has to deal with a variety of difficulties at different levels and the present pandemic is not an exception.

This study adopts a stochastic approach – following a line of thought developed over the last century [2–5] – that rests on counting populations in its natural form and evolving their numbers at characteristic events. In relation to them, ordinary differential equation models represent the evolution of average fractions of populations in a large-population limit [6–9].<sup>1</sup> In this respect, one of our aims is to explore, at least within the limited scope of the situation considered, the relevance of stochasticity in our perception of the pandemic.

A good part of the literature has addressed the phenomena of asymptomatic carriers of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [11–17]. Unfortunately, the label ‘asymptomatic’ has been used with different meanings, going from ‘not presenting the expected symptoms at the moment of infecting someone else’, as in [11, 13] to ‘never, in the course of the infection, presenting symptoms’ [18]. In all cases, asymptomatic and pre-symptomatic are considered as objective categories pertaining to the relation between the infected person and the infectious agent. The influence of the actions of the public health system and the perception of illness by the patients in building these categories has not been properly examined, thus preventing any improvement of these actions.

In previous modelling studies either asymptomatic carriers of SARS-CoV-2 have not been considered or they have been incorporated using an ad-hoc hypothesis, such as that the ratio between asymptomatic and symptomatic cases is constant (see, e.g. [19]). In contrast, our model incorporates a detection component based on what it is known of detection policies. Another sharp difference with an earlier study is that we model a variable contagiousness and not only a variable contagious period. Furthermore, intrinsic stochasticity is included

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

<sup>1</sup>For an epistemological discussion in the context of COVID-19, the reader may consult [(work in press) ref. 10].

in contrast to the extrinsic stochasticity (added *a-posteriori*<sup>2</sup>) included in [19] and few other studies. A search in PubMed<sup>3</sup> with keywords *covid-19*, *mathematical* and *model* offered 540 articles. A refinement with keywords *covid-19*, *model*, *asymptomatic* and *stochastic* ends in six scientific publications (as of 2020-09-09) plus one news article (not a research article) for a specialised magazine. Of the later six, reference [19] is the most closely related to our study, hence our decision to indicate only the differences of the current study with a related one among the pre-existing papers.

In this study, we will take a complex systems view. We begin by acknowledging that the COVID-19 epidemic is no longer a free-running epidemic but rather one in which there is a strong interaction between the public health system and the population dynamics of the outbreaks. Changes in the evolution of an outbreak trigger changes in the consideration of which characteristics of the COVID-19 cases should (or should not) trigger public action. This indicates that there is a clear interaction between these systems and they cannot be considered independent. To illustrate the point, we will use the various criteria of ‘COVID-19 case’ used in our home country (HGS), Argentina, following recommendations by the World Health Organization (WHO). We will produce compartments that relate to the evolution of the case in medical or biological terms as well as to the categories corresponding to the different protocols to be applied to the case.

The response to an epidemic requires not only the mobilisation of public resources but the participation of the public as well. To organise the actions required for each individual case COVID-hot-lines and web-servers have been organised worldwide. Such help services indicate which measures to take by those that suspect they are developing COVID-19, and prompt official actions if needed. Hospitals and health centres, as well as help services, are coordinated in their actions by protocols. A main tool of these protocols is the *suspected-case* criterion. This criterion regulates state intervention and depends on clinical symptoms of the (potential) patient and other circumstances. The criterion constitutes a difficult balance between the administration of resources (e.g. use of reverse transcription polymerase chain reaction (RT-PCR) kits and laboratories), the developmental stage of the epidemic, the mortality risk of the case and more. As in any decision taken under real circumstances (limited resources), establishing the suspected-case criterion implies trade-offs. When diagnostic resources – such as RT-PCR tests – are limited, a conflict emerges: should we reserve them for individual diagnosis (e.g. to confirm diagnostic by symptoms in cases of doubt or concern) or perhaps use them in epidemiological surveillance (triggering actions such as contact tracing or sample pooling monitoring) as well? In any intermediate cases: in which proportions?

Should the general criterion depend on being a *contact* of a COVID-19 case? Does it make sense to require weaker symptoms for the population which is aware of having epidemiological contact with COVID-19 cases rather than for the *communitarian* cases that cannot account for how they could have been infected? Actually, it could make sense if by such measures we were able to achieve a more efficient use of a scarce resource to be reserved for diagnosing related to treatment (a private/individual criterion contrasting to public/epidemic criteria). The question must be put: is it correct to focus our attention on travellers and their contacts at the beginning of the outbreak? Is efficiency really boosted

by requiring two relevant symptoms of a list for potentially communitarian cases and only one to people with epidemiological contacts? In the context of the propagation of SARS-CoV-2, what are the consequences of such decisions? We will address these questions implementing a model apt for answering them.

To set the grounds for our model, we analyse data collected by the Public Health Ministry of Argentina, made available to us through the COVID-19 initiative under the Ministry of Science and Technology. The model incorporates medical findings regarding the transmission of SARS-CoV-2 as well as actions taken by the health authorities and to a certain extent the social behaviour of the population. We apply the model to small slums (variously called in South America: *villas miseria*, *villas de emergencia*, *cantegril*, *favelas*, etc.) where the conditions of homogeneous contact, frequently used to simplify the modelling task, are closer to be fulfilled. We show how the model predicts epidemic circulation below the detection level for surprisingly long periods of time. Also, we illustrate that ‘average epidemics’ are not good representatives to grasp the dynamics, and that the undetected (mild, unrecognised, presymptomatic, ‘asymptomatic’) cases are in good proportion the result of public policies coupled to the characteristics of the illness. The outcome of three forms of surveillance and public action are comparatively analysed.

In Section ‘The Model’, we describe the model, from its basis – supported in both biology and social behaviour – all the way to the algorithm implementing a Markov-jump process [3, 4]. Results are presented in Section ‘Results’ and discussed in the following section. Section ‘Conclusions’ finally sums up the conclusions.

## The model

### Biological and social input

#### What is a COVID-19 case?

We review the evolution of the definition of ‘case’ along with the development of the pandemic. In many countries, this definition emerges from the national health authorities, following recommendations from the WHO. By 27 January 2020, there were comparatively few cases outside China. Apart from special considerations for sanitary operators, the definition of *suspected case* from the Italian health authorities<sup>4</sup> considered two situations: (A) severe acute respiratory infection (fever, cough and request for hospitalisation) and presence in risk zones a few days before the onset (at that moment mainly Wuhan/Hubei), or (B) acute respiratory infection and either recent presence at the Wuhan live animal market or recent close contact with a *confirmed* (positive PCR test) or *probable case* (a PCR-tested suspected case without a conclusive result). By 22 February,<sup>5</sup> severity and hospitalisation were no longer required for (A) and dyspnoea was recognised among possible symptoms. By 9 March,<sup>6</sup> the considered situations were three: acute respiratory infection (with at least one among fever, cough and difficulty in breathing) without other aetiology and either (A) recent presence in the areas of local transmission of the disease or (B) close contact with probable or confirmed cases. The third situation considered (C) cases presenting severe acute respiratory infection (fever and at least one symptom of respiratory disease) requiring hospitalisation and without another aetiology that fully explains the clinical presentation. This new item acknowledges

<sup>2</sup>Extrinsic stochasticity misses the root of the stochastic phenomena [8, 9, 20].

<sup>3</sup>PubMed’s website accessed on 2020-08-28.

<sup>4</sup>Resolution 27 January, accessed on 2020-08-26.

<sup>5</sup>Resolution 22 February accessed on 2020-08-26.

<sup>6</sup>Resolution 9 March accessed on 2020-08-26.

**Table 1.** Health workers

PCR	Health workers			$P_{2s}(\pm)$	$F+$	$P_{F+}(\pm)$	$H$	$P_H(\pm)$
	$\geq 1 s$	$P_{1s}(\pm)$	$\geq 2 s$					
Pos	29 465	0.9835	23 688	0.7907	13 415	0.4478	2814	0.0939
Neg	32 290	0.9796	23 347	0.7083	9015	0.2735	2299	0.0697
OR	1.0040		1.1163		1.6373		1.3467	
Tot	61 755	0.9815	47 035	0.7475	22 430	0.3565	5113	0.0813

Number of confirmed PCR-positive and -negative cases displaying at least one or two symptoms (1s, 2s) from the set given in the text.  $F+$ : cases with fever plus another symptom.  $H$ : cases requiring hospitalisation. Positivity odds in dataset are 0.9089.  $P_X(\pm)$  stands for the probability of having one symptom or more of SARS-CoV-2 being positive (negative) for each category. Odds ratios (ORs) are the ratio of the odds under the symptoms condition to the odds in the full set.

the existence of the illness regardless of any presence in risk zones or close contact with probable or confirmed cases.

Also the concept of *close contact* evolved during the period. By 31 January, 'risk contacts'<sup>7</sup> considered only recent (within 14 days) travel or cohabitation with a COVID-19 patient (apart from special considerations for sanitary operators). The concept evolved to that of close contact, becoming highly detailed in what regards social distance (2 m, 15 min) and hygiene already by 27 February 2020.<sup>8</sup>

At the end of May, specific instructions for *contact tracing*<sup>9</sup> (already operative, although) had been developed.

The criteria for identification of cases shifts focus along the pandemic. At the beginning, the focus is on the 'virus import' from other regions where it is active, while the local diffusion becomes relevant only some weeks/months later. The trade-off in the identification generates 'classes' of contagion depending on the criterion.

Along with the case criteria, surveillance and control criteria are developed. At the beginning of the pandemic, *passive surveillance* (i.e. to wait for the spontaneous appearance of patients, except perhaps for travellers) was the most common attitude. Soon after, many countries developed different degrees of contact tracing (with varying success), even revealing preexistent flaws in the various national health and care systems.

In [Appendix A](#), we show the evolution of the criteria in Argentina and its relation to Italy's case.

The decision of what to consider a *suspected case*, and when further actions are to be taken, is a critical one. However, it is not clear which is the overall criteria, meta-criteria, adopted by Italy or Argentina, presumably upon recommendations of the WHO. It appears that the meta-criterion is to keep an even level of certainty of being a COVID-19 case for each individual case. It is then pertinent to explore whether this goal is achieved or not and if such goal is epidemiologically sound.

We discuss this issue with data from Argentina.<sup>10</sup> In [Table 1](#), we report PCR results for health workers after 6 June, from the dataset of 5th October 2020 with 62 920 cases with symptoms information (29 958 positive and 32 962 negative).<sup>11</sup> Health workers can be assumed to be more accurately monitored than other

patient groups. On 6 June, the criterion for suspicious case for health workers was changed to presenting one symptom belonging to the set: fever, cough, anosmia, dysgeusia, dyspnoea, odynophagia (see [Appendix A](#)). In 4 August 2020, the set of symptoms was extended to headache, diarrhoea and vomits.

