

The estimated impact of the COVID-19 epidemic in the general population of France

Laurent TOUBIANA^a Jacques BOUAUD^{b,a}

^a *Inserm, Sorbonne Université, Université Sorbonne Paris Nord, UMR S_1142, LIMICS, Paris, France,*

^b *AP-HP, DRCI, Paris, France*

All authors contributed equally to the work

Correspondence to:

Laurent TOUBIANA

INSERM UMRS 1142 LIMICS

Campus des Cordeliers

15, rue de l'école de médecine

75006 Paris

France

laurent.toubiana@inserm.fr

Abstract

Background Model-based predictions for the COVID-19 outbreak revealed the potential for extraordinary mortality and saturation of health care systems if no action was taken. The pandemic hit France in late January 2020. Lockdown was implemented on March 17, 2020. The goal was to drastically reduce the number of infections, but at the cost of fear of a second epidemic wave when easing the lockdown. Our aim was to characterize the dynamics of the COVID-19 spread in France and estimate the proportion of coronavirus-infected individuals using ground truth from syndromic surveillance data.

Methods National health authorities provide data from syndromic surveillance of the diagnosis of suspected COVID-19 reported by a sample of primary care physicians and from epidemic surveillance of confirmed cases, originating from hospitals. By extrapolation, COVID-19 incidence in the general population can be estimated. In turn, a back-calculation model can infer the number of contagious individuals, providing an idea of the spread of the epidemic before the implementation of lockdown measures.

Results This study estimated that 12.3 million individuals were diagnosed 'suspected COVID-19' on May 6, 2020. At lockdown start, 2.5 million people were already contagious. The infection attack rate peaked on March 27, ten days after lockdown. The predicted sharp drop was not observed. The dynamics of the epidemic followed a continuous curve with a decline phase 2.35 times slower than the growth phase. 80% of infections occurred after lockdown.

Conclusions These results call into question the effectiveness of lockdown. The epidemic would have followed its 'natural trajectory', beginning long before the health system detected the first cases. This hypothesis does not dispute the caution required with regard to the extraordinary spread of the epidemic, with less affected geographic areas becoming a source of susceptible individuals.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Introduction

From December 31, 2019, information became available on the existence of cases of an acute respiratory disease in the city of Wuhan, in the Hubei Province of China.¹ It was announced that it was caused by an emergent coronavirus, subsequently called SARS-CoV-2. The number of cases increased rapidly. In a few days, these not many Wuhan cases led to an epidemic, which spread around the world in a few weeks. On March 11, 2020, it was declared a pandemic by the World Health Organisation, and took the official name of COVID-19. As the virus was new, there were no means of treatment or vaccination. In France, the first cases of COVID-19 were initially confirmed on January 24, 2020.² This was followed by a slow spread becoming exponential, similar to what was observed with a time lag in other countries, particularly in Italy and Spain.

As soon as data became available, first on the Chinese epidemic and then on its outbreaks in other regions, numerous studies were carried out in an emergency to identify the parameters of this new pandemic.

Theoretical estimates primarily focused on the impact of the epidemic in terms of morbidity, mortality and need for hospital care for COVID-19 patients according to the implementation of various containment strategies. Predictions obtained for worst-case scenarios have revealed the potential for extraordinary mortality and the total saturation of health systems. For example, 2,200,000 covid-19 deaths in the US, 510,000 in the UK were predicted if no action was taken.³ For France specifically, studies have reported figures ranging from 74,000⁴ to over 300,000 deaths.^{5,6}

Health authorities and governments of the impacted countries followed the China's initial response to this health crisis that have consisted on social distancing of individuals. Regardless of the kind of measures taken, sometimes drastic such as the confinement of an entire country, the principle consisted of reducing contacts between infected cases and susceptible cases to a minimum. France thus implemented a containment of the general population on March 17, 2020.

The epidemic peak was reached in France between March 23 and 29.⁷ Knowledge of the proportion of the population already infected is of primary importance in order to anticipate the occurrence of one or more new epidemic waves. The herd immunity required for natural extinction of the epidemic is estimated at 60-70% of the population. Several predictions have been published for different countries based on epidemiological models. For France, the Imperial College London team estimates this percentage at 3.48% as of May 7, 2020.⁸ A French team predicts this percentage to be 4.4% as of May 11, 2020.⁶ For the same date and according to another French team, this percentage would be around 14%.⁵

Syndromic surveillance methods make it possible to estimate the incidence of a pathology in the general population.⁹ Such syndromic surveillance are routinely used in France to seasonal influenza, gastroenteritis, chickenpox among others from activity tracing of samples of general practitioners.¹⁰ The surveillance network for influenza syndromes revealed between late February and early March, at the end of an influenza epidemic, an unexpected increase in reported influenza-like illness correlated with covid-19 cases, suggesting a greater circulation of SARS-CoV-2 than that inferred from confirmed cases.¹¹ The back-calculation methods was initially proposed to forecast the spread of AIDS in the population based on surveillance data on diagnosed cases and on estimates of the incubation period.^{12,13,14}

The aim of this study was to assess the dynamics of the COVID-19 spread and to estimate the prevalence of coronavirus-infected cases using extrapolations to the French population and back-calculation methods. We used the National health authorities combining syndromic surveillance of the diagnosis of suspected COVID 19 at primary health care level and epidemic surveillance of confirmed cases, originating from hospitals.

