

Underdetection of cases of COVID-19 in France threatens epidemic control

<https://doi.org/10.1038/s41586-020-03095-6>

Received: 11 August 2020

Accepted: 8 December 2020

Published online: 21 December 2020

 Check for updates

Giulia Pullano^{1,2,7}, Laura Di Domenico^{1,7}, Chiara E. Sabbatini¹, Eugenio Valdano¹, Clément Turbelin¹, Marion Debin¹, Caroline Guerrisi¹, Charly Kengne-Kuetche¹, Cécile Souty¹, Thomas Hanslik^{1,3,4}, Thierry Blanchon¹, Pierre-Yves Boëlle¹, Julie Figoni⁵, Sophie Vaux⁵, Christine Campèse⁵, Sibylle Bernard-Stoecklin⁵ & Vittoria Colizza^{1,6}✉

As countries in Europe gradually relaxed lockdown restrictions after the first wave, test–trace–isolate strategies became critical to maintain the incidence of coronavirus disease 2019 (COVID-19) at low levels^{1,2}. Reviewing their shortcomings can provide elements to consider in light of the second wave that is currently underway in Europe. Here we estimate the rate of detection of symptomatic cases of COVID-19 in France after lockdown through the use of virological³ and participatory syndromic⁴ surveillance data coupled with mathematical transmission models calibrated to regional hospitalizations². Our findings indicate that around 90,000 symptomatic infections, corresponding to 9 out of 10 cases, were not ascertained by the surveillance system in the first 7 weeks after lockdown from 11 May to 28 June 2020, although the test positivity rate did not exceed the 5% recommendation of the World Health Organization (WHO)⁵. The median detection rate increased from 7% (95% confidence interval, 6–8%) to 38% (35–44%) over time, with large regional variations, owing to a strengthening of the system as well as a decrease in epidemic activity. According to participatory surveillance data, only 31% of individuals with COVID-19-like symptoms consulted a doctor in the study period. This suggests that large numbers of symptomatic cases of COVID-19 did not seek medical advice despite recommendations, as confirmed by serological studies^{6,7}. Encouraging awareness and same-day healthcare-seeking behaviour of suspected cases of COVID-19 is critical to improve detection. However, the capacity of the system remained insufficient even at the low epidemic activity achieved after lockdown, and was predicted to deteriorate rapidly with increasing incidence of COVID-19 cases. Substantially more aggressive, targeted and efficient testing with easier access is required to act as a tool to control the COVID-19 pandemic. The testing strategy will be critical to enable partial lifting of the current restrictive measures in Europe and to avoid a third wave.

Surveillance and detection aim to rapidly identify and isolate cases to prevent onward transmission of SARS-CoV-2 in the community and to avoid a substantial resurgence of cases of COVID-19. After an initial period—during which, because of a limited capacity, testing for SARS-CoV-2 infections mainly focused on severely ill patients—a new testing policy was implemented in France to systematically screen for potential infections with SARS-CoV-2 and enable lifting of the lockdown restrictions on 11 May 2020⁸.

The specific characteristics of COVID-19, however, hinder the identification of cases^{9–11}. Large proportions of asymptomatic infectious individuals¹², and the presence of mild or paucisymptomatic infections that easily go unobserved^{9,11}, present serious challenges to the detection and control of SARS-CoV-2^{9,10,13}. Missing a substantial portion of infectious individuals compromises the control effort, enabling the

virus to silently spread^{10–12}. Synthesizing evidence from virological³ and participatory syndromic surveillance⁴ with mathematical models^{2,14} that account for behavioural data^{15–18}, we assessed the performance of the new testing policy in France and identified its main limitations for actionable improvements.

COVID-19 surveillance

Management of the COVID-19 pandemic in France after lockdown in spring (May–June) 2020 involved the generation of a centralized database that collected all data on virological testing (SI-DEP³, the information system for testing). All individuals with symptoms that were compatible with COVID-19¹⁹ were invited to consult their general

¹INSERM, Sorbonne Université, Institut Pierre Louis d’Epidémiologie et de Santé Publique, IPLESP, Paris, France. ²Orange Labs, Sociology and Economics of Network and Services (SENSE), Chatillon, France. ³UFR des Sciences de la Santé Simone-Veil, Université Versailles-Saint-Quentin-en-Yvelines, Versailles, France. ⁴AP-HP, Service de Médecine Interne, Hôpital Ambroise Paré, Boulogne Billancourt, France. ⁵Santé publique France, Direction des maladies infectieuses, Saint-Maurice, France. ⁶Tokyo Tech World Research Hub Initiative, Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan. ⁷These authors contributed equally: Giulia Pullano, Laura Di Domenico. ✉e-mail: vittoria.colizza@inserm.fr

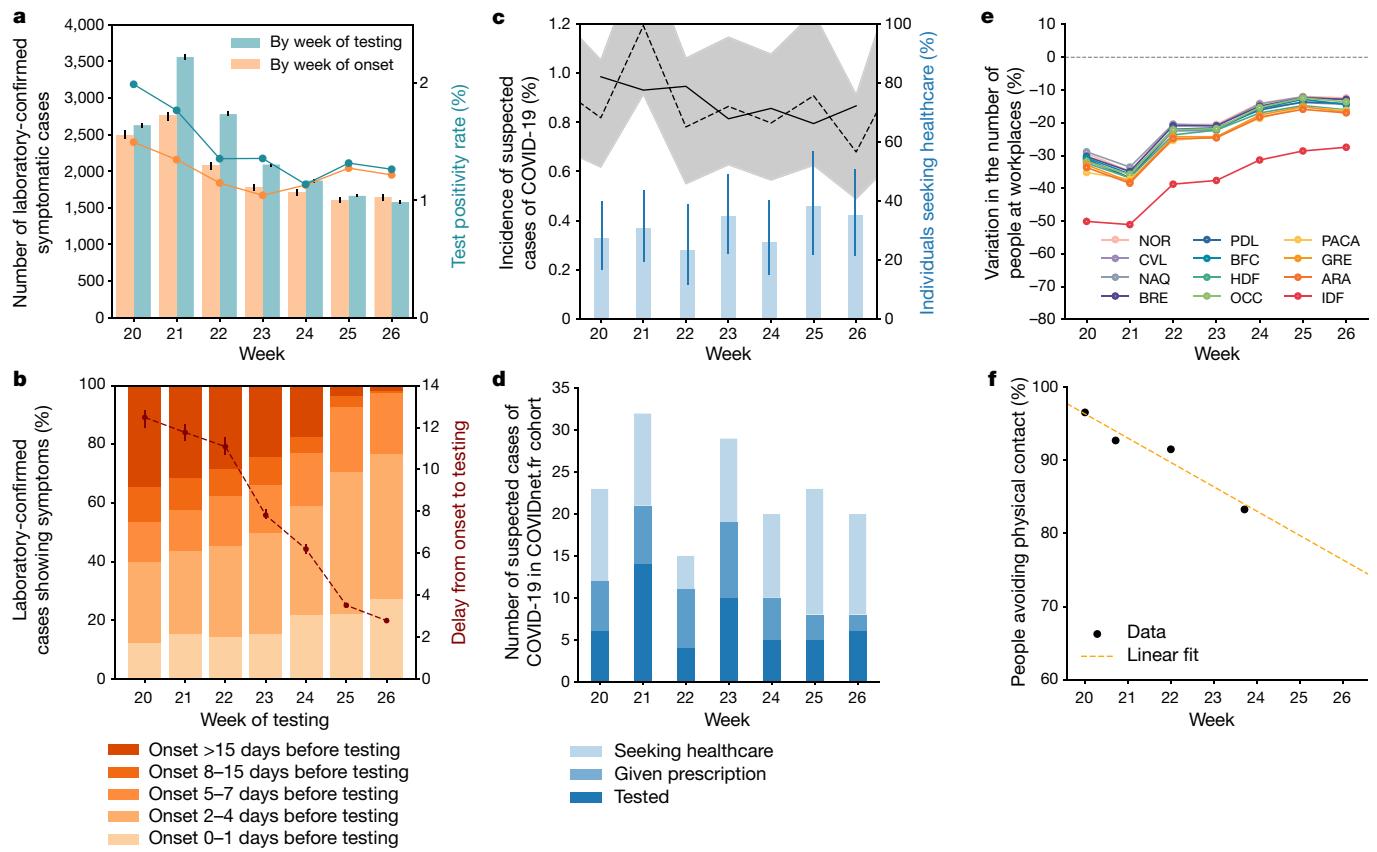


Fig. 1 | Virological surveillance, participatory syndromic surveillance and behavioural data for model parameterization. **a**, Estimated number of virologically confirmed symptomatic cases in mainland France by week of testing and week of onset (bar graphs), and test positivity rate (line graphs). Estimates are based on the imputation of individuals without symptoms who tested positive at the time of testing into asymptomatic or presymptomatic; imputation of missing data on clinical status at the time of testing into asymptomatic, presymptomatic or symptomatic; and imputation of the date of onset of symptoms for presymptomatic and symptomatic cases (Methods). Imputations were performed $n=100$ times. Uncertainties (black bars) correspond to the 95% confidence intervals. Test positivity rates were computed for cases with complete information. Data for weeks 20–26 were consolidated in week 30. **b**, Breakdown of virologically confirmed cases with symptoms and complete information in the SI-DEP database by week of testing according to the declared onset of symptoms (left y axis; $n=5,514$).