For health workers, 98% of the cases that reported symptoms<sup>12</sup> presented at least one symptom in the extended set. Among them, 48% were diagnosed as COVID-19 cases using RT-PCR. Considering cases reporting at least two symptoms, the number of cases falls by 23% but the positive cases within the group move up only to 50%. If the criterion is 'fever and one symptom', the case fall is 64% while the positivity within this smaller set raises to 60%. Similar trends are found for the whole patient dataset.

The data indicate that requiring more symptoms results in missing positive cases. The improvement in positivity rates is outnumbered by the large or very large fall in detected cases, with no significant improvement in the use of resources. At the early stages of the epidemic only hospitalised patients with pneumonia were considered as possible COVID-19 cases. In such cases, the detection ability drops to less than 10% of the cases showing symptoms.

#### *Viraemia, symptoms and contagiousness*

An important ingredient of any model concerning the evolution of the disease requires the description of a contagion mechanism at the individual level. It is important to relate when, how much and how long a person is in a contagious condition to the evolution of the disease in the agent.

Upon contagion, the infected individual gradually develops larger and larger levels of virus, in pace with the viral reproduction capabilities in the infected patient. Eventually, a maximum level is reached and the viraemic load subsequently declines along with a recovery from infection. This process may be interrupted at any time because of complications, be them virus-based or any other.

We assume, therefore, that the viraemic load is the biological origin of both the severity of illness for an average infected individual and the capability to transmit the virus. In simpler words, the quantity of virus in each individual regulates how ill she/he is and with which efficiency the infection can be passed along.

Symptoms, severity and contagiousness are different from person to person, but they follow an approximate sequence from zero up to a maximum value, subsequently decaying towards zero again. From the day of clear symptom onset, we adopt a model for the viraemic load, based on early findings [18, 21] from the

<sup>7</sup>Resolution 31 January accessed on 2020-08-26.

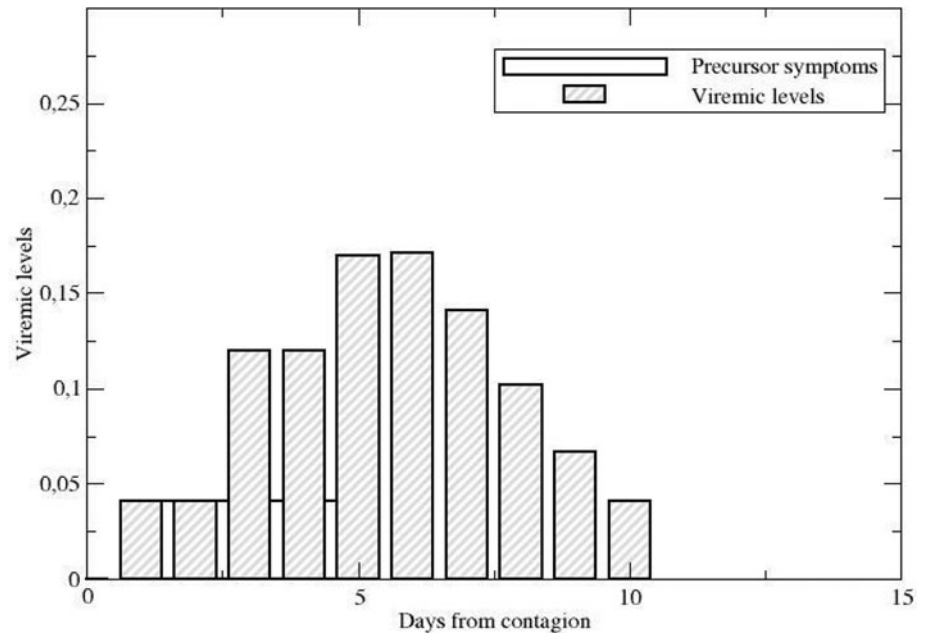
<sup>8</sup>Resolution 27 February accessed on 2020-08-26.

<sup>9</sup>Resolution 29 May accessed on 2020-08-26.

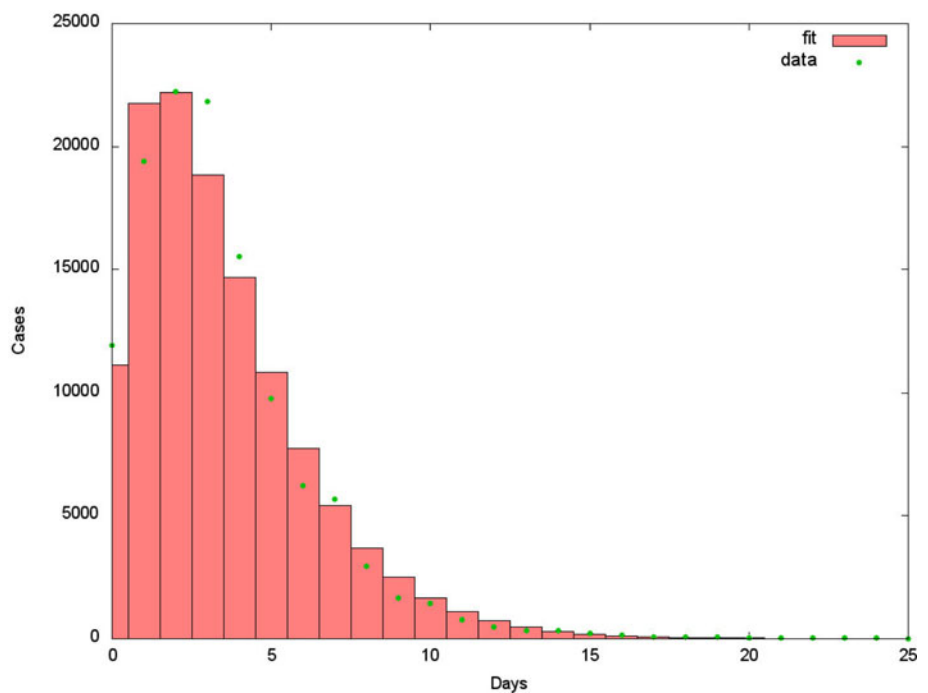
<sup>10</sup>The Argentine Ministry of Health provides on a daily basis an anonymised copy of the dataset corresponding to the nation-wide reported cases in epidemic outbreak for the National Science Council (CONICET).

<sup>11</sup>There are 493 cases with reported symptoms where none of the symptoms match the HS expectation.

<sup>12</sup>Report of symptoms is not an obligation for the sanitary units.



**Fig. 1.** Viraemic levels displaying a 2-day stage with low virus levels, followed by a stage with higher viraemic levels but still no detection symptoms. Bars represent days, not stages.



**Fig. 2.** Distribution of time between initial symptoms and swabbing calculated using 121 223 entries corresponding to the cases during the month of August 2020 (dots) and curve fit  $y(x) = 421\,627$  (histogram). The observed average time in Argentina is of  $\sim 3.86 \pm 0.5$  days. The fitted curve is the composition of two exponentially distributed stages, 2.10 and 1.86 days long in average. It is important to understand that the data reflects not only a biological matter, but it is also affected by public health decisions, the information of the population and self-diagnosing of the patient concerning the initial symptoms. As such, the statistical error is not the most relevant error. At the beginning of the outbreak the average time was longer than 5 days.

initial period of the pandemic where individual cases could be traced in detail. We model the viraemia from days 5 to 10 using a gamma distribution. The *presymptomatic period* (a period usually of weak symptoms) is modelled in three stages, a first non-contagious compartment lasting a day in average, followed by a low-contagion compartment, lasting about 2 days, with the same viraemic level than the last day of contagion and finally followed by a compartment with higher contagiousness lasting also about 2 days.

The duration and distribution of the presymptomatic days, from contagion to symptoms, described in this form is supported by the distribution of times between the appearance of earlier symptoms and the day of diagnostic for the data collected in

Argentina (see Fig. 2). In fact, the observed mean for the data points is  $3.86 \pm 0.5$  days, plus 1 day without any symptoms, yielding slightly less than 5 days before the onset of recognised symptoms and the decision of swabbing.

After the presymptomatic period, symptoms usually appear clearly until they gradually decline. We assume the symptomatic compartments to last in average 1 day each, with viraemic levels as in the final part of Figure 1.

For the sake of dealing with a pandemic, symptoms in themselves are only an ingredient. They facilitate the possibility of detecting infected patients, especially when the pandemic constrains the sanitary authorities to keep a passive attitude. In any case, the appearance of symptoms on each individual depends

**Table 2.** Normalised viraemic levels describing an average evolution of the disease

Stage	Duration (days)	Normalised viraemic level
0	1	0
1	2.10	0.0406
2	1.86	0.1201
3	1	0.1695
4	1	0.1713
5	1	0.1413
6	1	0.1019
7	1	0.0667
8	1	0.0406

The probability of contagion is assumed to be proportional to the viraemic levels along the different stages. The levels enter the modelling of the probability of detection as well (along with other important influences such as sanitary policies). The levels are normalised so that  $\sum_i V_i t_i = 1$ , where  $t_i$  is the duration and  $V_i$  is the viraemic levels.

not only on the viraemic load, but also on the individual condition of each patient.

On the contrary, regardless of symptoms (if and when they appear), the two processes driving the evolution of the pandemic are *contagiousness* and *detection*. The first one is of course mandatory since there is no pandemic without infections. Both these processes have a social component and a biological component. The biological component was discussed above: we assume both the probability of detection and the probability of contagion to be proportional to the viraemic levels of infected individuals, modelled according to Figures 1 and 2. This assumption rests partly on the observations in [18, 21]. However, a recent study ([22], appeared after this submission) suggests that there are other mechanisms operating as well, depending on the patient's response to the infection and deserving a deeper analysis. The modelling profile is summarised in Table 2. The social component reflects the ability of the sanitary authorities to enforce measures in order to (a) detect infected individuals and reduce the chances of contagion (by isolation, hospitalisation, etc.) and (b) effectively influence social behaviour, aiming to reduce the chances that infected, undetected individuals may transmit the disease.

#### About slums

Slum areas have a very specific social structure. They have high-population density and intense poverty, resulting in homes shared by several generations (sometimes with just one bedroom), larger number of homeless people in comparison with the rest of city, precarious services (water, electricity, sewage, etc.), as well as strong internal ties and social organisations such as community run food assistance and other internal solidarity networks. 'Stay at home' policies cannot be sustained, to the point that sometimes the whole slum area has been locked allowing for its internal life to continue undisturbed. The contact structure for individuals in slums is more homogeneous and with larger contact rate than the surrounding cities.