The predicted immunity rate has been discussed with respect to other published estimates. This was an opportunity to question the effectiveness of the lockdown measure on the dynamics of the epidemic and the plausibility of the resurgence of the epidemic after lockdown exit.

Materials and methods

Data sources

The data used for this study are available online on official websites, including the French Government website¹⁵ and the website of Santé Publique France (SPF), the National Agency of Public Health.¹⁶ The study period started on February 24, 2020, when data became available.

Hospital data consisted in the total number of patients hospitalised for confirmed COVID-19 and of those in intensive care units on any given day at the country level.

Syndromic data (from SurSaUD) relied on the activity of the physicians of SOS Médecins,¹⁷ a nationwide medical emergency service. Such data related to COVID-19 started to be recorded on March 3, 2020 in 62 SOS Médecins local associations¹⁸ scattered throughout the National territory and routinely transmitted on a daily basis to SPF. Time series of the number of daily medical acts with the diagnosis of 'suspected COVID-19' as well as the total number of daily medical acts were used for the extrapolation of the total estimated number of COVID-19 cases in the general population.

National and International data concerning the cumulative number of confirmed cases, deaths, and recoveries were accessible on the websites of SPF and of Johns Hopkins University (JHU),¹⁹ as well as in WHO reports²⁰ and numerous other online sources.

Definitions

We adopt the following definitions for the rest of the article:

'health care system':	any medical authority, health professionals detecting or counting COVID-19 patients, whether confirmed or suspected,
'detected case':	any individual with clinical signs, detected through syndromic surveillance by the health care system, actually considered and recorded by a healthcare professional as a suspected or confirmed case of COVID-19,
'confirmed case':	any individual counted as officially affected by COVID-19; this category is a subset of the prior 'case' category,
'hospitalised case':	any 'confirmed case' who required hospitalisation,
'estimated cases',	the 'cases' in the general population resulting from the extrapolation of detected 'cases'.

For the purpose of this study, we also define the different stages in which an individual can be in relation to the disease over time. These stages roughly have the same names than those used in classical SIR-based dynamic models to account for epidemic spread. However, these names do not entirely cover the same concepts. While in a dynamic model, they refer to the variables of systems of differential equations, here they refer to the states in a state-transition diagram describing at the individual level the progression of the disease using a finite-state machine. The different possible stages are as follows and their transitions are shown in Figure 1.

- 'S' (susceptible stage): At the onset of the epidemic, an individual has never been exposed to the virus. The person is not sick, but susceptible to contracting the disease.
- 'T' (incubation stage): The person is exposed to the virus and becomes infected. The person presents no symptoms of the disease. During a so-called latency period (p_i), the person is neither sick nor contagious.
- 'C' (contagious stage): After this latency period, viral shedding starts and the individual passes into a contagious stage without having entered a clinical state (no symptoms of the disease yet). However, the individual can now transmit the virus to all his/her contacts.
- 'D' (disease stage): Some individuals start feeling the effects of the disease after the incubation period (p_i) and then enter the clinical or symptomatic stage. It is only from this moment that the individual becomes detectable and may be identified by the health care system and counted as case. Since the patient is a confirmed case, he or she is contagious, before and during the clinical signs over the duration of the contagious period (p_c), but normally no longer transmits the disease from this date as he or she is taken care of and is being isolated.
- 'R' (remission stage): At the end of the contagious period, the person enters the final stage of remission in the model since he cannot spread the virus anymore, regardless of the clinical outcome.

Hereafter, we will use the following notations:

- t_i : Date of infection.
- t_c : Start of the contagious period.
- t_d : Date of detection by the healthcare system.
- t_r : End of the contagious period.

- p_i : Incubation period corresponding to the time between the date of infection t_i (start of stage 'T') and t_d of detection by the health system (start of stage 'D').
- p_c : Contagious period, going from t_c to t_r (exit at the 'R' stage)
- p_e : Period of effective contagiousness going from the date of contagiousness t_c to t_d detection by the health system (transition to stage 'D').
- p_l : Latency period corresponding to the time between the date of infection t_i (entry into stage 'T') and the start of contagiousness t_c (passage to stage 'C').

By extension we will refer to the incidences (i.e. the number of new cases per period) in each category by the letter of the corresponding stage: S, I, C, D, R. It is important to note that

ground truth data are only available for stage 'D', from which the individual is taken care of by the health care system and becomes reported in the surveillance system as a case.

Back-calculation methods for estimating the number of contagious individuals

Back-calculation methods allow deducing C based on D using a duration relationship between the two stages. In fact, as soon as the individual is taken care of by the health system, the case is counted as a new category 'D' case. If the patient was detected on date t_d , this means he was contaminated on date $t_i = t_d - p_i$ and became contagious on date $t_c = t_d - p_e$. We consider that an individual detected by the health care system remains contagious during the entire period of contagiousness. However, the patient should no longer be counted as contagious from date t_d as he is taken care of and therefore isolated. This hypothesis most certainly reduces the number of contagions. We note $C(t)$ and $D(t)$ the incidence of C and D at date t , and $\chi(t)$ the cumulative number of contagious individuals at date t .