practitioner and obtain a prescription for a virological test⁸. Contacts of confirmed cases were traced and tested. A total of 20,777 virologically confirmed cases were notified from 13 May (week 20) to 28 June (week 26) in mainland France. These cases included individuals with or without symptoms at the time of testing who tested positive for SARS-CoV-2 or individuals who tested positive for SARS-CoV-2 for whom information on clinical status at the time of testing was missing (Extended Data Fig. 1). Accounting for presymptomatic individuals among those presenting with no symptoms at the time of testing and after imputation of missing data (Methods), an estimated 16,165 (95% confidence interval, 16,101–16,261) symptomatic cases were tested in the study period (Fig. 1a). The average delay from symptom onset to testing decreased from 12.5 days in week 20 to 2.8 days in week 26 (Fig. 1b and Extended Data Fig. 1). Accounting for this delay (Methods and Extended Data Fig. 2), we estimated that 14,061 (13,972–14,156) virologically confirmed symptomatic cases had an onset of symptoms in the study period, showing a decreasing trend over time (2,493 in week 20, 1,647 in week 26). The test positivity rate decreased in the first weeks and stabilized at around 1.2% (mean over weeks 24–26).

The estimated time from onset to testing is also shown (right y axis; median and 95% confidence interval obtained from $n=100$ imputations of the onset date). **c**, Weekly incidence of suspected cases of COVID-19 (median (dashed line), 95% confidence interval (shaded area) and 3-week moving average (solid line)), and percentage of individuals seeking healthcare (median and 95% confidence interval), estimated from the participatory surveillance system, COVIDnet.fr (average weekly $n=7,481$). **d**, The number of suspected cases of COVID-19 in the participatory cohort who sought healthcare, and among those individuals, the number of individuals who received a prescription and performed a virological test when given the prescription. **e**, Estimated change in presence at workplace locations over time and by region based on Google location history data¹⁷. Region acronyms are listed in Table 1. **f**, Percentage of individuals avoiding physical contact with respect to lockdown, estimated from a large-scale survey conducted by Santé publique France¹⁸.

A digital participatory system was additionally considered for COVID-19 syndromic surveillance in the general population²⁰, including those who did not consult a doctor. Called COVIDnet.fr, it was adapted from the platform GrippeNet.fr (which is dedicated to the surveillance of influenza-like illnesses⁴) to respond to the COVID-19 health crisis in early 2020. It is based on a set of volunteers who weekly self-declare their symptoms, along with sociodemographic information. On the basis of symptoms declared by an average of 7,500 participants each week, the estimated incidence of suspected cases of COVID-19¹⁹ decreased from about 1% to 0.8% over time (Fig. 1c). Of 524 suspected cases, 162 (31%) consulted a doctor in the study period. Among them, 89 (55%) received a prescription for a test, resulting in the screening of 50 individuals (56% of those given the prescription) (Fig. 1d).

COVID-19 pandemic trajectories and detection rates

We used stochastic discrete age-stratified epidemic models^{2,14} based on demography, age profile²¹ and social contact data¹⁵ of the 12 regions of mainland France to account for age-specific contact activity and role in

Article

Table 1 | Population, confirmed and projected symptomatic cases, estimated detection rate and trends

Region	Acronym	Population (millions)	Number of laboratory-confirmed symptomatic cases by week of onset	Number of projected symptomatic cases by week of onset		Estimated detection rate (%) for symptomatic cases	Trend in detection rate
				Week 20	Week 26		
Île-de-France	IDF	12.3	737	574	12,427 (8,104–14,136)	1,704 (1,258–2,004)	6 (5–9) +146%
Grand Est	GRE	5.5	323	135	4,868 (2,992–5,848)	756 (568–914)	7 (6–11) +99%
Hauts de France	HDF	6.0	308	225	4,476 (2,381–6,648)	396 (219–538)	7 (5–13) +186%
Auvergne-Rhône-Alpes	ARA	8.0	204	181	3,552 (2,017–5,283)	312 (173–451)	6 (4–10) +244%
Occitanie	OCC	5.9	166	106	851 (397–1,400)	128 (57–235)	19 (12–42) +165%
Provence-Alpes-Côte d'Azur	PACA	5.1	164	73	3,040 (1,665–4,625)	157 (83–239)	5 (4–10) +289%
Pays de la Loire	PDL	3.8	127	96	1,158 (463–1,846)	255 (103–423)	11 (7–27) +45%
Bourgogne-Franche-Comté	BFC	2.8	118	36	1,591 (854–2,379)	154 (88–235)	7 (5–14) +95%
Nouvelle Aquitaine	NAQ	6.0	115	43	1,040 (482–1,691)	94 (38–166)	11 (7–24) +54%
Centre-Val de Loire	CVL	2.6	94	44	1,706 (812–2,511)	79 (34–142)	6 (4–12) +187%
Brittany	BRE	3.3	80	23	672 (294–1,155)	113 (51–206)	12 (7–27) -28%
Normandy	NOR	3.3	55	112	725 (322–1,194)	153 (63–258)	8 (5–17) +342%
France ^a		64.6	2,493	1,647	35,704 (30,290–40,748)	4,319 (3,773–4,760)	7 (6–8) +142%

Regions are ranked by decreasing number of confirmed cases in week 20. The trend is computed comparing the average of the estimated detection rate in the weeks of June (weeks 23–26) with the average in the weeks of May (weeks 20–22). For the number of projected symptomatic cases and the estimated detection rate, data are medians and 95% confidence intervals obtained from $n = 500$ independent stochastic runs.

^aData are for mainland France; Corsica and overseas territories were excluded.

COVID-19 transmission. Disease progression is specific to COVID-19^{2,14} and parameterized using the current knowledge to include presymptomatic transmission²², and asymptomatic¹² and symptomatic infections with different degrees of severity^{9,11,23,24}. The model was shown to capture the transmission dynamics of the pandemic in Île-de-France in the first wave and was used to assess the effect of lockdown and exit strategies^{2,14}. Full details are reported in the Methods.

Intervention measures were modelled as mechanistic modifications of the contact matrices, accounting for a reduction in the number of contacts engaged in specific settings, and were informed from empirical data. Lockdown data were obtained from previously published studies^{2,14}. The exit phase was modelled considering region-specific data of school attendance based on the data from the Ministry of Education¹⁶, partial presence at workplaces based on estimates from location history data of mobile phones¹⁷ (Fig. 1e), a reduction in the adoption of physical distancing over time and the increased risk aversion of older individuals based on survey data¹⁸ (Fig. 1f), and the partial reopening of activities. A sensitivity analysis was performed on the reopening of activities, as data were missing for an accurate parameterization of associated contacts. Testing and isolation of detected cases were implemented by considering a 90% reduction in contacts for the virologically confirmed cases of COVID-19^{2,14}. Region-specific models were fitted to regional hospital admission data (Fig. 2) using a maximum likelihood approach. Further details are reported in the Methods and Supplementary Information.

The projected number of cases decreased over time in all regions, in agreement with the decreasing tendency reported in hospital admissions during the study period (Fig. 2 and Extended Data Fig. 3). Overall, 103,907 (95% confidence interval, 90,216–116,377) new symptomatic infections were predicted in mainland France in weeks 20–26 (from 35,704 (30,290–40,748) in week 20 to 4,319 (3,773–4,760) in week 26). Île-de-France was the region with the largest predicted number of symptomatic cases (from 12,427 (8,104–14,136) to 1,704 (1,258–2,004) from week 20 to week 26), followed by Grand Est and Hauts-de-France (Table 1 and Extended Data Table 1).