#### The detection of cases as a function of the surveillance protocol

The decision of admitting a case as a probable case of COVID-19 depends not only on the biological/health condition of the case

(i.e. the viraemic level, presence of symptoms, etc.), but also on the expectations of the health services (HS) as we have discussed in Section 'What is a COVID-19 case?' and Appendix A. Since the chances for a contagious person to produce new cases depend on *a-priori* expectations, the expectations change the removal rate of contagious people (e.g. by isolating the person). Furthermore, the condition of being suspected *a-priori* is mostly hereditary. The suspicion increases the probability of detection and the detection of a case makes those infected by the case more likely to be detected. Let us call *T*, *traceable*, those with larger probabilities of detection *a-priori*, and *U*, *untraceable*, those with smaller probabilities of detection. Let us further consider the limit situation where all *T* are traced and detected with certainty and no *U* is ever detected. Such an idealised, limit situation will result in two independent epidemics, for no *T* can ever produce a *U* case and reciprocally, no *U* can produce a *T* case. No real situation is expected to reach this limit case, hence, in a more accurate description *U* cases are detected with lower probability and later than *T* cases. Also, some *T* may escape tracing and detection when still contagious. The inheritance of the tracing classes is then imperfect and there is only one, mixed-type, epidemic. We represent this situation by a probability table (written in the matrix form):

$$\begin{pmatrix} P(T \text{ by } T) & P(T \text{ by } U) \\ P(U \text{ by } T) & P(U \text{ by } U) \end{pmatrix} = \begin{pmatrix} 1 - \epsilon & s \\ \epsilon & 1 - s \end{pmatrix}$$

The probabilities,  $P(X \text{ by } Y)$ , indicate the probability for a susceptible person infected by a contagious case of type *Y* of becoming a case of type *X* assuming it was effectively infected. The non-negative quantities  $\epsilon$ ,  $s$  are not new parameters since we have to satisfy that  $P(T \text{ by } T)$  is exactly equal to the probability of a *T* case being effectively detected. The same can be said of  $P(U \text{ by } U)$  with respect to the undetectable cases *U*.

Since all health systems have limited resources and suffer different epidemic impacts, different strategies are likely to appear. One of the goals of this paper is to explore the impact of different strategies on the (local) evolution of the pandemic. Schematically, we will consider three scenarios, labelled *passive*, *intermediate* and *active* representing different policies for the detection process.

In the *passive* policy, intervention starts when and if the symptoms are clear. The intensity of the perceived symptoms is assumed to be, on average, proportional to the viraemic state. A distinction is made between the *T* and *U*, being the HS's more prone to act for the *T* group than for the *U* group. The passive policy represents the policies adopted during the early days of the pandemic (mid-February to mid-March 2020 in Europe), where the HS focused attention on imported cases (travellers) and their contacts. The *intermediate* policy reflects the situation in which the HS become aware of the problem of presymptomatic contagious cases, and begin to track oligosymptomatic cases in the *T* group (contacts of known cases). At the same time, it has been observed a lowering on the requirements, in terms of a lower number and a larger set of symptoms, required for sanitary intervention (isolation). The *active* intervention consists of one of the two possibilities: either the *T* class is substantially enlarged by including in it the contacts of contacts, as was done in Italy or by dropping the distinction between *U* and *T* and acting (or strongly exhorting to individual action) on cases presenting any symptom compatible with COVID-19, no matter how weak, as

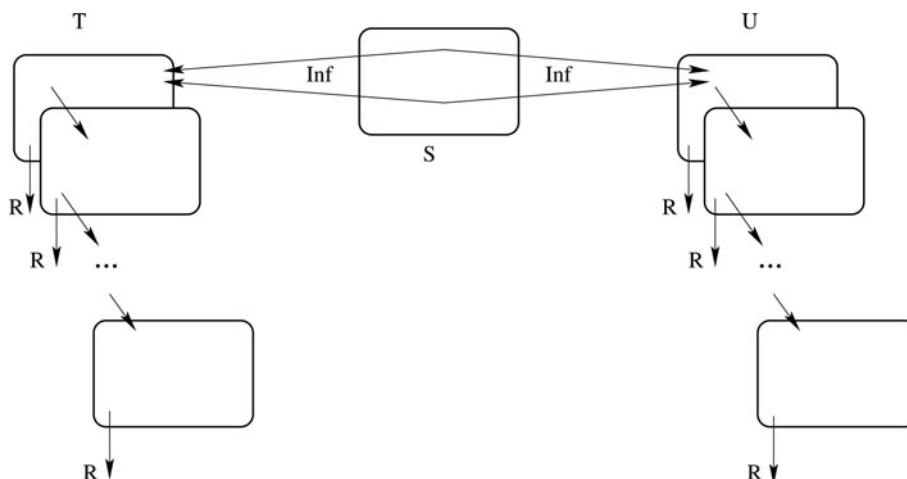


Fig. 3. Schematic view of compartments and events.

it was the public advise of e.g. the Swedish HS.<sup>13</sup> We will only model the second case.

### Mathematical/computational support

The general approach is based on a Markov-jump process following the setup of the Feller–Kendall [3, 4] algorithm. The compartments  $X_i$ ,  $i = 1, \dots, N$  involved in the process are the different classes of individuals taken into account (described below) and the stochastic dynamics evolves by expressing the number of individuals on each compartment as a function of time. Transitions between compartments are given by Markov jumps triggered by different events and characterised by an event probability rate  $W_\alpha(X)$ ,  $\alpha = 1, \dots, E$ . The relation between events, compartments (populations) and stochastic dynamics is given by

$$X_i(t) = X_i^0 + \sum_{\alpha=1}^E n_\alpha(t) \delta_i^\alpha \quad (1)$$

where  $X_i^0$  is the initial condition for compartment  $i$ ,  $n_\alpha(t)$  indicates the number of occurrences of event  $\alpha$  up to time  $t$  and  $\delta_i^\alpha$  is an integer indicating how each occurrence of event  $\alpha$  modifies the population in compartment  $i$ . For the present problem,  $\delta$  will take the values  $-1, 0, 1$ , meaning that e.g. one infected individual is removed from the contagious process by isolation, etc. The stochastic dynamics proceeds by establishing the behaviour of  $n_\alpha(t)$ .

General properties of Markov-jump processes are assumed to hold for this problem, in particular that events are independent of each other (although related indirectly by the dependence of the rates on the populations). These properties add up to the following two results [3, 23–25]:

1. The waiting time to the next event is exponentially distributed with rate  $R = \sum_{\alpha=1}^E W_\alpha(X)$ .
2. At the occurrence time indicated above, the probability of occurrence for event  $\alpha$  is  $P_\alpha = W_\alpha(X)/R$ .

A realisation of this stochastic dynamical process requires a good knowledge of the probability rates  $W_\alpha$  and the computation of one random number (exponentially distributed) for the time of occurrence of the next event and another (uniformly distributed)

<sup>13</sup>Recommendation of 13 March 2020 (in Swedish) General information (English) accessed on 2020-10-13.

for selecting the event happening at that time. Upon occurrence of each event, populations and consequently transition rates are updated according to Eq. (1). Random numbers were generated with the *Double precision SIMD-oriented Fast Mersenne Twister (dSFMT)* algorithm [26], implemented in C.

### Details

The algorithm is implemented as a C-programme, fully available from GitHub. The compartmental structure is as follows (see Fig. 3).

There exist three classes of compartments, namely *susceptible*  $S$ , *traceable infected*  $T$  and *untraceable infected*  $U$ . Infected individuals belong to several sub-compartments describing the degree of evolution of their disease (or rather their infective period). At each stage, they may proceed in the disease to the next stage of infection (diagonal arrows) or be removed from the system by any reasonable means, e.g. by being detected and isolated by the HS, by self-isolation, hospitalisation, etc. (vertical arrows labelled  $R$ ), thus ceasing in all such situations to be a source of contagion. Infection may proceed either by contact of  $T$  or  $U$  individuals with an  $S$  individual, or by ‘importing’ the infection from outside the system in consideration.

What regards infection by contact, the tracing of infections is usually not complete, for various reasons. To take this fact into account, we assume that a portion of infections by  $T$  individuals (of size  $\delta < 1$  in Table 3) may remain undetected and also that a portion of infections by  $U$  individuals (of size  $s < 1$  in Table 3) will eventually become detected. The quantities  $s$  and  $\delta$  are discussed in Section ‘The detection of cases as a function of the surveillance protocol’ and will be further specified below. Two additional uniformly distributed random numbers  $r_1$  and  $r_2$  (in  $[0, 1]$ ) are computed to decide these outcomes (double arrows from  $S$  in Fig. 3), representing the probability pairs  $\{1 - \delta, \delta\}$  and  $\{s, 1 - s\}$ , respectively, for  $T$  and  $U$  infected individuals.

For the ‘imported’ infections taking place outside the system, the uniform random number  $r_0$  distributes the resulting infected individuals among  $T$  and  $U$  with proportions  $\{1 - \eta, \eta\}$ .

### Rates and actions

In Table 3, we describe the expressions adopted for the different rates and their action on the population (i.e. the non-zero values of the incidence matrix  $\{\delta_i^\alpha\}$ ). Considering the nature of the available data, the time-unit is  $(\text{day})^{-1}$ , i.e. transition rates are given per day.

**Table 3.** Transition rates and incidence matrix

Event $\alpha$	$W_\alpha(X)$	Incidence matrix $\delta_j^\alpha$
External contagion	$a \frac{S}{N}$	$S: -1; \text{if } (r_0 > \eta) T_0: +1$ else $U_0: +1$
Contagion $T \rightarrow S$	$\beta_T V(k) T_k \frac{S}{N-1}$	$S: -1; \text{if } (r_1 > \delta) T_0: +1$ else $U_0: +1$
Contagion $U \rightarrow S$	$\beta_U V(k) U_k \frac{S}{N-1}$	$S: -1; \text{if } (r_2 > s) U_0: +1$ else $T_0: +1$
Disease evolution $T$	$pr(T_k)$	$T_k: -1; T_{k+1}: +1$
Disease evolution $U$	$pr(U_k)$	$U_k: -1; U_{k+1}: +1$
Infective removal $T$	$rem(T_k) = B_T^k V(k) T_k$	$T_k: -1$
Infective removal $U$	$rem(U_k) = B_U^k V(k) U_k$	$U_k: -1$

$V(k)$  indicate the viraemic levels at stage  $k$ .  $r_1, r_2$  are the uniformly distributed random numbers in  $[0, 1]$ . Other quantities are described in the text.