The following temporal equations are obtained:

$$\begin{aligned} (1) \quad C(t) &= D(t_2) && \text{with } t_2 = t + p_e \\ (2) \quad \chi(t) &= \sum_{x=t_s}^{t_f} D(x) && \text{with } t_s = t \text{ and } t_f = t + p_e \end{aligned}$$

The periods required for the numerical application of these equations are not well known. We took the average value and their range found in the literature²¹.

p_i : Incubation period:	7 (95% CrI: 4–10) days
p_l : Latency period:	2 (95% CrI: 1–3) days
p_e : Period of effective contagiousness:	5 (95% CrI: 3–9) days

Extrapolation of recorded syndromic data to the general population

We applied the analytical methods used by epidemiological reference observatories in France for the general population, such as the French Syndromic Sentinel Network^{10, 22} or the Research Institute for the exploitation of health data (IRSAN)²³. IRSAN has in particular utilised data from the records of the physicians of SOS Médecins since 2012. Using survey methods on diagnoses collected daily from a sample of general practitioners in France, these observatories estimate the incidence of a disease in the general population, noted $D_g(t)$. Over 1,000 physicians participate in the collection of some 10,000 medical acts per day on average. The algorithms for calculating^{24,25} the aforementioned indicators have been implemented and routinely obtain results disseminated in real time for more than thirty years for many syndromes circulating in the population (e.g. influenza-like illnesses, bronchiolitis, gastroenteritis, and chickenpox).

The available time unit used is the day. The COVID-19 incidence rates were estimated from the population data available on the website of the French National statistics database INSEE.²⁶ All results are available on the IRSAN website.²⁷

Correlating ground truth data with population-based estimates

Ground truth data from all sources (cases, confirmed cases, and hospitalised cases) show the same dynamics, as well as the estimated cases in the general population derived from cases. This is hardly surprising given that hospitalised people are necessarily a reflection of people suspected of having been affected by the virus. We have established the proportionality coefficients of the incidences according to the source of the data. These coefficients allow us to estimate the values of the incidences in the general population regardless of the data source.

$$\beta_{gc} = \frac{\sum_{t=t_i}^{t_f} D_c(t)}{\sum_{t=t_s}^{t_f} D_g(t)}$$

$$\beta_{gh} = \frac{\sum_{t=t_s}^{t_f} D_h(t)}{\sum_{t=t_s}^{t_f} D_g(t)}$$

With:

$D_c(t)$: Number of new detected cases confirmed by SPF per 24-hour time unit.

$D_h(t)$: Number of new hospitalised cases.

$D_g(t)$: Estimated number of new cases diagnosed as (suspect COVID-19 in the general population in one day)

These coefficients then allow us to estimate the values of the incidence D_g in the general population from all ground truth data source. Under numerous assumptions about this model transposition to other population, this allows to estimate the epidemic cumulative incidence in other countries from, for instance, the number of their confirmed cases.

Sensitivity analysis

The incubation period and the effective period of contagiousness are not knowing with certainty. We conduct a sensitivity analysis of these parameters according to the values found in the literature. We estimated the number of infected individuals over time by varying each value of these parameters within the limits of the confidence intervals (see above).

This analysis allowed us to estimate an average value $\chi(t)$ with its confidence interval, a minimum and maximum value for each date between February 24 and May 6, 2020. Figures 4 and 5 represent the evolution of these values over the time. The same calculations were performed on the data from confirmed and hospitalised cases, giving the same orders of magnitude.

Results

Description of ground truth epidemic and syndromic data

At the time of writing (May 6, 2020) and since the epidemic onset in France, a total of 137,150 confirmed cases had been declared on the SPF website,¹⁵ 93,027 persons have been hospitalised and 16,360 have undergone intensive care.

From the syndromic surveillance network by SOS Médecins, a total of 47,242 patients have been detected and reported. Figure 2 shows the dynamics of the epidemic through the evolution of the daily incidences of collected indicators, detected, confirmed, and hospitalized cases of COVID-19.

Regardless of the source of the data, all curves follow the same profile: a growth phase followed by a phase of decline after reaching a peak. Peaks occurred with a few days interval, detected cases on March 27, hospitalised cases on April 1st, and confirmed cases on April 7, 2020.

The curves have very different amplitudes. This is not surprising, however, since the number of entries into intensive care is a subset of the number of hospitalised cases, which itself is a subset of the number of confirmed cases (fortunately, not all COVID-19 cases are hospitalised). As for detected cases, diagnosed as 'suspected COVID-19' at the primary care level, the curve is relatively regular. On the other hand, the curve representing the number of confirmed cases shows significant variations from one day to another. In fact, given the relatively low number of new cases confirmed during the epidemic period in France, these variations are due to differences in the recording of cases. The same applies to the number of hospitalised patients, which also shows a weekly structure, which is probably due to the differences in hospital activity during weekends. The growth phase of the epidemic is very rapid. In France, it reached its peak on March 27, 2020, i.e. only 25 days after the start of the data collection on March 3, 2020.

Estimation of COVID-19 spread in the French general population

From incidence data on detected cases, diagnosed as 'suspected COVID-19' by the primary care physicians of SOS Médecins, the equivalent numbers were inferred for the general population. Figure 3 provides the daily evolution of these estimated cases. The profile of the curve is the same as the one of detected cases, but the scale is larger. The epidemic reached its peak with 473,000 new estimated cases per day.

On May 6, 2020, we estimated that 12.4 (95% CrI: 11.62 –13.18) million people have been hit by COVID-19 in the general population. This represents 19% of the French metropolitan population. Table 1 summarizes the cumulative cases according to the indicator for 3 key dates: lockdown (March 17, 2020), epidemic peak (according to indicator), date of this study (May 6, 2020).