Projections were substantially higher than the number of virologically confirmed cases (Figs. 2, 3). The estimated detection rate for

symptomatic infections in mainland France in the period of weeks 20–26 was 14% (12–16%), suggesting that about 9 out of 10 new cases with symptoms were not identified by the surveillance system. A lower detection rate was found for asymptomatic infections (Extended Data Fig. 5). The estimated detection rate increased over time (7% (6–8%) in week 20, 38% (35–44%) in week 26) (Table 1). By the end of June, five regions had a median detection rate above 50%, and six regions had a detection rate within the confidence interval of model projections (Fig. 3b–d). All regions except Brittany displayed average increasing trends in the estimated detection rate in June compared with May. We did not find any significant associations between the detection rate and the number of detected cases, or the test positivity rate (Extended Data Fig. 4). However, the detection rate was negatively associated with model-predicted incidence (Spearman correlation, $r = -0.75$, $P < 10^{-15}$) (Fig. 3f). In addition, the data followed a power-law function, $\pi = 66 \times i^{-0.51}$, where π is the weekly detection rate of symptomatic cases (expressed as a percentage) and i the projected weekly incidence (number of cases per 100,000). This function quantifies the relationship between the detection capacity of the test–trace–isolate system and the circulation of the virus in the population. It clearly shows that the detection capacity rapidly decreases as the incidence of COVID-19 increases.

Validation of the model was performed in two ways. First, we compared our model projections of the percentage of the population infected with the results of three independent seroprevalence studies performed after the first wave in France^{7,25,26} (Methods). Modelling results are in agreement with serological estimates at the national and regional level (Fig. 3e and Extended Data Fig. 6). Second, we compared the projected incidence of symptomatic cases of COVID-19 in week 26 (6.69 (5.84–7.37) cases per 100,000) with the value obtained from the number of virologically confirmed cases (2.55 (2.48–2.61) cases per 100,000) and two estimates based on COVIDnet.fr data (Fig. 3g). The first estimate applies the measured test positivity rate to the incidence of self-reported suspected cases of COVID-19 (estimate 1, which yielded 8.6 (95% confidence interval, 6.2–11.5) cases per 100,000); the second additionally assumes that only 55% would be confirmed as a suspected case by a physician and prescribed a test (according to COVIDnet.fr

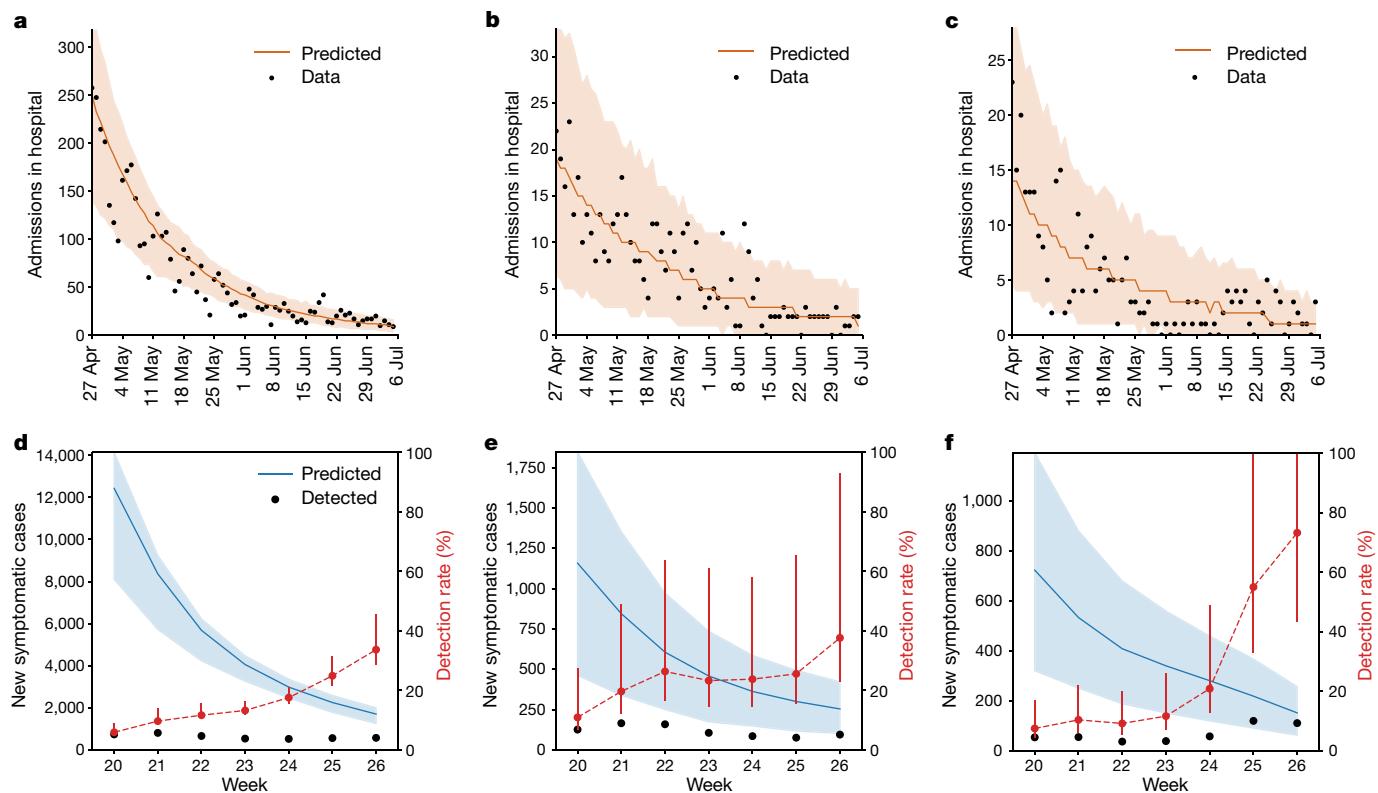


Fig. 2 | Hospital admissions and number of new symptomatic cases. **a–c**, Hospital admissions over time; data (points) and simulations (median and 95% confidence intervals) for Île-de-France (**a**), Pays de la Loire (**b**) and Normandy (**c**). Hospital admission data up to week 27 (consolidated in week 28) were used to infer parameter values. **d–f**, Projected number of new symptomatic cases over time (median and 95% confidence interval) and estimated number of virologically confirmed symptomatic cases by week of

onset (points), for the same regions (Île-de-France (**d**), Pays de la Loire (**e**) and Normandy (**f**) (left y axis). The estimated detection probability of symptomatic cases (%) is also shown (red points, median and 95% confidence interval; right y axis). In all panels, 95% confidence intervals were obtained from $n = 500$ independent stochastic runs. Plots for the remaining regions are shown in Extended Data Fig. 3.

data; estimate 2, which yielded 4.7 (3.4–6.3) cases per 100,000). Our projections are in line with plausible estimates from COVIDnet.fr, and suggest that, on average, at least 80% of suspected cases should be tested to reach the predicted incidence.

Sensitivity analysis showed that the findings were robust to elements of the contact matrices that could not be informed by empirical data (Supplementary Figs. 8, 9). Furthermore, a model selection analysis showed that changes in contact patterns over time due to restrictions and the activities of individuals of different age classes after lockdown (for example, partial attendance at school and remote working) are needed to accurately capture the transmission dynamics (Supplementary Table 2 and Supplementary Fig. 5).

Discussion

Despite a test positivity rate in mainland France well below the recommendations (5%) of the WHO⁵, a substantial proportion of symptomatic cases (9 out of 10) remained undetected in the first 7 weeks after lockdown.

Low detection rates in mid-May were in line with estimates for the same period from a seroprevalence study in Switzerland²⁷. Surveillance improved substantially over time, leading to half of the French regions reporting numbers of cases that were compatible with model projections. The framework progressively strengthened with increasing resources over time, as shown by a more-rapid detection of cases (78% reduction in the average delay from symptom onset to testing from May to June). At the same time, the system benefited from a substantial and concurrent decrease in epidemic activity in all regions.

Despite this positive trend, our findings highlight structural limitations and a critical need for improvement. Some areas remained with limited diagnostic exhaustiveness. This is particularly concerning in those regions that were predicted to have large numbers of weekly infections (Île-de-France, in which only one out of three symptomatic cases was detected by the end of June, and Grand Est, in which one out of five was detected). Almost all patients (92%) who were clinically diagnosed by sentinel general practitioners as suspected cases of COVID-19 were prescribed a test²⁰. However, only 31% of individuals with COVID-19-like symptoms consulted a doctor according to participatory surveillance data. Overall, these figures suggest that a large number of symptomatic cases of COVID-19 were not screened because they did not seek medical advice despite the recommendations. This was confirmed by serological studies. In France, only 48% of symptomatic participants with antibodies against SARS-CoV-2 reported consulting a general practitioner⁷; in Spain, between 16% and 20% of individuals with antibodies against SARS-CoV-2 reported a previous virological screening⁶. By combining estimates from virological and participatory surveillance data, we extrapolated an incidence rate from crowd-sourced data that is compatible with model projections, under the hypothesis that the large majority of suspected cases would get tested (>80%). This finding further supports testing of all suspected cases of COVID-19. Large-scale communication campaigns should reinforce recommendations to raise awareness in the population and strongly encourage healthcare-seeking behaviour especially in patients with mild symptoms. At the same time, investigations to identify reasons for not consulting a doctor could be quickly performed through the participatory surveillance system.