In the table,  $N$  is the size of the population, typically a neighbourhood or other region that can be safely assumed to behave homogeneously (basically, that any individual may, in principle, meet any other individual; a natural assumption for working places, schools, etc.). Initial conditions for all simulations is that most individuals ( $N - c$ ) are in compartment  $S$  while the remaining  $c$  are in  $T_0$  or  $U_0$  (we assume  $c$  is typically around 2 for  $N$  up to a few thousands). The ‘import’ rate  $a(S/N)$  describes infected individuals undergoing contagion outside the system. We include in this event the possibility of travellers bearing the infection when returning to the system after a temporary absence, a group that has been important in the global scale to transfer the disease across continents, but is comparatively small for such stable communities as those we consider.

The evolution of the illness is given by stages  $T_k, k = 0, \dots, K - 1$  (and similarly for  $U$ ) describing the viraemic level  $V(k)$  at each stage (Fig. 1 and Table 2). In this study,  $K = 9$ .  $pr$  describes the rate of passage to the next viraemic stage i.e.  $(pr)^{-1}$  is the average permanence of an individual on each stage (second column in Table 2). The factor  $\beta V(k)$  describes the contagion rate between a susceptible and an infected individual. In principle,  $\beta_T$  and  $\beta_U$  may be different but we have not explored that possibility. Since the viraemic levels are normalised, the weighted infective period is one and the factor  $\beta$  corresponds to the basic reproductive number  $R_0$  of the SIR (Susceptible, Infected, Recovered) model. The approximate value of 2.5 was taken from early reports from China [27] and contrasted against data from the initial days of the first outbreak in slums in Argentina (Barrio Padre Mujica, also known as Villa 31 in Retiro, a neighbourhood of Buenos Aires city).<sup>14</sup>

The constant  $\delta$  indicates what portion of the individuals infected by a  $T$  will not be detected by the HS during the contagion phase of that case, while similarly  $s$  indicates what portion of the individuals infected by a  $U$  will eventually be detected by the control procedures. Similarly, the constant  $\eta$  describes the distribution of imported cases among  $T$  and  $U$  compartments. Since  $\eta$  is largely unknown we will only consider the extreme cases. We have  $\eta = 0$  usually associated with long distance travels to/from regions of viral circulation. For the case of slums, casual contagion

within the same city but in e.g. different neighbourhood is expected to be the most frequent case, hence we adopt  $\eta = s$ .

Finally,  $rem$  is the rate of removal of an individual from stage  $k$  out of the contagion chain. This rate also depends both on the viraemic level and on the HS strategies (giving different choices for the factors  $d_T(k), d_U(k)$ ). This part of the model will be described in detail in the next subsection.

The data usually discussed in the news and websites are the number of confirmed COVID-19 cases. In the present model, these data are represented by the total number  $D$  of detected individuals, i.e. the outcome of all removal events within the infective period. The model provides an estimate of the silent cases, i.e. infected individuals of which the HS has no records. In the model, these non-detected infected individuals  $ND$  are given by the identity  $N = S + D + ND$ , where  $N$  is the population size and  $S$  is the number of susceptible individuals.

The outcome of the model is presented by computing a few realisations (typically 100) of the Feller–Kendall algorithm. No matter how parameters are chosen, there exists a non-zero probability of early disease extinction (as it is in any Markov-jump process), particularly when the onset of the epidemics contains very few infected individuals (1 or 2 in a population of a few thousands). The model allows for ruling out early extinctions, considering that the epidemics that are tracked, and concern us, are those that avoid early extinction and come to be notable.

The actual evolution of the pandemic is intrinsically stochastic. Borrowing from the modelling language, there is only one ‘realisation’ of the real process, namely the one we are currently experiencing. There is no ‘second run’, although many weakly coupled contagion chains may be running simultaneously within e.g. a larger city.

With this in mind, we stress that the averaging of realisations is not a substitute for the real process. It has a limited value, in that it highlights features that are recurrent, while it smears out what is less frequent. Moreover, no realisation of the stochastic process is ‘more true’ than any other. Predictions based only on the averaging of realisations may serve as a clue about what to do, but policy decisions should take into account the whole picture.

### Contagion, removal and HS policies

We consider in detail the mechanisms of contagion and removal, as well as their relation to both the evolution of the disease in the infective individual and the HS policies.

Contagion within the system is taken to be strictly proportional to the viraemic levels  $V(k)$ . The proportionality constants  $\beta_T, \beta_U$  may vary according to social strategies and attitudes.

The eventual removal of an infected individual in the model is governed by the competition between two mutually exclusive events. Either the individuals evolve to the next stage in the viraemic levels (i.e. they are still infected and capable of contagion) or they are removed from the contagion chain for whatever reason (detection, isolation, full recovery or death). At stage  $k$ , the probabilities  $P_m^X(k)$  of moving to the next stage in the contagion chain and  $P_r^X(k)$  of being removed from the chain for an individual of class  $X = \{T, U\}$ , can be described (in the notation of Table 3) as:

$$\begin{aligned}
 P_m^X(k) &= \frac{pr(k)}{pr(k) + B_X^k V(k)} \\
 P_r^X(k) &= \frac{B_X^k V(k)}{pr(k) + B_X^k V(k)} \\
 &= 1 - P_m^X(k), \quad k = 0, \dots, K - 1
 \end{aligned}$$

<sup>14</sup>Data available at [buenosaires.gob.ar/datosabiertos](https://buenosaires.gob.ar/datosabiertos).

**Table 4.** Parameter values used in various simulation scenarios, in the variants low, medium and high respectively

Name	Ideal <i>I</i>	Fever + <i>F</i> +	Hospital <i>H</i>
$\alpha$	0.002	0.002	0.002
$\beta_T = \beta_U$	2.5	2.5	2.5
$B_T$	0.24,1.00,1.76	0.24,1.00,1.76	0.24,1.00,1.76
$k_{0T}$	1,1,1	1,1,1	1,1,1
$B_U$	0.24,1.00,1.76	0.10,0.34,0.46	0.03,0.09,0.12
$k_{0U}$	1,1,1	1,1,1	3,3,3

*B* is called *det* in the code and  $k_0$  is called *delay*. Under the present normalisation, the contagion rate  $\beta$  corresponds to the basic reproduction number  $R_0$  of the SIR model.

where  $B_X^k$ ,  $X = \{T, U\}$ , model the HS policy adopted. In the present implementations, the factors  $B_X^k$  are set to zero for an initial subset  $[0, \dots, k_0 - 1]$  of stages ( $k_0 \geq 1$ ) and take the same positive value  $B_X$  for the remaining stages,  $k \in [k_0, K - 1]$ .  $B_X$  relates to the probability of  $X$  being effectively detected (named  $1 - \delta$  and  $s$  in Section ‘The detection of cases as a function of the surveillance protocol’) through Eq. (2).

At the final stage,  $K - 1$ , the overall action of both competing events is a removal from the contagion chain. Individuals that have not been removed at any previous stage have effectively participated in the contagion chain during all of their contagious period. These individuals were not detected by the HS policies while they still were active in the contagion chain. The overall probability of detection can be computed as follows. Let  $Q_k^X$  be the probability of removal up to and including stage  $k$  for infected individuals of class  $X = \{D, U\}$ . Set further,  $Q_{-1}^X \equiv 0$ . For any stage  $i$ ,

$$Q_i^X = Q_{i-1}^X + (1 - Q_{i-1}^X)P_r^X(i), \quad i = 0, \dots, K - 1,$$

which can be restated as  $Q_i^X = (1 - P_r^X(i))Q_{i-1}^X + P_r^X(i)$ . The total probability of removal during the infective period is  $Q_{K-1}^X$ , while the probability of being *detected* at some point during the infective period for individuals in class  $T, U$  is given by

$$Q_{K-1}^T = 1 - \epsilon, \quad Q_{K-1}^U = s. \quad (2)$$

Note that  $Q_{K-1}^T$  and  $Q_{K-1}^U$  are rational functions, the ratio of two polynomials of degree  $K - k_0$  in  $B_X$ .

Equation (2) relates the value of the constants  $B_T, B_U$  for the different HS policies with the probability of detection. The differences in dealing with  $T$  and  $U$  infected individuals follows from the differences between  $Q_{K-1}^T$  and  $Q_{K-1}^U$ , being the HS's more prone to act for the  $T$  group than for the  $U$  group. We distinguish three main policies:

*Passive policy:* Intervention on the  $U$  class concerns only severe cases (e.g. requiring hospitalisation) in a situation where the viraemic levels of the patient are comparatively high. In the model, intervention for the  $U$  class starts at  $k_0 = 3$  (stage 3 in 2).

*Intermediate policy:* The conditions required for sanitary intervention (isolation) in the  $U$  group are broadened in terms of a lower number and a larger set of symptoms and possibly intervention at an earlier stage. Active (preventive) intervention, as in contact tracing, starting at stage  $k_0 = 0$  (or 1), is implemented for the  $T$  group. It reflects a situation in which the HS become aware of the problem of presymptomatic contagious cases, and

begin to track oligosymptomatic cases in the  $T$  group (contacts of known cases).

*Active policy:* No distinction is made between  $U$  and  $T$  regarding actions of the HS. One symptom is enough to trigger sanitary actions. Interventions start at stage  $k_0 = 0$  (or 1).

*Simulation scenarios:* In the next section, we discuss a few scenarios based on these policies, relating to data from Table 1. We consider three detection efforts, that we may call *Low*, *Medium* and *High* ( $L, M, H$ ) for each scenario. We identify the  $T$  group with health workers, who for practical reasons were better monitored than other individuals. A first scenario, labelled *I* (for ‘Ideal’) corresponds to the active policy, with three different effort levels, represented by  $B_X$  such that the detection probability  $1 - \delta = s$  takes the values 0.21, 0.61 and 0.79. The latter corresponds to the fraction of confirmed cases among health workers displaying one symptom of the extended list of Section ‘What is a COVID-19 case?’. However, registration of symptoms was optional. Therefore, 0.79 is only a crude lower bound to the ability of detecting cases among health workers (which are subject in part to routine testing). A second scenario of intermediate character, labelled *F+* (for ‘fever plus other’), corresponds to the same detection probabilities as above for the  $T$  group, whereas for the  $U$  group the detection probabilities  $s$  are set to 0.10, 0.28 and 0.36. The latter corresponds roughly to the proportion of confirmed cases among health workers displaying fever plus another symptom in Table 1. The third scenario, labelled *H* (for ‘hospital’) is still unaltered for the  $T$  group relative to the previous two, while the  $U$  group is subject to the passive policy (thus assuming that only highly viraemic cases have a chance of being detected, and only from stage 3), with detection probabilities  $s$  set to 0.02, 0.06 and 0.08. The latter corresponds roughly to the proportion of confirmed health worker cases that were hospitalised as in Table 1. Hence, the *High* intensity level of the three scenarios relate to detection policies adopted by HS's at different periods of time. The lower estimate for the HS actions is roughly 1/4 of the upper estimate and the intermediate estimate was taken to obtain an intermediate level of detection. In the unfortunate situations where the epidemic is out of control, the effectiveness of the HS measures could be even lower.