We therefore calculated that the number of confirmed cases as recorded by the authorities only represents $\beta_{gc} = 1.50\%$ (95% CrI: 1.07%–1.92%) of the impact of the epidemic estimated according to the number of cases diagnosed as 'suspected COVID-19' in the general population. As for the number of hospitalised patients, they only represent $\beta_{gh} = 0.80\%$ (95% CrI: 0.73%–0.87%) of this estimate.

Estimated number of contagious individuals in the general population

We carried out assessments of the cumulative numbers of contagious contaminated individuals $\chi(t_c)$ as of March 17, 2020, the day of lockdown in France, based on the estimated incidences of individuals diagnosed with 'suspected COVID-19' in the general population. To compensate for the imprecision of the information concerning these periods, we made several estimates considering plausible values for these parameters.

On March 17, 2020, while the estimated value of new suspected Covid-19 individuals is 202,262 individuals, the estimated number of contagious individuals was 2.5 (95% CrI: 2.3–2.7) million.

Theoretical projections of COVID-19 spread from confirmed cases for other countries

Coefficient $\beta_{gc} = 1.50\%$ (95% CrI: 1.07%–1.92%) allows an estimation of the number of 'suspected COVID-19' diagnoses in the general population by extrapolating from the number of confirmed cases recorded by the health authorities. We applied it to the numbers of confirmed cases available internationally. Table 3 shows for a few countries and for the world

as a whole, estimates of the total number (in millions) of individuals that would have been diagnosed with 'suspected COVID-19' if these countries had a health data collection system equivalent to that used in France.

Discussion

In this study, we mainly analysed real data from nationwide syndromic surveillance on COVID-19, made available by the French authorities in the same way as hospital and confirmed case data. Extrapolated results to the whole population were obtained from the same network of primary care physicians (SOS Médecins) and by the same methods as those used since 2013 on epidemics of influenza-like illness.²⁸ These same methods have been used to monitor the progression of recurrent epidemics in France for almost four decades.²⁹ To the authors' knowledge, this is the first assessment of the impact of the SARS-CoV-2 epidemic based on the measure of syndromic-related medical activity.

These ground truth data provide information on the impact of the epidemic on the French population in proportions that have never been shown before. Thus, we estimate that the number of individuals diagnosed with 'suspected COVID-19' in France on May 6, 2020 to be 12.4 million. One could object that this value be overestimated since suspected cases are counted and have not been confirmed. However, the detected suspected cases are not self-declared, but reported by health care professionals. They are expected to be somehow reliable. Besides, doubts also exist about the reliability of case confirmation³⁰ (sur la reliability des tests). Nevertheless, studies on flu epidemics have shown that influenza-like illness had a high predictive value for influenza although there might be other causes for this syndrome.²⁹ In our study, we used the same assumption, which could be a limitation since not validated of course. Even considering an error margin of 20%, our results for COVID-19 still remain considerable. By way of comparison, it should be noted that for influenza-like illness, on average less than 3 million symptomatic cases are diagnosed under the same conditions per episode. It is a fact that the COVID-19 epidemic has affected at least four times more people in France, confirming its unprecedented nature.

12.4 million people affected by the COVID-19 represent 19% of the French metropolitan population. Let us remind this estimate is produced on the factual basis of reported ground data; irrespective of any non-pharmaceutical intervention, including the lockdown and its expected outcome. This estimate obviously contrasts with the lower predictions of 3.48% (as of May 7, 2020)⁸ and 4.4% (as of May 11, 2020)⁶ of infected people. These latter models take into account the lockdown expected effect with a drop in hospitalisation needs in the line of sight. In terms of their dynamics, projections anticipated an immediate recession of infections on the day of lockdown from around 200,000 to 50,000 new daily cases. These predictions informed the political authorities in many countries, including France, and weighted on their decision to enact lockdowns.

We have shown that, on March 17, 2020, date of the lockdown, the viral circulation in the general population had already reached a considerable level. By back-calculation based on symptomatic individuals, we evaluated the number of contagious individuals to be 2.5 million on average at that date. This estimate is probably lower than reality for at least two reasons. First, we assumed viral transmission was interrupted as soon as the patient was taken care of by the health system. It can be presumed that some of the patients continued to transmit the virus, for instance within their family cluster or to their carers. Second, some individuals have

an asymptomatic or a paucisymptomatic form of the disease. They go undetected and can potentially transmit the virus silently during the entire period of contagiousness. It is acknowledged that the proportion of asymptomatic cases is the big unknown of this pandemic.³¹ Taking into account any value for this proportion would increase the number of contagious individuals and, as a consequence, considerably augment the further spread of the disease.

Indeed, the very principle of lockdown is to reduce the contact rate to stop the spread and protect the population of non-infected individuals. This principle automatically generates a perverse effect. As the virus is still circulating, lifting lockdown will automatically imply that this subpopulation will be exposed under the same conditions as before lockdown, assuming the immunity of the already infected subpopulation. Hence the fear of a strong second wave and the many precautions taken to progressively ease the lockdown.

If there is no second wave, this will support our hypothesis. The epidemic would have followed its 'natural trajectory', which began long before the health system detected the first cases.^{5,6} It is conceivable that a large part of the initially susceptible population was infected very early on. The fear of having artificially maintained a huge number of susceptible people due to lockdown would have to be qualified if our hypothesis proves to be correct.