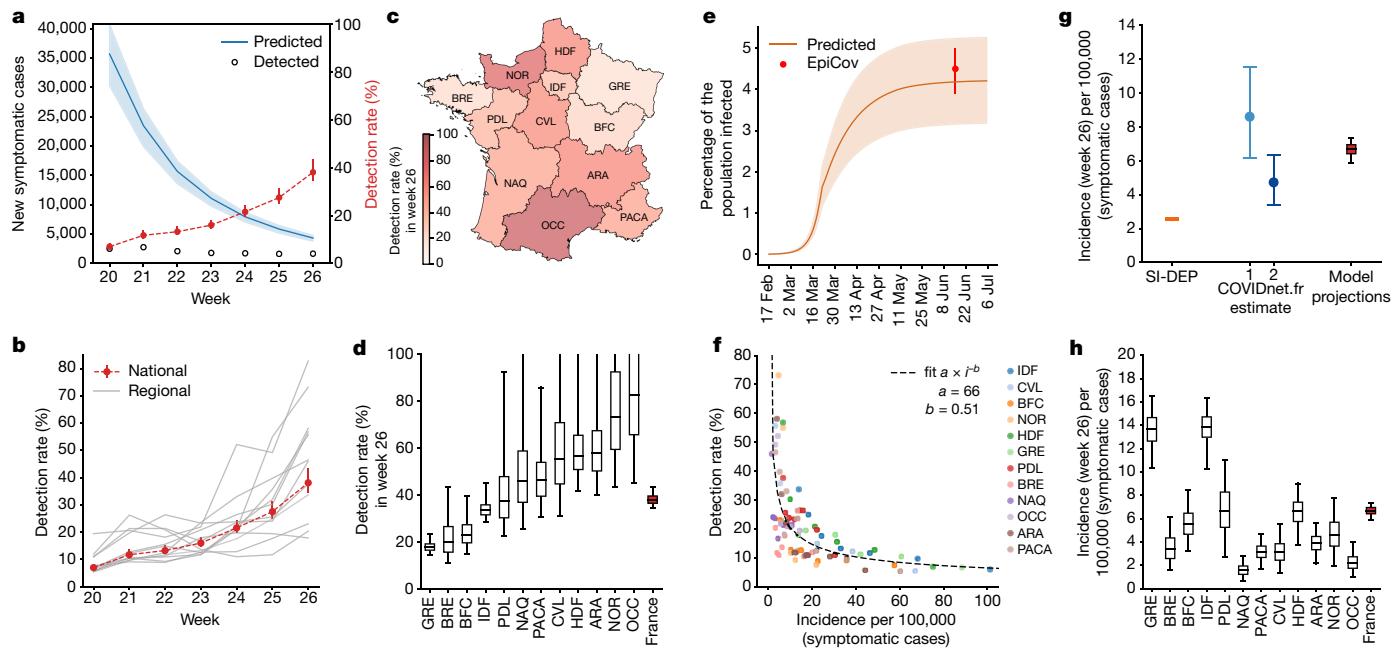


Fig. 3 | Detection rate and incidence. **a**, Projected number of new symptomatic cases over time (median and 95% confidence interval) and estimated number of virologically confirmed symptomatic cases by week of onset (points) in mainland France (left y axis). The estimated detection rate of symptomatic cases (%) is also shown (red points, median and 95% confidence interval; right y axis). **b**, Estimated detection rate of symptomatic cases (%) and 95% confidence intervals over time for mainland France (red dots and bars), and for all regions (grey lines, only median values are shown for visualization). **c**, Map of the estimated detection rate (%) by region in week 26 (22–28 June 2020). **d**, Estimated detection rate per region compared to the national estimate. Regions are ranked by increasing median detection rate. Box plots represent the median (line in the middle of the box), interquartile range (box limits) and 2.5th and 97.5th percentiles (whiskers). **e**, Predicted percentage of the population infected (median and 95% confidence interval) compared with

estimates from the serological study EpiCov²⁶ performed on a representative sample of the population in mainland France. **f**, Estimated detection rate of symptomatic cases (%) by region and by week compared with the projected incidence by region and by week. The curve shows the result of a least-square fit to the data with a power-law function, $\pi = a \times i^{-b}$, where π is the detection rate (expressed as a percentage), i is the weekly incidence (cases per 100,000), $a = 66$ (95% confidence interval, 52–85) and $b = 0.51$ (0.41–0.60). **g**, Estimated incidence of symptomatic cases and 95% confidence intervals in mainland France in week 26 from different sources: virological surveillance data (SI-DEP), participatory surveillance data (COVIDnet.fr, with two estimates) and model projections. **h**, Projected incidence per region compared to the national estimate. Regions are ranked as in **d**. Box plots are as defined in **d**. In all panels, medians and 95% confidence intervals for model projections were obtained from $n = 500$ independent stochastic runs.

Red tape might have contributed to low testing rates. Prescription of a test was deemed compulsory in the new testing policy to prevent misuse of diagnostic resources⁸; however, this involved consultation, prescription and a laboratory appointment, which may have discouraged mildly affected individuals who do not require medical assistance. To facilitate access, testing should not require a prescription, as later established by authorities²⁸. Some local initiatives emerged over summer that increased the number of drive-through testing facilities, promoted massive screening in certain areas and offered mobile testing facilities to increase proximity to the population²⁹. The use of antigen tests will further facilitate access. These initiatives are particularly relevant to counteract socioeconomic inequalities in access to care in populations that are vulnerable to COVID-19^{30,31}. However, such strategies should not hinder a testing protocol that targets suspected index cases. Our results show that high testing efforts, measured by low test positivity rates, are not associated with high rates of detection. This was also observed in the UK during the first wave, when detection remained low despite large numbers of tests and a low positivity rate³². Without strong case-based surveillance, the risk is to disperse resources towards random individuals without symptoms who are unlikely to be positive. This could saturate the test–trace–isolate system, as observed during summer³³, without slowing down the circulation of SARS-CoV-2 that is required to safeguard the hospital system.

Given presymptomatic transmission, notification of contacts should be almost immediate to enable the effective interruption of transmission chains²². For testing to be an actionable tool to control the

transmission of SARS-CoV-2, delays should be suppressed and screening rates greatly increased but better targeted. Over May–June, the average weekly number of tests was 250,000—remaining well below the objective that was originally set by authorities (700,000 tests). The number of tests increased over summer, but proportionally to the increased circulation of the virus. The capacity of detection of the test–trace–isolate system scaled as the inverse of the square root of the incidence, already deteriorating rapidly at low incidence levels. More aggressive testing that targets suspected index cases should be performed at low viral circulation to avoid case resurgence. The system was predicted to be able to detect more than two out of three cases (rate >66%) only if the incidence was lower than one symptomatic case per 100,000, a figure that is 50 times smaller than estimated at the exit from lockdown. As detection of at least 50% of cases is needed to control the pandemic while avoiding strict social distancing², these results indicate that the system was insufficient to perform comprehensive case-based surveillance, as has been recommended when aiming to phase out restrictions⁵. Current restrictions applied in Europe to curb the second wave offer a second opportunity to improve testing policies and support the lifting of these measures in the upcoming weeks. Failing to do so may lead to a rapid and uncontrolled increase in the number of cases of COVID-19^{2,34}. Such risk is even stronger in the winter season and with the existing fatigue with regard to adhesion to the restrictions¹⁸.