Unless otherwise stated, all simulations are performed with 5000 individuals, of which two are initially contagious in the  $T$  compartment (it makes only imperceptible difference to set the initial contagion in the  $T$  group or the  $U$  group), while the contagion rate is set to  $\beta = 2.5$  and there is a small rate of external contagion ( $ext = 0.002$ ).

A list with the parameter values used in different scenarios is provided in Table 4. Other necessary input data for running the simulations are: number of realisations (usually 100), length of simulation in days, initial condition for populations  $S, T, U$



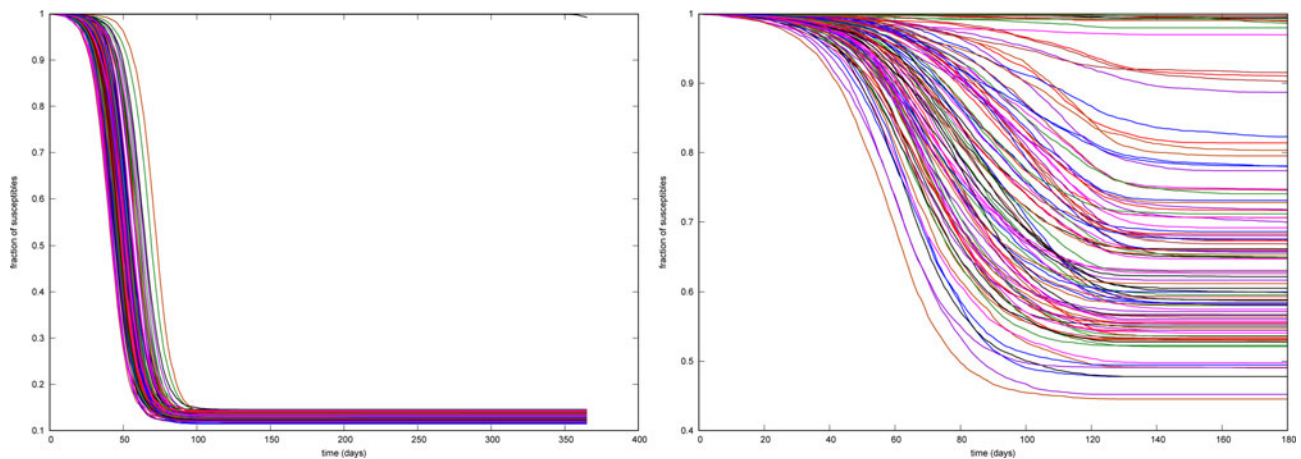


Fig. 4. Fraction of susceptible individuals for 100 realisations. Left: Low, constant detection. Right: Increasingly sharp detection.

(usually 4998, 2, 0), random number seed, flag to discard ‘early’ extinctions (positive integer) and maximal duration to be considered ‘early’ (usually 19 days).

## Results

### General results

The following results follow from the structure of the model. There is essentially nothing left to prove, just following the construction in Section ‘Contagion, removal and HS policies’.

*Lemma 1.*  $Q_{K-1}^X$  is a monotonically increasing function of  $B_X$ .

In modelling language,  $B_X$  senses the efficiency of the detection process.

*Lemma 2.* For fixed  $B_X$ ,  $Q_{K-1}^X$  is decreasing with increasing  $k_0$ .

A late start stage for the detection process can be interpreted as a HS policy that is only capable of taking care of seriously ill cases, with highly developed viraemic levels. The more stages an individual passes without any policy action, the lower the overall chances of detection within the infective period.

### Simulations

Simulation results allow us to compare the outcomes of different policies on an equal footing.

#### Spread

Before considering averaged results let us sense the spread of outcomes from different realisations of the process. In Figure 4, we show the fraction of susceptible individuals as a function of time for 100 realisations of the stochastic process in two different configurations.

The left panel corresponds to a situation where the probability of detection while still contagious is 21% for the  $T$ -group and 9% for the  $U$ -group, with  $\beta = 2.5$ . All outcomes display a sharp fall in the number of susceptible individuals. Note, however, the spread in time: the fastest and slowest realisations differ in about 40 days, corresponding to 100% at the 0.5 level. The right panel corresponds to a weaker contagion situation ( $\beta = 1.75$ ), where the probability of  $T$ -detection increases every 60 days, from 0.6 through 0.71 up to 0.79 (all detections starting on stage 1), while the  $U$ -detection goes from 0.07 through 0.58 (with detection starting on stage 3) up to 0.79, with detection starting on stage 1. Note

here the spread in the outcome. Although some realisations display almost no variation in the fraction of susceptible individuals, some others achieve a fall of over 50%. It is worth to keep in mind that the policy of progressively increasing the detection effort has been the rule in practical cases.

#### Initial growth

As in most models with homogeneous contact, the initial growth of the epidemic outbreak is almost exponential and this regime lasts for about 2 months in the present simulations with about 5000 initial susceptible individuals. However, it is worth to indicate that the growth exponent of infected cases and that of detected cases is not the same, being the latter smaller than the former, especially in less effective regimes as  $H$  and  $F+$ . As a consequence, basic reproductive numbers inferred from the early development of the pandemic that had assumed that detected cases are roughly proportional to the actual cases underestimate the growth rate (see Fig. 5). Note that the gap between growth rates is larger for the lower detection effort as compared with the higher (red/green vs. blue/magenta pairs).

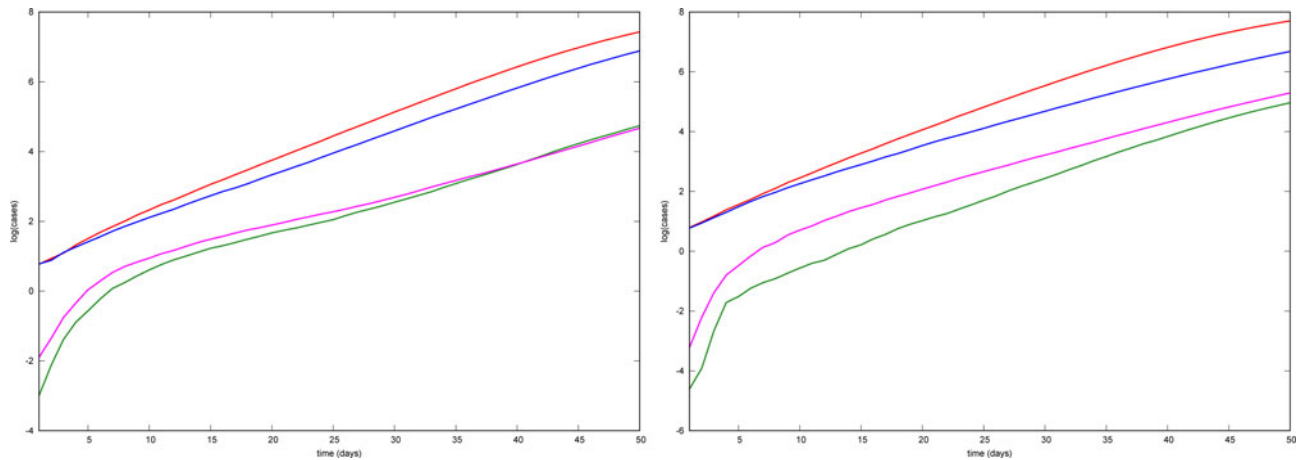
#### Undetected/detected ratios

The ratio between total cases and detected cases of COVID-19 has been the subject of several studies. In particular, Malani *et al.* [28] and Muñoz *et al.* [29] address the situation in slums, reporting ratios of 10:1 [28] and 5:1 [29].<sup>15</sup> The latter study was performed at least 1 month after ‘most cases’ occurred, although with the outbreak still running. The majority of the registered cases had occurred before 6 June, when the tracking method in use was of type  $F+$ . After 6 June, the tracking sharpened to ‘any two symptoms’, a medium form of  $I$ . We show averaged ratios in Figure 6, but it is worth to keep in mind that there are usually large fluctuations present. The figure shows that in all situations there is a tendency to a sharp increment of the ratio at the beginning of the outbreak followed by a maximum level and subsequently a monotonic decrement.

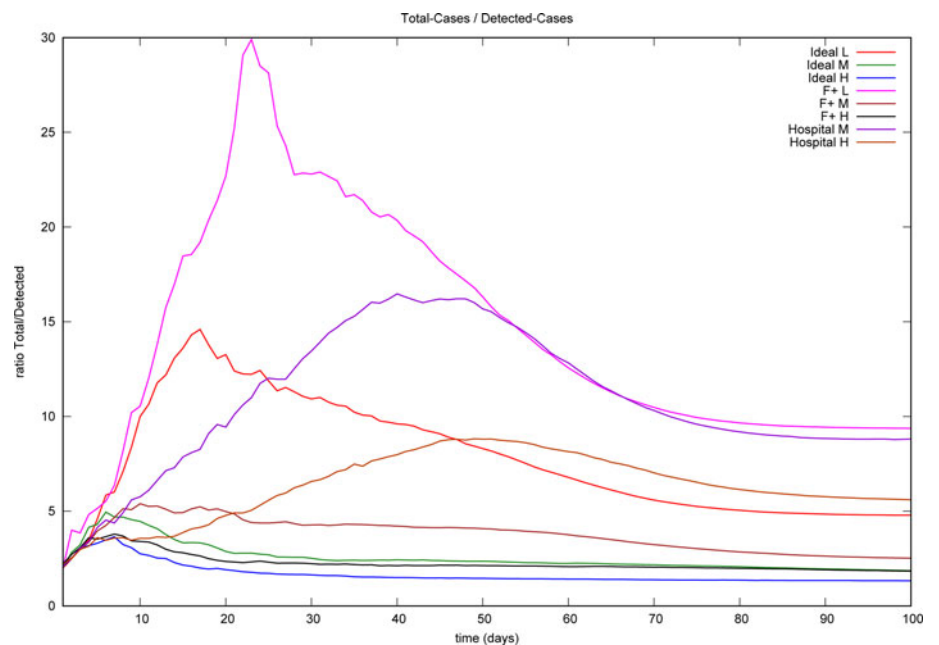
#### Dynamical mechanisms

The most remarkable features present in the simulations are the diverse forms in which the stochasticity and the particularities

<sup>15</sup>The accuracy of serological tests used to evaluate *a-posteriori* the prevalence of COVID-19 can be questioned based on studies that indicate that not all people presenting immunological reactions to the virus develop humoral antibodies [22, 30, 31].