However, this fear is not entirely lifted. Indeed, the epidemic spread has an asymmetrical geographical configuration. Parts of France have been strongly affected, in particular eastern and northern regions and the Paris area, other parts much less so. To date, there aren't many alternative hypotheses to the lockdown effect. It would therefore be conceivable that the western and southern areas of France, which were less affected during the first period of the epidemic, might be affected after the lift of lockdown. In that case, we would observe localized increases of the incidence of symptomatic cases from the week that follows.¹⁰

Once the epidemic is over, it will be possible to assess the impact of different strategies to contain the epidemic by precisely comparing the data that will then be available for all countries. In the meantime, the data we have to date already give us enough information to question the effectiveness of the lockdown strategy on the progression of the COVID-19 epidemic, at least in France.

References

1. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
2. Stoecklin, S. B. *et al.* First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Eurosurveillance* **25**, 2000094 (2020).

3. Ferguson, N. *et al.* *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.* 20
<http://spiral.imperial.ac.uk/handle/10044/1/77482> (2020) doi:10.25561/77482.
4. Massonnaud, C., Roux, J. & Crépey, P. COVID-19: Forecasting short term hospital needs in France. *medRxiv* 2020.03.16.20036939 (2020) doi:10.1101/2020.03.16.20036939.
5. Hoertel, N. *et al.* Lockdown exit strategies and risk of a second epidemic peak: a stochastic agent-based model of SARS-CoV-2 epidemic in France. *medRxiv* 2020.04.30.20086264 (2020) doi:10.1101/2020.04.30.20086264.
6. Salje, H. *et al.* Estimating the burden of SARS-CoV-2 in France. *Science* (2020) doi:10.1126/science.abc3517.
7. Santé Publique France. COVID-19 : point épidémiologique du 14 mai 2020. /maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-14-mai-2020 (2020).
8. Flaxman, S. *et al.* *Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries.* 35
<http://spiral.imperial.ac.uk/handle/10044/1/77731> (2020) doi:10.25561/77731.
9. Mandl, K. D. *et al.* Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience. *J. Am. Med. Inform. Assoc. JAMIA* **11**, 141–150 (2004).
10. Valleron, A. J. *et al.* A computer network for the surveillance of communicable diseases: the French experiment. *Am. J. Public Health* **76**, 1289–1292 (1986).
11. Boëlle, P.-Y. *et al.* Excess cases of influenza-like illnesses synchronous with coronavirus disease (COVID-19) epidemic, France, March 2020. *Eurosurveillance* **25**, (2020).
12. Brookmeyer, R. & Gail, MitchellH. Minimum size of acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *The Lancet* **328**, 1320–1322 (1986).

13. Deuffic-Burban, S. & Costagliola, D. Including pre-AIDS mortality in back-calculation model to estimate HIV prevalence in France, 2000. *Eur. J. Epidemiol.* **21**, 389–396 (2006).
14. Rice, B. D. *et al.* Monitoring of the HIV Epidemic Using Routinely Collected Data: The Case of the United Kingdom. *AIDS Behav.* **21**, 83–90 (2017).
15. French Government website, Info Coronavirus Covid 19 (in French); (2020).
16. Santé publique France (GEODE). (2020).
17. SurSaUD, the French surveillance system of emergencies and deaths. (2020).
18. Meurice, L. *et al.* General Practitioner House Call Network (SOS Médecins): An Essential Tool for Syndromic Surveillance – Bordeaux, France. *Prehospital Disaster Med.* **35**, 326–330 (2020).
19. CSSEGISandData csse_covid_19_time_series. (2020).
20. WHO novel coronavirus 2019 situation-reports.
21. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
22. Flahault, A. *et al.* Virtual surveillance of communicable diseases: a 20-year experience in France. *Stat. Methods Med. Res.* **15**, 413 (2006).
23. Toubiana, L. & Griffon, N. Some Innovative Approaches for Public Health and Epidemiology Informatics. *Yearb. Med. Inform.* 247–250 (2016) doi:10.15265/IY-2016-047.
24. Method IRSAN Recherche for Impact assessment in général population (French). <http://recherche.irsan.fr/fr/documentation/index/voir/58-Origines-des-donn%C3%A9es-et-m%C3%A9thodes>.
25. French syndromic Sentinel Network; <https://www.sentiweb.fr>.
26. website of the national statistics database INSEE.

27. IRSAN COVID-19 Research Institute for the exploitation of health data Covid-Website.
(2020).
28. Toubiana, L., Guerin, P., Chansard, P. & Demoor, C. Detection and monitoring of outbreaks of influenza-like illness using data of SOS Médecins France 2006-2014. in *IMED 2014* (2014).
29. Souty, C. *et al.* Influenza epidemics observed in primary care from 1984 to 2017 in France: A decrease in epidemic size over time. *Influenza Other Respir. Viruses* **13**, 148–157 (2019).
30. Holmdahl, I. & Buckee, C. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *N. Engl. J. Med.* **0**, null (2020).
31. Gandhi, M., Yokoe, D. S. & Havlir, D. V. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. *N. Engl. J. Med.* **0**, null (2020).