Models were region-based and did not consider a possible coupling between regional epidemics caused by mobility. This choice

was supported by stringent movement restrictions during lockdown³⁰, and by the limited mobility increase in May–June, before important inter-regional displacements took place at the start of the summer holidays in July. Foreign importations of the virus³⁵ were neglected as France reopened its borders with EU member states on 15 June, and the Schengen area remained closed until July. The COVIDnet.fr cohort is not representative of the general population; however, a previous study on influenza-like illnesses has shown that the adjusted incidence was in good agreement with sentinel estimates⁴. Under-detection may also continue because of the imperfect characteristics of the reverse-transcription PCR tests used to identify infections of SARS-CoV-2³⁶. Some cases tested for SARS-CoV-2 could have had false-negative results, for example, because they were tested too early after the infection, thus further increasing the rate of underdetection. Previous work assessed the rate of underdetection in 210 countries³², but this study mainly focused on the early global dynamics. Our model gives up geographical extent for higher data quality in a specific country, providing a synthesis of data sources that characterizes human behaviour over time and space together with virological and participatory surveillance data to identify the weak links in the pandemic response.

Our findings identify critical needs for the improvement of the test-trace-isolate response system to control the COVID-19 pandemic. Substantially more aggressive and efficient testing that targets suspected cases of COVID-19 needs to be achieved to act as a way to control the COVID-19 pandemic. Associated communication and logistical needs should not be underestimated. These elements should be considered to enable the lifting of restrictive measures that are currently used to curb the second wave of COVID-19 in Europe.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-03095-6>.

- OECD. Testing for COVID-19: a way to lift confinement restrictions. <https://www.oecd.org/coronavirus/policy-responses/testing-for-covid-19-a-way-to-lift-confinement-restrictions-89756248/> (2020).
- Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Med.* **18**, 240 (2020).
- Data.gouv.fr. Données relatives aux résultats des tests virologiques COVID-19 (SI-DEP). <https://www.data.gouv.fr/en/datasets/donnees-relatives-aux-resultats-des-tests-virologiques-covid-19/> (2020).
- Guerrisi, C. et al. The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: the case of France. *Euro Surveill.* **23**, 1700337 (2018).
- WHO. Considerations for implementing and adjusting public health and social measures in the context of COVID-19. <https://www.who.int/publications/item/considerations-in-adjusting-public-health-and-social-measures-in-the-context-of-covid-19-interim-guidance> (2020).
- Pollán, M. et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* **396**, 535–544 (2020).
- Carrat, F. et al. Seroprevalence of SARS-CoV-2 among adults in three regions of France following the lockdown and associated risk factors: a multicohort study. Preprint at <https://doi.org/10.1101/2020.09.16.20195693> (2020).
- Ministère des Solidarités et de la Santé. Prise en charge par les médecins de ville des patients atteints de COVID-19 en phase de déconfinement. <https://solidarites-sante.gouv.fr/IMG/pdf/prise-en-charge-medecine-ville-covid-19.pdf> (2020).
- Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
- Pinotti, F. et al. Tracing and analysis of 288 early SARS-CoV-2 infections outside China: a modeling study. *PLoS Med.* **17**, e1003193 (2020).
- Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
- Lavezzo, E. et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429 (2020).
- Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151 (2004).
- Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. *Nat. Commun.* (in the press).
- Béraud, G. et al. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE* **10**, e0133203 (2015).
- Ministère de l'Education Nationale de la Jeunesse et des Sports. Déconfinement phase 2: point de situation au 28 mai. <https://www.education.gouv.fr/deconfinement-phase-2-point-de-situation-au-28-mai-303813> (2020).
- Google. Community Mobility Report - COVID-19. <https://www.google.com/covid19/mobility?hl=fr> (2020).
- Santé publique France. COVID-19: une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie. <https://www.santepubliquefrance.fr/etudes-et-enquetes/covid-19-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie> (2020).
- Haut Conseil de la Santé Publique. Signes cliniques d'orientation diagnostique du Covid-19. <https://www.hcsp.fr/explore.cgi/avisreportsdomaine?clfr=812> (2020).
- Santé publique France. COVID-19: point épidémiologique du 25 juin 2020. <https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-25-juin-2020> (2020).
- INSEE. Pyramide des âges 2020 - France et France métropolitaine. <https://www.insee.fr/en/statistiques/3696316> (2020).
- Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
- Riccardo, F. et al. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. *Eurosurveillance* **25**, 2000790 (2020).
- Salje, H. et al. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
- Vu, S. L. et al. Prevalence of SARS-CoV-2 antibodies in France: results from nationwide serological surveillance. Preprint at <https://doi.org/10.1101/2020.10.20.2021316> (2020).
- Ministère des Solidarités et de la Santé. En mai 2020, 4,5% de la population de France métropolitaine a développé des anticorps contre le SARS-CoV-2. Premiers résultats de l'enquête nationale EpiCov. <https://drees.solidarites-sante.gouv.fr/IMG/pdf/er1167.pdf> (2020).
- Stringhini, S. et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCOV-POP): a population-based study. *Lancet* **396**, 313–319 (2020).
- Ameli. Covid-19: les tests de dépistages sont possibles sans prescription médicale. <https://www.ameli.fr/assure/actualites/covid-19-les-tests-de-depistages-sont-possibles-sans-prescription-medicale> (2020).
- Covid-19: une campagne de tests à grande échelle débute dans 32 communes d'Île-de-France. *Le Parisien* (29 June 2020).
- Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit. Health* **2**, e638–e649 (2020).
- Garnier, R., Benetka, J. R., Kraemer, J. & Bansal, S. Socio-economic disparities in social distancing during the COVID-19 pandemic in the United States. Preprint at <https://doi.org/10.1101/2020.11.07.20201335> (2020).
- Russell, T. W. et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med.* **18**, 332 (2020).
- Explosion des délais d'attente des résultats aux tests Covid-19. *Le Figaro* (22 September 2020).
- Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
- Ruktanonchai, N. W. et al. Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* **369**, 1465–1470 (2020).
- Wernike, K. et al. Pitfalls in SARS-CoV-2 PCR diagnostics. *Transbound. Emerg. Dis.* <https://doi.org/10.1111/tbed.13684> (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Virological surveillance data

The centralized database SI-DEP for virological surveillance³ collects all tests performed in France for any reason. In the period under study, guidelines recommended individuals to consult a general practitioner at the first sign of COVID-19-like symptoms and to obtain a prescription for a virological test (a prescription was compulsory to access the test)⁸. In addition, routine testing was performed for patients admitted to the hospital with any diagnosis, healthcare personnel and individuals at other facilities (for example, in some care homes for older people or long-term healthcare facilities). Data include detailed information for the individuals tested in France, including (1) the date of the test; (2) the result of the test (positive or negative); (3) location (region); (4) the absence or presence of symptoms at the time of testing; (5) self-declared delay between onset and test in presence of symptoms. The delay is provided with the following breakdown: onset date occurring 0–1 day before date of test, 2–4 days before, 5–7 days before, 8–15 days before, or more than 15 days before. For some tests, information on points (4) and (5) is missing. The SI-DEP database provided complete information for 23,210 (66%) out of 35,264 laboratory-confirmed cases of COVID-19 tested between week 20 (11–17 May) and week 30 (19–26 July), with an increasing trend of complete information over time (from 49% in week 20 to 76% in week 30) (Extended Data Fig. 1). Among confirmed cases with complete information, 12,716 (55%) showed no symptoms at the time of testing (Extended Data Fig. 1). The study referred to the period from week 20 to week 26. Data up to week 30 were used to consolidate the data in the study period accounting for the delays.

Imputation of asymptomatic versus presymptomatic cases, onset date and missing information

Individuals who tested positive on a given date were recorded in the SI-DEP database as: cases with symptoms at the time of testing, with a self-declared delay from onset of symptoms; cases without symptoms at the time of testing; or cases with no information on presence or absence of symptoms at the time of testing. These three subsets of cases were analysed to account for the presence of presymptomatic individuals among those with no symptoms at the time of testing, imputation of missing data and the estimation of dates of infection or symptom onset.

For laboratory-confirmed cases of COVID-19 who had symptoms at the time of testing, we estimated their date of onset using the information on the date of test and the time interval of onset-to-test delay, which was self-declared by the patients (Fig. 1b). In the time period between weeks 20 and 30, 20% of cases had an onset-to-test delay of ≤ 1 day, 63% had a delay of ≤ 4 days, 83% had a delay of ≤ 7 days and 88% had a delay of ≤ 15 days (Extended Data Fig. 1). We fitted a Gamma distribution to the onset-to-test delay data with a maximum likelihood approach, using three different periods of time (May, June and July), to account for changes in the distribution of self-declared delays over time (that is, longer delays at the beginning of the study period, shorter delays at its end) (Extended Data Fig. 2). The estimated average delay in May, June and July was 12.9 (95% confidence interval, 7.0–16.1), 5.1 (3.7–6.3) and 2.7 (2.0–3.1) days, respectively. July data were used to consolidate data corresponding to infections with onset in June and tested with delay. Given a confirmed case with symptoms testing on a specific date, we assigned the onset date by sampling the onset-to-testing delay from the fitted distribution for that period, conditional to the fact that the delay lies in the corresponding time interval declared by the patient. We assumed that onset did not occur before the implementation of the national lockdown, on 17 March 2020 (week 12); we therefore truncated

the Gamma distribution accordingly, when assigning the date of onset for cases with onset-to-test delay > 15 days. The imputation procedure was carried out 100 times. Results were aggregated by week of onset.