**Fig. 5.** Average growth rates for 100 realisations at the beginning of the epidemics. Left:  $H$  simulations. Red ( $\log(N - S)$ ) and green ( $\log D$ ) curves correspond to medium effort while blue ( $\log(N - S)$ ) and magenta ( $\log D$ ) curves correspond to high effort. Right:  $F+$ . Red ( $\log(N - S)$ ) and green ( $\log D$ ) curves correspond to low detection effort while blue ( $\log(N - S)$ ) and magenta ( $\log D$ ) curves correspond to medium detection level.



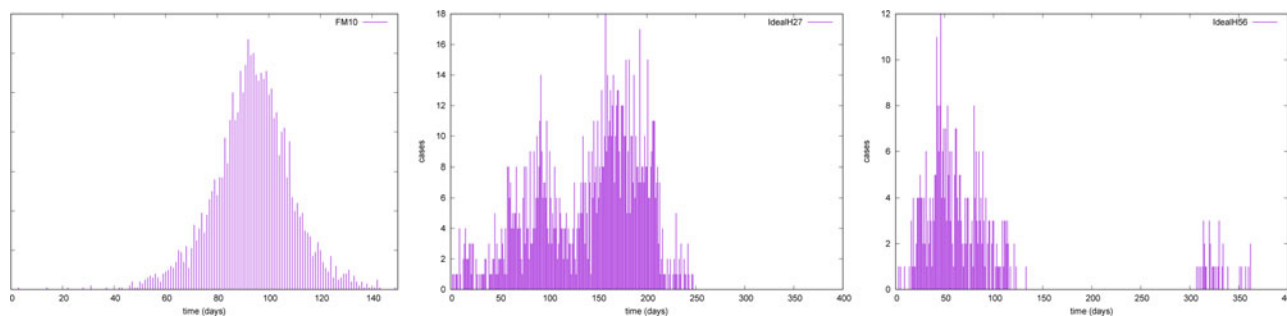
**Fig. 6.** Average ratio of total cases to detected cases (100 realisations) under eight different control measures implemented with moderate efficiency.

of the contagious process manifest globally. Despite being seeded with two traceable cases, it is not uncommon (probability larger than 0.01) to observe the outbreak to remain with sporadic cases up to 50 days and only then the recognisable bell-shaped of the daily cases begins. We show one of such cases in Figure 7, left panel. In the centre panel, we show a ‘two waves’ outbreak under policy  $IH$  and a different shape of ‘two waves’ with a long delay between them is shown in the right panel. The difference among realisations suggests that stochastic epidemic outbreaks are not just an ‘average outbreak’ plus noise.

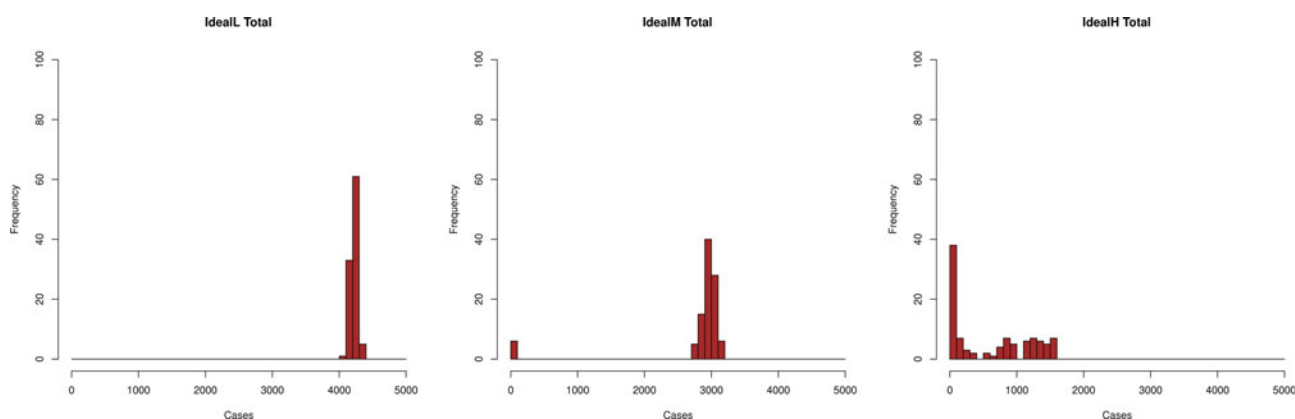
As expected, the epidemic size depends strongly on the policy applied. It is interesting to show the transition of the probability distribution as a function of the intensity of the control measures. In Figure 8, we show histograms after 100 simulations for the total number of infections in a population of 5000 individuals. Although extinctions of the epidemic with a low number of cases are always possible, they are infrequent in the low intensity case, they begin to

be notable in the medium intensity and are dominant in the high intensity situation. This transition is known as the stochastic equivalent of the transition in deterministic equations when the basic reproductive number moves from above one to below one and has been discussed elsewhere [32–34].

The number of *detected cases*, i.e. people diagnosed as infected with SARS-CoV-2, depends on no trivial form of the detection policy and intensity. We show in Figure 9 that the relation is not monotonic. In general, an increase in detection efficiency from medium to high intensity may result in a decrease in the number of cases detected (policies  $F+$  and  $I$ ) but also in an increase ( $H$ ). In fact, the design of the model only assures that the *probability* of detection increases with increasing intensity. If the total number of cases is low, the total number of detected cases will also be low, despite a higher detection probability. We can see as well that the efforts made with a passive policy ( $H$ ) produce only little changes in the development of the epidemic.



**Fig. 7.** Daily cases. Observable delays in the development of outbreaks for three individual realisations. Left,  $F + M$  policy presenting a ‘long’ waiting time until the outbreak develops. Centre,  $IH$  policy with immediate second wave. Right,  $IH$  policy with delayed second wave.



**Fig. 8.** Transition of probability (frequency, %) distribution for the  $I$  (Ideal) policies as a function of the intensity of control measures. Left: Low detection rate. Centre: Medium detection rate. Right: High detection rate.

## Discussion

In the present model, biological aspects are intertwined with sanitary policies. These policies are not considered in terms of their desired effects translated as effective parameters of an otherwise free-running epidemic but rather mechanistically, changing not only parameter values but the structure of the model as well. By doing so, we allow policies to manifest not only what they were intended for, but also unexpected features. The same can be said with respect to the coupling between intrinsic randomness and dynamics which results not only in the expected daily fluctuations of the outbreak but presents low-frequency fluctuations as well. These low-frequency fluctuations account, by themselves, for the possibility of silent circulation of the virus for prolonged intervals of time as well as for the awakening of extinguished outbreaks due to contagion outside the simulated community. The effective coupling of control measures and intrinsic randomness represents an additional difficulty for any attempt at predicting the evolution of a single/particular epidemic outbreak.

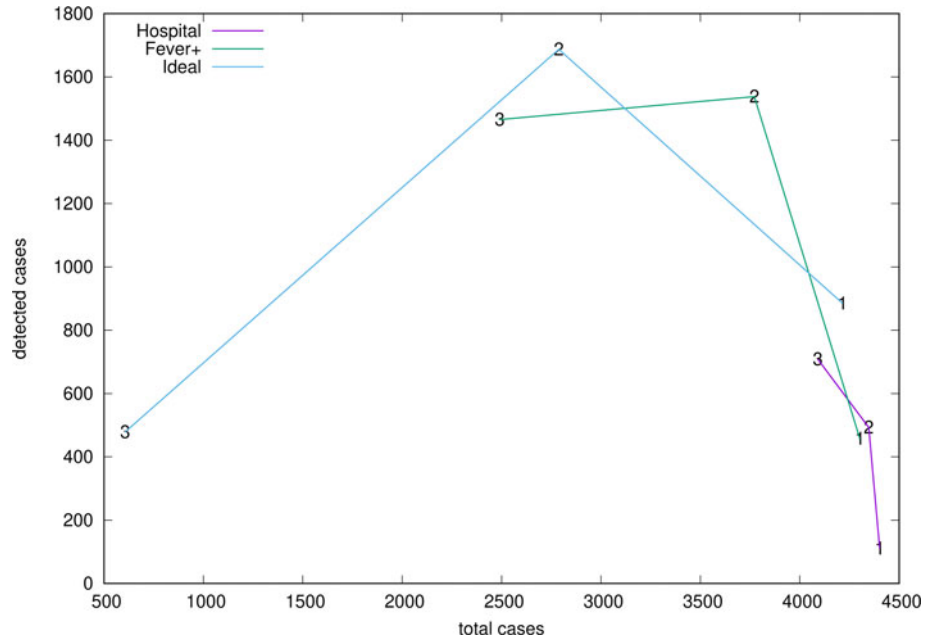
Although it should be clear from the setup adopted in this study, it is worth recalling that all realisations of a stochastic model are on an equal footing. Any of them respond to the process in its own right. A strength of the present approach is the capability of displaying a variety of possible epidemic outcomes. Indeed, Figures 4 (right) and 7 show that dramatic differences in epidemic size (for the same stochastic process), ‘second waves’ and late development of outbreaks are not unlikely to occur.