Tables and figures

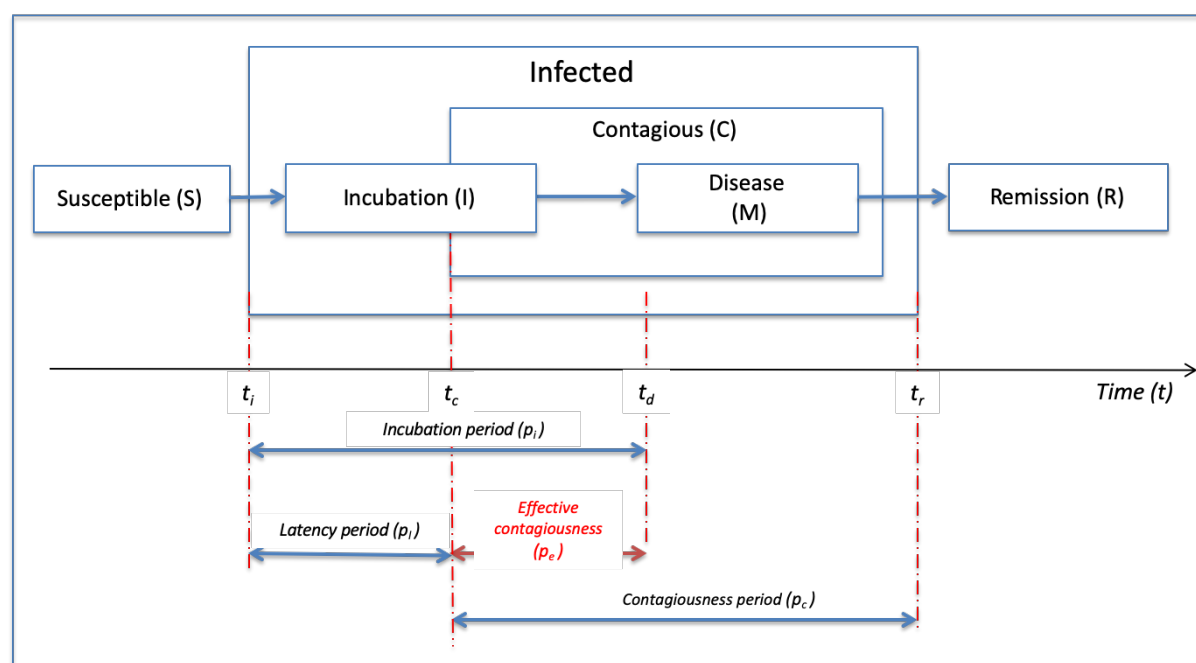


Figure 1: Time transition diagram of states of an individual infected with a virus

This diagram describes the different states over time and the associated transitions for an infected individual.