For laboratory-confirmed cases of COVID-19 with no symptoms at the time of testing, we assumed that on average 40% of them were asymptomatic¹² (see the ‘Transmission model summary’ section), whereas the remaining 60% were presymptomatic who tested early thanks to contact tracing. Imputation was done by sampling from a binomial distribution and repeated 100 times. Data on contact tracing could not be used to inform data on infection or symptom onset, because of national regulatory framework on privacy preventing the matching of the two databases (virological tests and contact tracing). Given the low sensitivity of PCR tests in the early phase of the incubation period, we considered that imputed presymptomatic cases belonged to the prodromic phase. Onset date for presymptomatic cases was estimated by sampling from an exponential distribution with a mean of 1.5 days, corresponding to the duration of the prodromic phase in our model (Supplementary Table 1). For imputed asymptomatic, we assumed the same delay from infection to testing as in cases with symptoms. Given the structure of our compartmental model and to match the definition of the time used for symptomatic individuals (week of onset), we considered a delay in the detection of asymptomatic individuals starting from the end of the prodromic phase (corresponding to the symptom onset time for symptomatic infections) to the date of testing. We assigned this date by sampling the delay from the monthly gamma distribution. Imputation of the dates was repeated 100 times.

For laboratory-confirmed cases of COVID-19 with no information on symptoms at the time of testing, missing data were imputed by sampling from a multinomial distribution with probabilities equal to the rate of occurrence of the outcomes (asymptomatic, presymptomatic or symptomatic with five possible time intervals for the onset-to-test delay) reported for cases with complete information and assuming the imputation of cases without symptoms into asymptomatic and presymptomatic, as described above. Imputation was performed by region and by week and repeated 100 times. Presymptomatic and symptomatic individuals were aggregated together by onset date (Fig. 1a) to estimate the rate of detection of symptomatic cases.

Participatory surveillance data and analysis

COVIDnet.fr is a participatory online system for the surveillance of COVID-19, available at <https://www.covidnet.fr/>. It was adapted from GrippeNet.fr⁴ to respond to the COVID-19 health crisis in March 2020. GrippeNet.fr is a participatory system for the surveillance of influenza-like illnesses available in France since 2011 through a collaboration between Inserm, Sorbonne Université and Santé publique France, supplementing sentinel surveillance^{4,37}. The system is based on a dedicated website to conduct syndromic surveillance through self-reported symptoms volunteered by participants resident in France. Data are collected on a weekly basis; participants also provide detailed profile information at enrolment³⁸. In addition to tracking the incidence of influenza-like illnesses^{4,37}, GrippeNet.fr was used to estimate vaccine coverage in specific subgroups³⁹, individual perceptions towards vaccination⁴⁰ and healthcare-seeking behaviour⁴¹. It was also used to assess behaviours and perceptions related to diseases other than influenza⁴², including COVID-19⁴³.

Participants are on average older and include a larger proportion of women compared to the general population^{38,44}. The participating population is, however, representative in terms of health indicators such as diabetes and asthma conditions. Despite these discrepancies, trends of the estimated incidence of influenza-like illnesses from GrippeNet.fr reports compared well with those of the national sentinel system^{4,37}. All analyses were adjusted by age and sex of participants.

To monitor suspected cases of COVID-19 in the general population, we used the expanded case definition recommended by the High Council of Public Health for systematic testing and described in their

20 April 2020 notice¹⁹, which included either of the two following definitions: (1) (sudden onset of symptoms OR sudden onset of fever) AND (fever OR chills) AND (cough OR shortness of breath OR (chest pain AND age > 5 years old)) or (2) (sudden onset of symptoms) OR (sudden onset of fever AND fever); and one of the three following conditions: (i) (age > 5 years old) AND ((feeling tired or exhausted) OR (muscle/joint pain) OR (headache) OR (loss of smell WITHOUT runny or blocked nose) OR (loss of taste)); or (ii) ((age ≥ 80 years old) OR (age < 18 years old)) AND (diarrhoea); or (iii) (age < 3 months old) AND (fever WITHOUT other symptoms).

Two independent estimates obtained from COVIDnet.fr cohort data for the incidence of symptomatic cases in week 26 are shown in Fig. 3. These estimates were computed as follows. Estimate 1 = (COVIDnet.fr estimated incidence of suspected cases in week 26) × (test positivity rate from SI-DEP in week 26); estimate 2 = (COVIDnet.fr estimated incidence of suspected cases in week 26) × (estimated proportion screened and confirmed as a suspected case of COVID-19 by a physician, and prescribed a test; estimates from COVIDnet.fr) × (test positivity rate from SI-DEP in week 26). The two estimates were used to validate model projections and identify the specific surveillance mechanisms that needed improvement.

Ethics statement

GrippeNet.fr/COVIDnet.fr was reviewed and approved by the French Advisory Committee for research on information treatment in the health sector (that is, CCTIRS, authorization 11.565), and by the French National Commission on Informatics and Liberty (that is, CNIL, authorization DR-2012-024)—the authorities ruling on all matters related to ethics, data and privacy in the country. Informed consent was provided by each participant at enrolment, according to regulations.

Transmission models summary

We used a stochastic discrete age-stratified transmission model for each region based on demographic, contact¹⁵ and age profile data of French regions²¹. Models were region-specific to account for the geographically heterogeneous epidemic situation in the country and given the mobility restrictions limiting inter-regional movement fluxes. The study focused on mainland France where the epidemic situation was comparable across regions, and excluded Corsica, which reported very limited epidemic activity and overseas territories characterized by increasing transmission²⁰.

Four age classes were considered: [0–11], [11–19], [19–65] and 65+ years old, referred to as children, adolescents, adults and older individuals. Transmission dynamics follows a compartmental scheme specific to COVID-19, in which individuals were divided into susceptible, exposed, infectious and hospitalized (Supplementary Information and Supplementary Figs. 1, 2). We did not consider further progression from hospitalization (for example, admission to intensive care units, recovery or death²) as it was not needed for the objective of the study. The infectious phase is divided into two steps: a prodromic phase (I_p) and a phase during which individuals may remain either asymptomatic (I_a , with probability¹² $p_a = 40\%$) or develop symptoms. In the latter case, we distinguished between different degrees of severity of symptoms^{9,11,23,24}, ranging from paucisymptomatic (I_{ps}), to infectious individuals with mild (I_{ms}) or severe (I_{ss}) symptoms. Prodromic, asymptomatic and paucisymptomatic individuals have a reduced transmissibility $r_\beta = 0.55$, as estimated previously¹¹, and in agreement with evidence from the field^{45–47}. A reduced susceptibility was considered for children and adolescents, along with a reduced relative transmissibility of children, following available evidence from household studies, contact-tracing analyses, serological investigations and modelling works^{48–53}. A sensitivity analysis was performed on the relative susceptibility and transmissibility of children, and on the proportion of asymptomatic infections (Supplementary Figs. 10–13). Full details are reported in the Supplementary Information.

The study was not extended to the summer months, because of (1) the challenge of mechanistically parameterizing the contact matrices during summer; (2) the increase of movement fluxes across regions weakening our assumption of region-specific models; and (3) the interruption of COVIDnet.fr surveillance during the summer break, which prevented the identification of the key factors behind case underascertainment.

Contact matrices

Age-stratified transmission uses a social contact matrix that reports the average contact rates between different age classes in France¹⁵. This refers to the baseline condition, that is, before lockdown. The contact matrix includes the following layers: contacts at home, school, workplace, transport, leisure activities and other activities, and discriminates between physical and non-physical contacts. To account for the change of contact patterns over time, contact matrices are mechanistically parameterized, by region and over time, with different data sources informing on the percentage of students going to school¹⁶, the percentage of workers going to the workplace¹⁷, the compliance to preventive measures¹⁸, with a higher compliance registered in older individuals¹⁸. Information on the progressive reopening of activities indicates that leisure and other activities were only partially open in the study period. Data, however, are not fine-grained enough to parameterize our model, so we assume a 50% opening of these activities and explore variations in the sensitivity analysis.