Figures 5 and 6 show that the ratio of undetected to detected cases is not constant in the course of an epidemic and even worse,

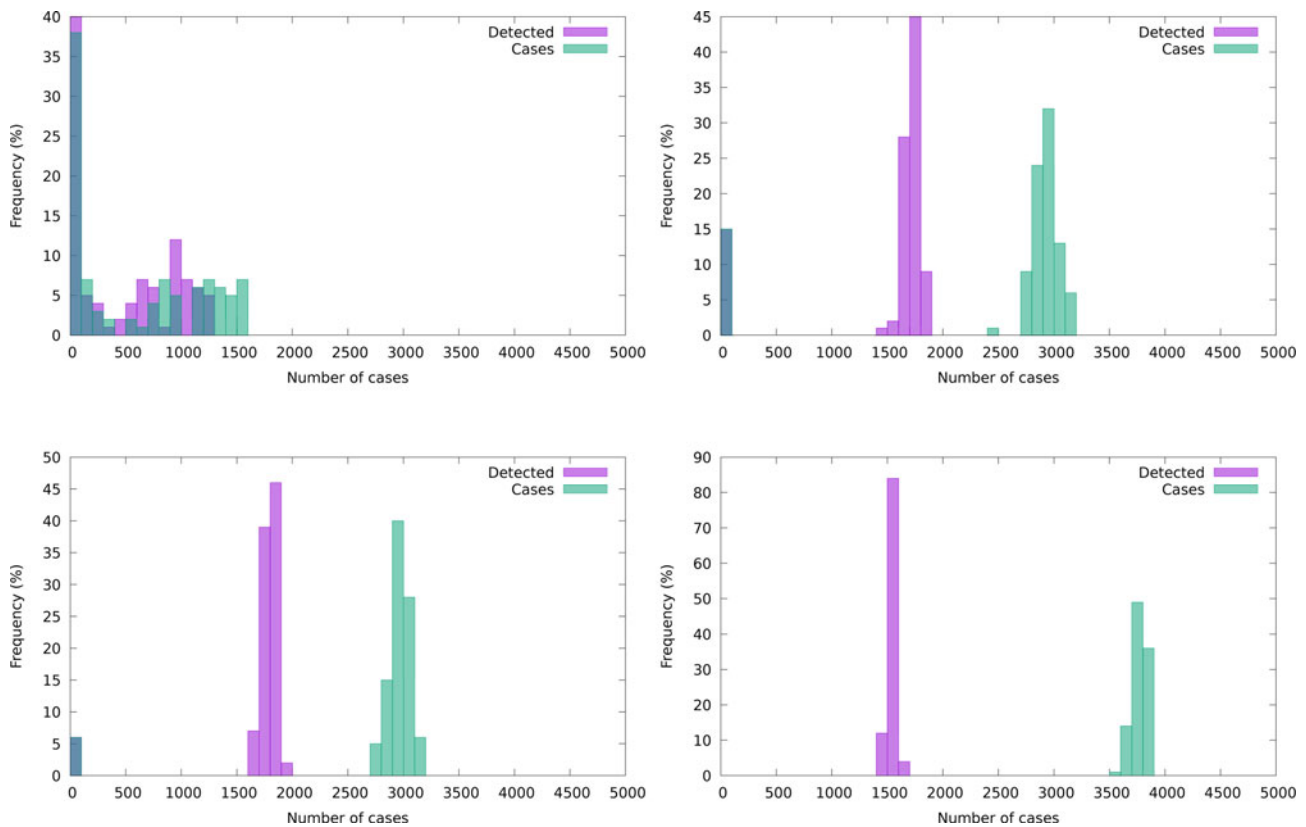
the growth rate of undetected cases is larger than that of the detected cases. Hence, epidemic size is likely to be underestimated when computed through recorded cases. Finally, Figure 8 illustrates how the stochastic outcomes can be translated into probabilities of e.g. having a given epidemic size.

One important goal of this study is to assist in the issue of resource allocation when dealing with a pandemic. HSs throughout the world differ in equipment, logistic capabilities, flexibility, etc., depending on the preexisting policies and infrastructure. The working conditions differ even locally within the same city, as discussed above. Where should resources go? Will low-cost (and lower efficiency) strategies under a longer period of time be preferred to high-cost (and higher efficiency) strategies with a shorter time-span?

In our model, we mimic the HS decisions by considering two groups of individuals: those that are early identified and recognised as potential patients,  $T$ , and the rest,  $U$ , of which the HS is initially unaware. We do not deal with financial costs, but we can compare highly-effective and less effective strategies throughout time. Preventive intervention strategies accrue costs in terms of isolating contagious people and testing. The kind of intervention considered in the current study is based upon tracking oligosymptomatic people but not searching for completely asymptomatic cases, thus a good indicator of costs is the total number of cases detected. The best strategy of all those considered in this respect is the  $I$  strategy with an efficient,  $H$ , search method, which is able to inhibit the development of outbreaks. The strategy has fewer detected cases and smallest overall size (see Fig. 8, right panel and Fig. 10, top left panel).



**Fig. 9.** Relation between detected and total cases corresponding to the average over 100 realisations of the policies *H*, *F+* and *I*, with three intensity levels labelled 1, 2, 3 in the plot and corresponding to low, middle and high detection levels.



**Fig. 10.** Frequency distribution (%) of detected and total cases. Top panel: *IH* and *F+H*. Bottom panel: *IM* and *F+M*.

We compare two distributions of the detected cases for equal efficiency of detection (see Fig. 10). The comparatively few detected in the *I* panel of Figure 10 constitute most of the outbreak, while *F+* adds a larger number of undetected cases. A short side of the *I* policy is that the effective suffocation of an epidemic outbreak in slum areas cannot by itself prevent

recurrent late outbreaks triggered by external contagion (see Fig. 11). Hence, the alert state of the HS will have to be maintained for longer times. However, the advantages of the *I* strategy under a medium or low detection success, thus having a higher failure rate in avoiding outbreaks, are not so considerable. As it can be seen in the figure, combining a suboptimal policy with a

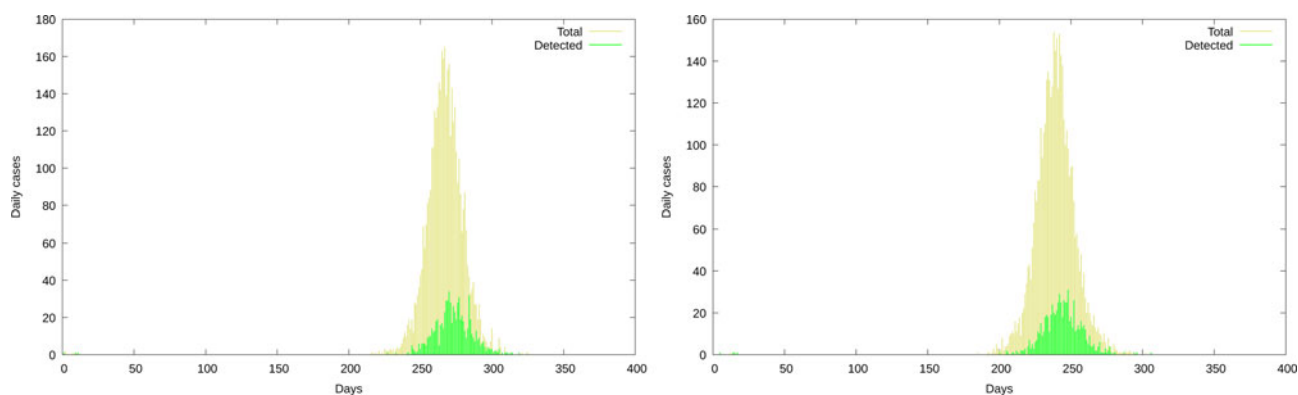


Fig. 11. Initial (seeded outbreak) dies out but after an external infection a 'delayed' outbreak emerges.

suboptimal tracking ( $F + M$ , bottom right panel) is expected to be more cost effective than an optimal policy with suboptimal tracking ( $IM$ , bottom left panel) in terms of detection, although the overall size of the epidemic is expected to be lower in the  $I$  situation, while the necessary social effort is larger.

Despite our simulations have been seeded in all cases with two infected people and only epidemic outbreaks that do not get extinguished for 19 days have been considered, some runs do not develop an outbreak and some others produce an outbreak only because of external contagion (from outside of the simulated neighbourhood), an effect that can occur in a completely different time scale as is shown in Fig. 11.

Also, we illustrate that the 'average epidemic' is not enough to grasp the relevant diversity of possible scenarios of real (unique) outbreak dynamics, and that the undetected (mild, unrecognised, presymptomatic and 'asymptomatic') cases are in good proportion the result of public policies coupled to the characteristics of the illness.

### Limitations

As mentioned previously, one of the assumptions of the present model is the homogeneity of contacts through the population. For that reason, it only makes full sense when applied to small communities. The proper path to surpass this constraint is to raise the level of details, identifying subpopulations with some common property (e.g. age segregation, mobility, local confinement, etc.) that are in weakly mutual interaction. This is a costly approach from the point of view of experimental design, since each new level of detail demands a detailed understanding of the specific interactions. Some effort in this direction has been to identify 'superspreaders', a possibility that recently became interesting [35].

### Conclusions

The modelling goal of this study was to conceive mechanisms for the interplay of the epidemic disease and the adopted social measures. The epidemic is not just biologically given in terms of e.g. a basic reproductive number or a herd-immunity level that are taken to be virus-specific and independent of social organisation. The COVID-19 pandemic is not a free-running epidemic or one addressed with pre-established measures, but rather one where dynamically evolving interventions are the rule. The way in which interventions change the dynamics and the observations (monitoring) of the epidemic must be considered. All too often

the analysis of the epidemic by the political components of the HS's, in practice consider a biologically given epidemic, and interventions that affect its evolution only in the way they were intended. This reasoning leads to false alternatives between herd immunity, vaccines and confinement, with the hidden assumption that social behaviour cannot (or must not) be modified. On the contrary, this study suggests that what we (collectively) do influences the level of risk to which we are exposed. Social behaviour can modify epidemic outcomes [36].

We observe that an increment in the number of daily detected cases does not necessarily imply an improvement in how the epidemic is being managed, nor a worsening of the outbreak. Case detection cannot be understood separately from the HS policy. Lower detection may be an indicator of success in the proper context. Hence, to translate the statistics for one country to another country is far from straightforward. More locally, the transfer of information from detected (registered) cases to estimated number of cases from seroprevalence studies is not independent of the adopted HS policy and depends as well of the timing with respect to the development of the outbreak.

Randomness plays a substantial role in COVID-19 dynamics, a role that departs from the signal + noise analysis framework. Low frequency, or coherent fluctuations, is relevant at the level of outbreaks in slums and there is no reason to believe the same is not going to be true in larger, heterogeneous, settings. The immediate consequence is that averaging and uncontrolled 'approximations' to the average outbreak will be aligned with intuitions but could be misaligned with a reality displaying a largely unpredictable form. The stochastic behaviour is affected as well by the social management of the epidemic, coupling two usually neglected contributions and making prediction of outcomes even more difficult.

The intervention of health authorities had been 'from below' in most countries. By 'from below', we mean a sequence of interventions going from non-intervention and passing through increasing levels of action until reaching lock downs in desperation. Such an approximation has to be revised; it is an approach that privileges something different than people's health. If our model is correct, it is possible to control the outbreaks with interventions that target mostly the symptomatic population. Such a method will have to target for isolation of any one presenting a single symptom of those compatible with COVID-19. The cost of more certainties is to lose control of the outbreak, being forced to apply lock downs, thus immobilising the productive forces of the healthy people rather than the comparatively small group that is potentially infected by SARS-CoV-2.