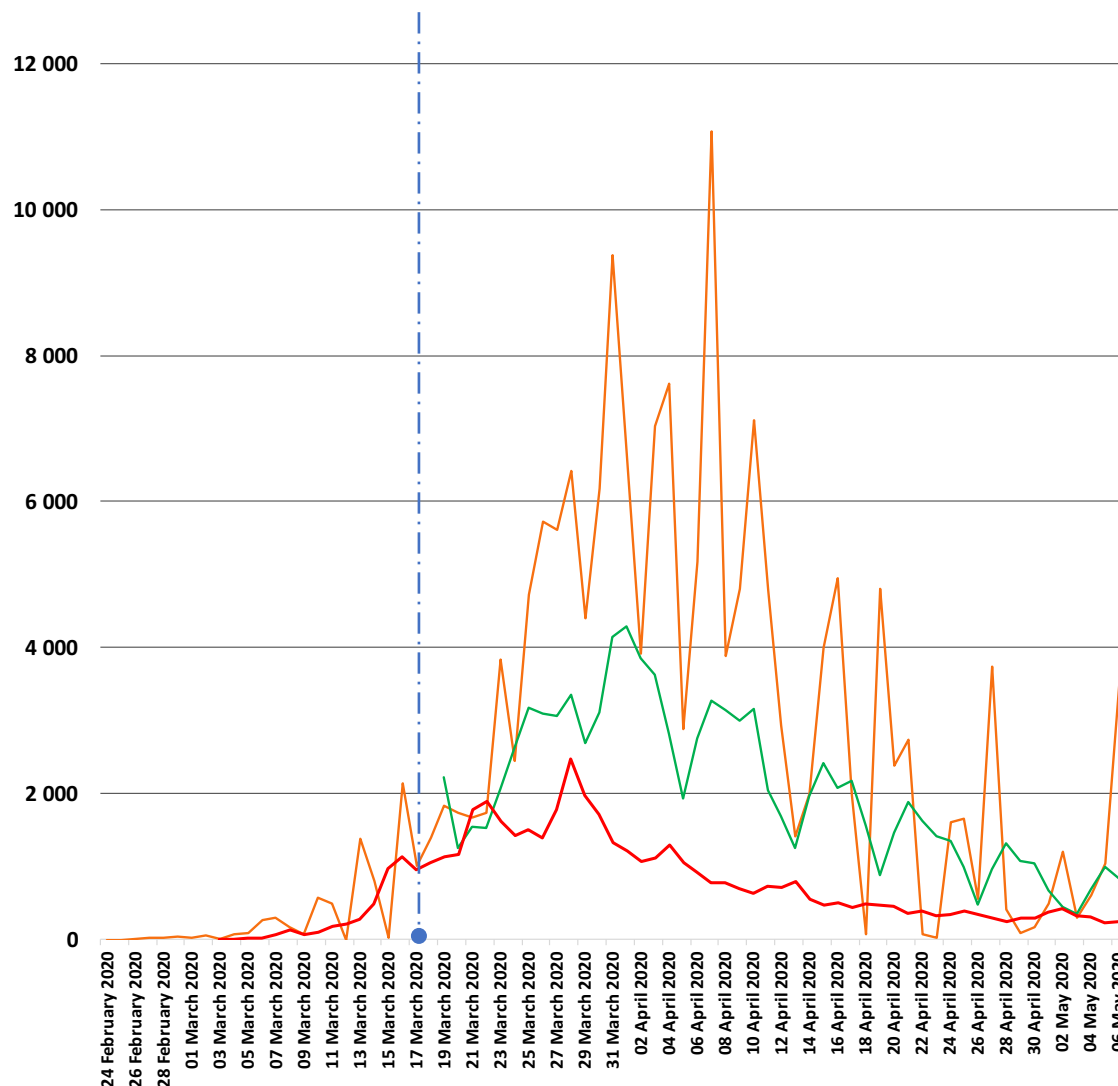


Figure 2: Evolution of the daily incidences of detected, confirmed and hospitalized cases

The orange curve represents the evolution of the daily incidences of cases of COVID 19 confirmed by health authorities in France between February 24 and May 6, 2020. The green curve represents the evolution of the daily incidences of hospitalized COVID 19 cases (data are only available from 19 March 2020). The red curve represents the evolution of detected cases. The vertical blue dotted line corresponds to the date of confinement (March 17, 2020).

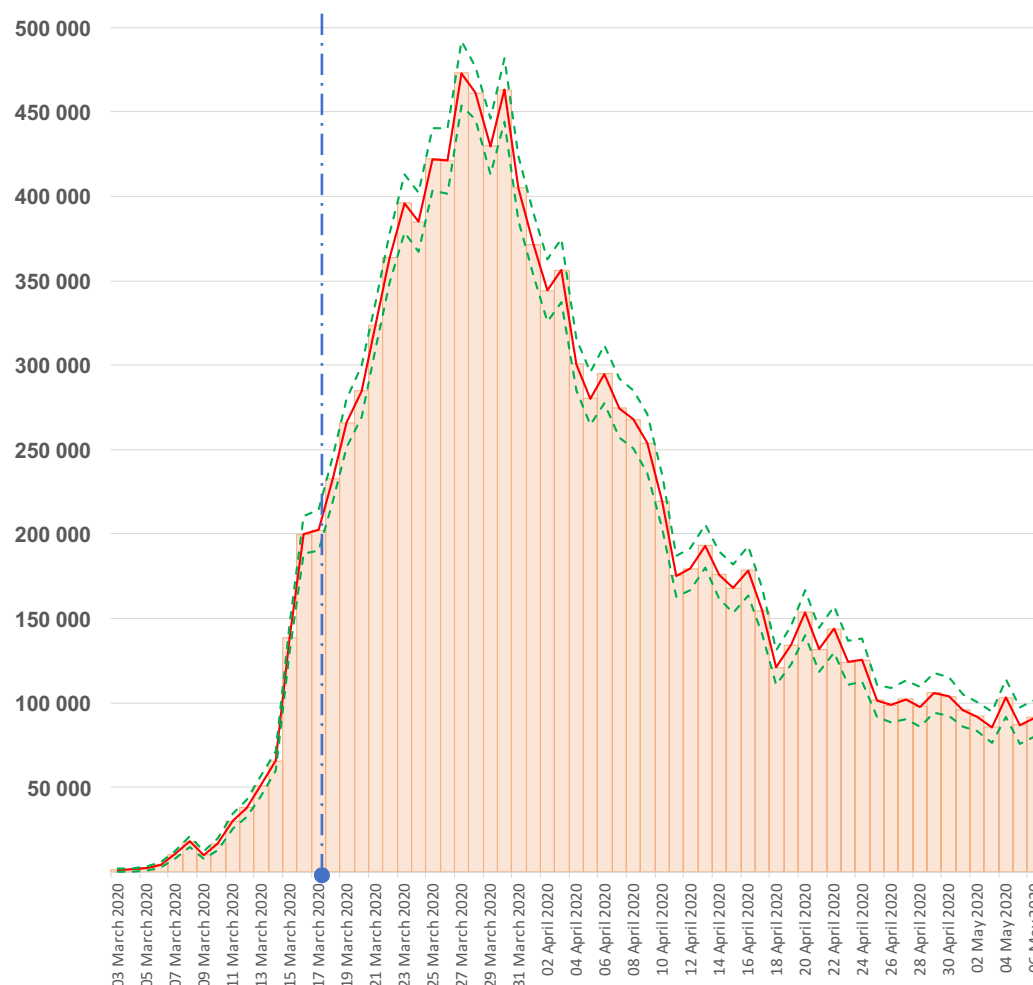


Figure 3: Evolution of the daily incidences of estimated cases of COVID 19 in the general population

The red curve represents the evolution of the daily incidences of cases of COVID 19 suspected by general practitioners and emergency physicians outside hospitals in France between March 3 and May 6, 2020. The vertical blue dotted line corresponds to the date of confinement (March 17, 2020). The green curve represents, on the same scale, the evolution of the daily incidences of COVID 19 confirmed cases by health authorities in France. At this scale, this incidence is barely visible.