School attendance. School reopening was parameterized by considering the percentage of reported attendance at school (pre-school and primary school; middle and high school) provided by the Ministry of Education¹⁶ (Supplementary Fig. 3). The number of contacts in the school matrix was modified to account for the attendance of students in each school level provided by data. That is, attendance of 14.5%, referring, for example, to the attendance registered in Île-de-France in pre-schools and primary schools, corresponds to a reduction of 85.5% in the number of contacts established at school by students belonging to that school level. Contacts for different modes of transport were modified accordingly.

Presence at work. To account for the percentage of individuals at work, given recommendations on remote working and activities that were not yet reopened, we used the estimated variation of presence at workplaces based on mobile phone location data provided by Google Mobility Trends¹⁷. Contacts at work and for different modes of transport were therefore modified according to this percentage, as described for contacts at school. Household contacts were increased proportionally to each adult staying at home based on statistics comparing weekend versus weekday contacts¹⁵ and the proportion of adults working during the weekend⁵⁴, as done previously².

Adoption of physical distancing. Our previous work showed that physical contacts during lockdown were fully avoided², in agreement with data collected afterwards¹⁸. To account for individual adoption of preventive behaviour after lockdown, we used the percentage of population avoiding physical contacts estimated from a large-scale survey conducted by Santé publique France (CoviPrev¹⁸). Data were fitted with a linear regression (Fig. 1) to provide the weekly percentage of individuals avoiding physical contacts. We therefore modified our contact matrices over time, removing the percentage of physical contacts corresponding to the survey estimates for that week.

Increased risk aversion of older individuals. Data from the Santé publique France survey CoviPrev¹⁸ also show that older individuals protected themselves further relative to other age classes. On average, they respected physical distancing 28% more than the other age classes (Supplementary Fig. 4). For this reason, we considered a further

Article

reduction of 30% in contacts for older individuals in the exit phase, informed by survey data.

Inference framework

The parameters of the transmission models to be estimated are specific to each pandemic phase.

Before lockdown, $\{\beta, t_0\}$, where β is the transmission rate per contact and t_0 is the date of the start of the simulation, seeded with 10 infectious individuals.

During lockdown, $\{\alpha_{LD}, t_{LD}\}$, where α_{LD} is the scaling factor of the transmission rate per contact and t_{LD} is the date when lockdown effects on hospitalization data became visible.

After lockdown, $\{\alpha_{exit}, \pi_a(w), \pi_s(w)\}$, where α_{exit} is the scaling factor of the transmission rate per contact, and $\pi_a(w)$ and $\pi_s(w)$ are the proportion of asymptomatic and symptomatic cases tested in week w of the exit phase, respectively. Detected cases in the simulations had their contacts reduced by 90% to mimic isolation, as done in previous studies^{2,14}.

We used simulations of the stochastic model to predict values for all quantities of interest (500 simulations each time). We fitted the model to the daily count of hospitalizations $H_{obs}(d)$ on day d throughout the period and the number of people testing positive by week of onset, split according to disease status (symptomatic or asymptomatic), denoted $Test_{s,obs}(w)$ and $Test_{a,obs}(w)$ in week w of the exit phase. We used hospital admission data up to week 27 (29 June–5 July) to account for the average delay from infection to hospitalization. Data in week 27 were consolidated by waiting for one additional week to account for updates and missing data (week 28, 6–12 July 2020).

We assumed a Poisson distribution for hospitalizations and a binomial distribution for the number of people getting the test, therefore the likelihood function is

$$L(\text{Data}|\Theta) = \prod_{d=t_0}^{t_n} P_{\text{Poisson}}(H_{obs}(d); H_{pred}(d), \beta, t_0, \alpha_{LD}, t_{LD}, \alpha_{exit}, \pi_a(w_d), \pi_s(w_d)) \\ \times \prod_{w \in \text{exit}} P_{\text{Binomial}}(Test_{s,obs}(w); i_{s,pred}(w), \pi_s(w)) \\ \times \prod_{w \in \text{exit}} P_{\text{Binomial}}(Test_{a,obs}(w); i_{a,pred}(w), \pi_a(w))$$

where $\Theta = \{\beta, t_0, \alpha_{LD}, t_{LD}, \alpha_{exit}, \{\pi_a(w)\}, \{\pi_s(w)\}\}$ indicates the set of parameters to be estimated, $H_{pred}(d)$ is the model-predicted number of hospital admissions on day d , $i_{s,pred}(w)$ and $i_{a,pred}(w)$ are the model-predicted weekly incidences of symptomatic and asymptomatic cases, respectively, in week w of the exit phase, P_{Poisson} is the probability mass function of a Poisson distribution, P_{Binomial} for a binomial distribution, $[t_0, t_n]$ is the time window considered for the fit, and w is the week in the exit phase (weeks 20–26).

We reduced the required computations with an optimization procedure in two steps, first maximizing the likelihood function in the pre-lockdown and lockdown phase to estimate the first four parameters, and then maximizing the likelihood in the exit phase by fixing the first four parameters that describe the epidemic trajectory before the exit phase to their maximum likelihood estimators (MLEs). This second step was further simplified through an iterative procedure, and we show through simulations that the simplified optimization procedure is consistent and well-defined. The parameter space was explored using NOMAD software⁵⁵. Fisher's information matrix was estimated at the MLE value to obtain the corresponding confidence intervals. Simulations were then parameterized with 500 parameter sets obtained from the joint distribution of transmission parameters at MLE (one stochastic simulation for each parameter set). A Bayesian estimate of the posterior parameter distribution using Markov chain Monte Carlo (MCMC) would also have been an alternative to maximum likelihood and confidence interval estimation. In this case, however, MCMC would have considerably slowed down parameter exploration, with negligible added value to the fitting procedure.

We repeated model fitting starting from several starting points and using different random number streams. Values of fitted parameters and full details on the different steps and the tests performed are reported in the Supplementary Information (Supplementary Figs. 6, 7 and Supplementary Table 3).

Simulation details

Simulations are initialized with 10 infected adults in the I_p compartment at time t_0 . We obtained 500 parameter sets from the joint distribution of transmission parameters at MLE and ran one stochastic simulation for each parameter set. Therefore, errors in the detection rates computed in the output account for the variability of the estimate of the parameters, in addition to the stochastic fluctuations of the model. We find that the errors in the estimation of the detection rates obtained including the variability of the parameters are slightly larger than the ones obtained with only stochastic fluctuation, suggesting that the stochasticity of the model is the main source of error in the estimation of the detection rate.

Model selection analysis

To assess the role of the mechanistic modification of the contact matrix informed by the different data sources in the exit phase, we compared our model to a simplified version assuming that contact patterns in the exit phase do not change from pre-epidemic conditions, and that all changes in the epidemic trajectory are explained exclusively by the transmissibility per contact. This is equivalent to normalizing the contact matrix to its largest eigenvalue and estimating the reproductive ratio over time. We compared the two models with the Akaike information criterion and found that accounting for changes in contacts better describes the epidemic trajectory (Supplementary Table 2 and Supplementary Fig. 5).

Comparison with serological estimates

We compared model projections with serological estimates from three independent studies^{7,25,26} (Fig. 3e and Extended Data Fig. 6).

Estimates by Carrat et al.⁷ used ELISA-S tests and ELISA-NP tests. The sample was not representative of the population, and estimates were weighted to account for this bias. In the comparisons, we used the results from a multiple imputation method performed by the authors and estimating a participant's positivity with a likelihood of positivity based on observed test results and covariates (see ref.⁷ for more details).

Estimates by Santé publique France²⁵ are based on at least one positive result in one of the following three tests: ELISA-S, ELISA-NP and a pseudo-neutralization test that detects the presence of pseudo-neutralizing antibodies, representative of the presence of neutralizing antibodies as conferring protection against infection. Analyses were performed on residual sera obtained from clinical laboratories, and estimates were weighted to account for the lack of representativeness.

Estimates by EpiCoV²⁶ (Enquête Épidémiologie et Conditions de vie liées à la Covid-19) used ELISA-S tests and further validated these with a seroneutralizing antibody test at higher specificity (see ref.²⁶ for more details). This was the only seroprevalence survey that was conducted in a representative sample of the population. For this reason, we used it as the reference study.

For all studies, we report in Fig. 3e and Extended Data Fig. 6 the estimates 14 days before the last blood collection to account for the time needed to mount a detectable presence of antibodies. For the EpiCoV survey, we used the last date at which samples were sent back to the laboratory.

Modelling results are in good agreement with the serological estimates at the national level (Fig. 3e) and in the large majority of the regions (Extended Data Fig. 6). Projections tend to be systematically smaller than serological estimates in two regions that were weakly affected by the epidemic (Pays de la Loire and Brittany), although they remained compatible with observations.