The decision of requiring more symptoms to declare a case as COVID-19 suspect whenever the patient has no identified contact with confirmed cases facilitates the circulation of the virus even when a highly efficient detection protocol is used.

Asymptomatic cases quite often are undetected cases as well, but the reverse is not true. Classifying or referring to undetected patients as asymptomatic can be viewed as an ethical matter. The term 'asymptomatic' puts the blame on the virus and helps to dispense social failures. In contrast, 'undetected' places the burden on society and should help to fix attention in what we can do better. Thus, if we are forced to err because of incomplete information the way we err must be ethically considered.

**Acknowledgements.** HGS acknowledges support from the University of Buenos Aires. MAN acknowledges support from Kungliga Fysiografiska Sällskapet i Lund (2018–2020). We thank the members of the 'Red de modelización de enfermedades infecciosas' (CONICET, Argentina) for valuable discussions, and particularly Verónica Simoy for help in tracking public policy in Argentina and Ignacio Simoy who has helped us to analyse the database using the R language.

**Data availability statement.** The data that support the findings of this study are available from the Argentine Ministry of Health, according to the law. Data were provided to the authors through the COVID-19 initiative of the Argentine National Science Council (CONICET). The authors can be contacted in case readers experience difficulties obtaining the data from the authorities.

## References

1. **China CDC** (2020) The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)-China, 2020. *Weekly Vital Surveillances* **2**, 113–122.
2. **McKendrick AG** (1914) Studies on the theory of continuous probabilities, with special reference to its bearing on natural phenomena of a progressive nature. *Proceedings of the London Mathematical Society* **s2–13**, 401–416.
3. **Feller W** (1940) On the integro-differential equations of purely discontinuous Markoff processes. *Transactions of the American Mathematical Society* **48**, 488–515.
4. **Kendall DG** (1950) An artificial realization of a simple 'birth-and-death' process. *Journal of the Royal Statistical Society. Series B (Methodological)* **12**, 116–119.
5. **Bartlett MS** (1964) The relevance of stochastic models for large-scale epidemiological phenomena. *Applied Statistics* **13**, 2–8.
6. **Kurtz TG** (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal Applied Probabilities* **7**, 49–58.
7. **Kurtz TG** (1971) Limit theorems for sequences of jump processes approximating ordinary differential equations. *Journal Applied Probabilities* **8**, 344–356.
8. **Kurtz TG** (1976) Limit theorems and diffusion approximations for density dependent Markov chains. *Mathematical Programming Studies* **5**, 67.
9. **van Kampen NG** (1981) The validity of nonlinear Langevin equations. *Journal of Statistical Physics* **25**, 431–442.
10. **Solari HG and Natiello MA** (2021) 'Normal science and the hidden lemmas of mathematical epidemiology'. In: *Mathematics for human flourishing in the time of COVID-19 and post COVID-19*. DeGruyter Proceedings. ISBN 978-3-11-073862-9 (In press). DeGruyter.
11. **Bai Y et al.** (2020) Presumed asymptomatic carrier transmission of COVID-19. *JAMA* **323**, 1406–1407.
12. **Cereda D et al.** (2020) The early phase of the COVID-19 outbreak in Lombardy, Italy 2020. *arXiv*. 2003.09320 [q-bio.PE].
13. **Huang L et al.** (2020) Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16–23 years outside Wuhan and characteristics of young patients with COVID-19: a prospective contact-tracing study. *Journal of Infection* **80**(6), e1–e13.
14. **Nishiura H et al.** (2020) Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Diseases* **94**, 154.
15. **Sakurai A et al.** (2020) Natural history of asymptomatic SARS-CoV-2 infection. *New England Journal of Medicine* **383**, 885–886.
16. **Mizumoto K et al.** (2020) Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **25**, 2000180.
17. **Yang R, Gui X and Xiong Y** (2020) Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in Wuhan, China. *JAMA Network Open* **3**, e2010182–e2010182.
18. **Böhmer MM et al.** (2020) Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases* **20**, 920–928.
19. **Hao X et al.** (2020) Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**, 420–424.
20. **Ethier SN and Kurtz TG** (1986) *Markov Processes*. New York: John Wiley and Sons.
21. **Woelfel R et al.** (2020) Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster. *medRxiv* doi: 10.1101/2020.03.05.20030502. eprint: <https://www.medrxiv.org/content/early/2020/03/08/2020.03.05.20030502.full.pdf>.
22. **Combes AJ et al.** (2021) Global absence and targeting of protective immune states in severe COVID-19. *Nature* **591**, 124–130.
23. **Durrett R** (2001) *Essentials of Stochastic Processes*. New York: Springer Verlag.
24. **Solari HG and Natiello MA** (2014) Linear processes in stochastic population dynamics: theory and application to insect development. *The Scientific World Journal – Journal of Probability and Statistics*. Vol **2014**, Article ID 873624, 1–15.
25. **Natiello MA and Solari HG** (2020) Modelling population dynamics based on experimental trials with genetically modified (RIDL) mosquitoes. *Ecological Modelling* **424**, 108986.
26. **Saito M and Matsumoto M** (2009) *Monte Carlo and Quasi-Monte Carlo Methods*. Springer. doi:10.1007/978-3-642-04107-5\_38.
27. **Li Q et al.** (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine* **382**, 1199–1207.
28. **Malani A et al.** (2020) Seroprevalence of SARS-CoV-2 in slums and non-slums of Mumbai, India, during June 29–July 19, 2020. *medRxiv* doi: <https://doi.org/10.1101/2020.08.27.20182741>.
29. **Muñoz L et al.** Vigilancia y seroprevalencia: evaluación de anticuerpos igg para SARS-CoV-2 mediante elisa en el barrio popular villa azul, quilmes, provincia de Buenos Aires, Argentina (2020). *SciELO Preprints*.
30. **Sekine T et al.** (2020) Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* **183**, 158–168.
31. **Mick E et al.** (2020) Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nature Communications* **11**, 1–7.
32. **Nåsell I** (1996) The quasi-stationary distribution of the closed endemic SIS model. *Advances in Applied Probability* **28**, 895–932.
33. **Nåsell I** (2002) Measles outbreaks are not chaotic. In Chavez CC (ed.), *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory*, vol. **126**. New York: Springer, pp. 199–227.
34. **Nåsell I** (2001) Extinction and quasi-stationarity in the Verhulst logistic model. *Journal of Theoretical Biology* **211**, 11–27.
35. **Edholm CJ et al.** (2018) *Searching for Superspreaders: Identifying Epidemic Patterns Associated with Superspreading Events in Stochastic Models*. Cham: Springer International Publishing, chap. 1, pp. 1–29.
36. **Bavel JVV et al.** (2020) Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* **4**, 460–471.

## Appendix A: Criteria of suspicious case in Argentina

The early evolution of the criteria (we omit the specifications for sanitary operators) for a case to be suspicious in Argentina is as follows (we indicate correspondence with Italy's resolutions):

29 February 2020 (Corresponds to Italy's January resolution)

There are two ways to be considered a suspected case:

1. The person has fever and signs of respiratory infection (cough and difficulty in breathing) and a requirement for hospitalisation and no other aetiology that fully explains the clinical presentation and a history of travel or residence in mainland China in the 14 days prior to the onset of symptoms.

2. The person has fever and signs of respiratory infection (cough, odynophagia and difficulty in breathing) and either a history of travel or residence in the province of Hubei (China) in the 14 days prior to the onset of symptoms or known close contact with a probable or confirmed case of COVID-19 infection, in the 14 days prior to the onset of symptoms or exposure in a health centre in a country where confirmed cases of COVID-19 have been attended, in the 14 days prior to the onset of symptoms or visited or worked in a live animal market in any city in China.

5 March 2020 (Corresponds to Italy's late-February resolutions)

The person has fever and one or more respiratory symptoms (cough, difficulty in breathing and odynophagia) without another aetiology that fully explains the clinical presentation, and that in the last 14 days either has been in contact with confirmed or probable COVID-19 cases or has a history of travel or presence in areas with local transmission of SARS-CoV-2

16 March 2020 (Corresponds approximately to Italy's 9 March resolution) In addition to the previous situation, a new suspected case begins to be considered: any person with severe acute respiratory disease who requires mechanical ventilation due to their respiratory symptoms, without other aetiology that explains it, even without epidemiological link.

21 March 2020

Travel history to specific countries is substituted by travel abroad. Specifications are given for severe acute respiratory disease, defined as pneumonia and one of the following:

Respiratory rate: >30/min

Sat O<sub>2</sub> <93% (ambient air)

Mechanical assistance requirement

Increase in infiltrates >50% in 24–48 h

Altered consciousness

CURB-65 ≥2 points

Unit of intensive therapy requirement,

and without another aetiology that explains the clinical picture.

—  
30 March 2020

The issue of travel or residence in areas of local transmission (either community or by conglomerates) of COVID-19 in Argentina is added.

In the case of pneumonia, the other conditions raised on 21 March are no longer required.

8 April 2020

Cases added:

Health workers presenting fever and one symptom among (cough, odynophagia or difficulty in breathing)

16 April 2020

To the requirements raised on 5 March, in the case of symptoms that could accompany fever, the following is added: anosmia/dysgeusia.

13 May 2020

Anyone presenting fever (37.5 °C or more) and one or more of (cough, odynophagia, difficulty in breathing, anosmia and dysgeusia) of recent presentation, without another clinical explanation AND precedents of travelling to (or residence in) places with viral circulation or contact with confirmed cases of COVID-19

Anyone presenting anosmia/dysgeusia is to be observed by 72 h and then tested.

Health workers with two or more of the described symptoms.

6 June 2020

Anyone presenting two or more of (fever – 37.5 °C or more–, cough, odynophagia, difficulty in breathing, anosmia and dysgeusia) AND (having being present in a zone with viral circulation OR residing in a 'popular neighbourhood' – slum –) OR (requiring hospitalisation)

Health workers and any with close contact with a COVID-19 case presenting at least one symptom.

4 August 2020

Three symptoms added to the set: headache and vomits and diarrhoea.

11 September 2020

Added symptom: myalgia

Anyone presenting fever (37.5 °C or more) and one or more of (cough, odynophagia, difficulty in breathing, anosmia and dysgeusia) of recent presentation, without another clinical explanation.

Anyone presenting anosmia/dysgeusia.

For health workers and inhabitants of 'popular neighbourhoods' the requisite is of one symptom.