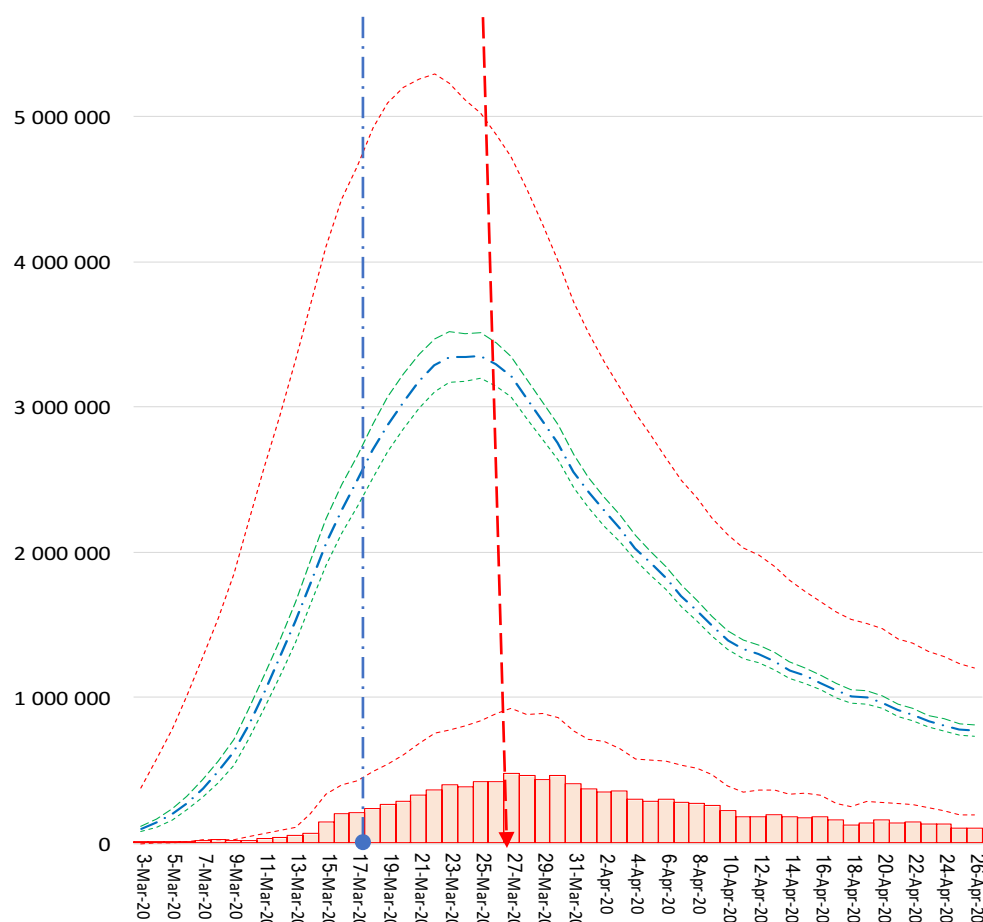


Figure 4: Evolution of the daily incidence of contagious cases $\chi(t)$ estimated in the general population

The red bar plot represents the evolution of the daily estimated incidences of cases of COVID 19 suspected by general and emergency physicians outside the hospital in France between February 24 and April 13, 2020. The blue dotted line corresponds to the date of confinement (March 17, 2020). The red vertical arrow indicates the epidemic peak reached on March 27, 2020. The blue dotted curve represents the evolution of the number of contagious individuals in the French population estimated by retro-calculation according to the incidences of symptomatic cases detected in the general population. On the day of confinement, around 2.5 (95% CrI: 2.3–2.7) million people were infected with the virus and contagious (see Table 2). The green dotted curves are the 95% confidence intervals. The red dotted curves are the minimum and maximum values of the estimated number of contagious individuals

Table 1: Cumulative number of cases in the general population at the date of lockdown, of epidemic peak and end of study

Event	Lockdown		Epidemic peak		End of study
	March 17	March 27	April 1 st	April 7	May 6, 2020
Detected cases	4 619	19 368			47 242
Confirmed cases	7 652			89 941	137 150
Hospitalised cases	NA		38 120		97 322
Estimated cases	788 510	4 355 248			12 399 005

Table 3: Estimates by country of the total number of individuals (in million individuals with 95% confident interval) who have been diagnosed with ‘suspected COVID-19’ as of May 6, 2020.

Country	Reported number of confirmed cases	Estimated number of ‘Suspected Covid-19’ (million)	95% CrI
China	83 970	5.6	[4.4 - 7.8]
Korea	10 806	0.7	[0.6 - 1]
France	172 465	11.5	[9 - 16]
Italy	214 457	14.3	[11.1 - 20]
UK	201 101	13.4	[10.5 - 18.7]
Sweden	23 918	1.6	[1.2 - 2.2]
USA	1 229 331	82.0	[63.9 - 114.4]
World	3 756 065	250.5	[195.2 - 349.4]

We show for 5 countries and for the whole world, an estimate of the total number (in million) of individuals who would have been diagnosed "suspected of COVID-19" if there was a syndromic surveillance system equivalent. Each number is accompanied by its confidence interval. For example, for France we find a value of 11.5 (95% CrI: 9–16) million individuals affected by applying the coefficient β_{gc} on international data from the Johns Hopkins University website, which is consistent with the 12.4 (95% CrI: 11.6–13.2) million individuals calculated directly on SURSAUD data collection.