Overall differences may be due to the limitations of the methods involved. First, the type of tests, the specificity levels, the samples of the population tested, and the weighting and imputation approaches considered in each serological study could lead to differences across the three investigations. We note, for example, that larger discrepancies are observed between EpiCov and Santé publique France results in those regions that experienced smaller epidemics. We used EpiCov as the reference study as it was the only one study that was conducted on a representative sample of the population. Second, there are limitations to the dataset of hospital admissions used to calibrate the models: the database infrastructure for data collection became operational in mid-March and was filled in retrospectively. Notification biases would inevitably alter the inference of parameters in the pre-lockdown phase. This may have differed region by region; however, we have no way to control for this potential bias; possible errors would have affected regions with small-size epidemics more than others. In support of this hypothesis, we note that a similar but independent mathematical model fitted to regional hospitalization data²⁴ in the first wave predicted small epidemics in Pays de la Loire and Brittany, similarly to our model.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data used for model parameterization (contact matrices¹⁵, school attendance¹⁶, presence at workplaces¹⁷ and avoidance of physical contacts¹⁸) are available at the cited sources. Hospitalization data used for model calibration are made available with the code. Virological data³ are available at the cited source; data on onset-to-test delay are made available with the code. COVIDnet.fr individual data cannot be shared owing to restrictions imposed by the French national data protection authorities. Requests for custom access to aggregated and post-processed data can be made to the GripeNet.fr/COVIDnet.fr Scientific Committee (<https://covidnet.fr/fr/covidnet/confidentialite-et-securite-des-donnees/>) through the submission of a scientific proposal describing the aims, methods, data format requested and team proposing the project. Decisions by the GripeNet.fr/COVIDnet.fr Scientific Committee will be based on pertinence of the scientific proposal to the objectives of COVIDnet.fr and on the constraints on privacy and data treatment imposed by national regulatory authorities. Adjusted data on COVIDnet.fr participants, incidence and healthcare-seeking behaviour are made available with the code. Source data are provided with this paper.

Code availability

Code for the transmission models used for the analyses is available at https://github.com/EPIcx-lab/COVID-19/tree/master/Underdetection_France.

37. Guerrisi, C. et al. Participatory syndromic surveillance of influenza in Europe. *J. Infect. Dis.* **214**, S386–S392 (2016).

38. Debin, M. et al. Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system for influenza in France. *PLoS ONE* **8**, e73675 (2013).
39. Loubet, P. et al. Influenza during pregnancy: incidence, vaccination coverage and attitudes toward vaccination in the French web-based cohort G-GrippeNet. *Vaccine* **34**, 2390–2396 (2016).
40. Boiron, K. et al. Opinion about seasonal influenza vaccination among the general population 3 years after the A(H1N1)pdm2009 influenza pandemic. *Vaccine* **33**, 6849–6854 (2015).
41. Ariza, M. et al. Healthcare-seeking behaviour in case of influenza-like illness in the French general population and factors associated with a GP consultation: an observational prospective study. *BJGP Open* **1**, bjgopen17X101253 (2018).
42. Mathieu, P. et al. Population perception of mandatory childhood vaccination programme before its implementation, France, 2017. *Euro Surveill.* **24**, 1900053 (2019).
43. Raudé, J. et al. Are people excessively pessimistic about the risk of coronavirus infection? Preprint at <https://doi.org/10.31234/osf.io/364qj> (2020).
44. Cantarelli, P. et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. *BMC Public Health* **14**, 984 (2014).
45. Buitrago-Garcia, D. et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med.* **17**, e1003346 (2020).
46. Qiu, X. et al. Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission - a living systematic review. Preprint at <https://doi.org/10.1101/2020.09.01.20135194> (2020).
47. Luo, L. et al. Contact settings and risk for transmission in 3410 close contacts of patients with COVID-19 in Guangzhou, China. *Ann. Intern. Med.* **173**, 879–887 (2020).
48. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
49. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).
50. Zimmermann, P. & Curtis, N. Coronavirus infections in children including COVID-19: an overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children. *Pediatr. Infect. Dis. J.* **39**, 355–368 (2020).
51. Jiehao, C. et al. A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features. *Clin. Infect. Dis.* **71**, 1547–1551 (2020).
52. Fontanet, A. et al. SARS-CoV-2 infection in primary schools in northern France: a retrospective cohort study in an area of high transmission. Preprint at <https://doi.org/10.1101/2020.06.25.20140178> (2020).
53. Fontanet, A. et al. Cluster of COVID-19 in northern France: a retrospective closed cohort study. Preprint at <https://doi.org/10.1101/2020.04.18.20071134> (2020).
54. INSEE. Durée et organisation du temps de travail. <https://www.insee.fr/fr/statistiques/3303396?sommaire=3353488> (2020).
55. Le Digabel, S. Algorithm 909: NOMAD: nonlinear optimization with the MADS algorithm. *ACM Trans. Math. Softw.* **37**, 44 (2011).

Acknowledgements This study was partially supported by ANR projects DATAREDUX (ANR-19-CE46-0008-03) and EVALCOVID-19 (ANR-20-706 COVI-0007); EU H2020 grants MOOD (H2020-874850) and RECOVER (H2020-101003589); and REACTing COVID-19 modelling and surveillance grants. We thank P. Crepey, C. Pelat, E. Chatignoux, J. Paireau and D. Levy-Bruhl for discussions, and all participants of the COVIDnet.fr system.

Author contributions V.C. conceived and designed the study. C.T., M.D., C.G., C.K.-K., C.S., T.H., T.B., S.B.-S. and V.C. developed and implemented the protocol of the participatory platform for COVID-19 surveillance. G.P., L.D.D., C.E.S., C.T., J.F., S.V., C.C. and S.B.-S. collected the data. G.P., L.D.D., C.E.S. and C.T. analysed the data. G.P., L.D.D., C.E.S., P.-Y.B. and V.C. developed the inference framework. G.P. and L.D.D. developed the code. G.P., L.D.D., C.E.S. and P.-Y.B. performed the numerical simulations, and analysed the results. G.P., L.D.D., C.E.S., E.V., C.T., M.D., C.G., C.K.-K., C.S., T.H., T.B., P.-Y.B., J.F., S.V., C.C., S.B.-S. and V.C. interpreted the results. V.C. drafted the Article. All authors contributed to and approved the final version of the Article.

Competing interests The authors declare no competing interests.

Additional information

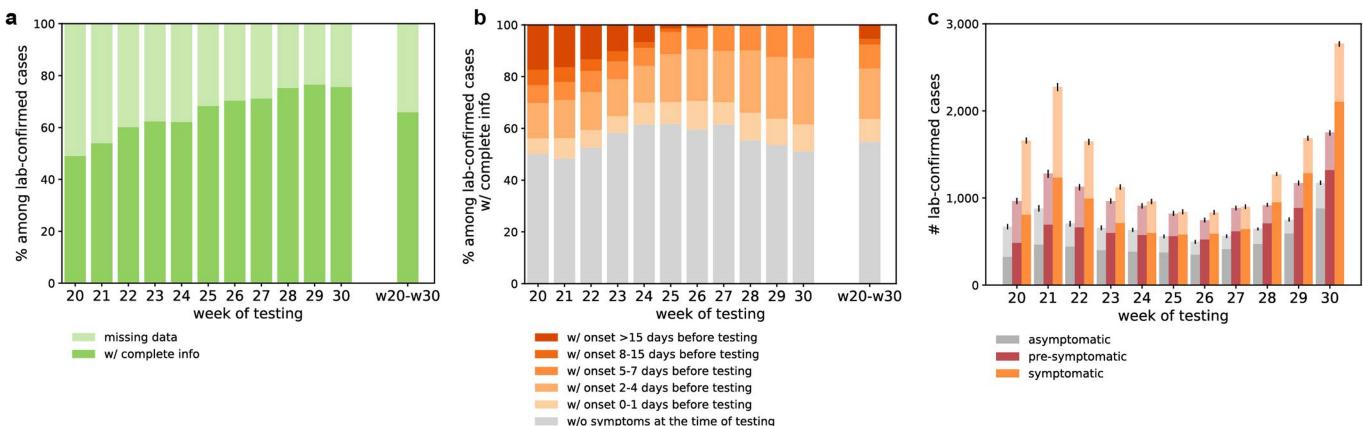
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-020-03095-6>.

Correspondence and requests for materials should be addressed to V.C.

Peer review information *Nature* thanks Ilaria Dorigatti, Niel Hens and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

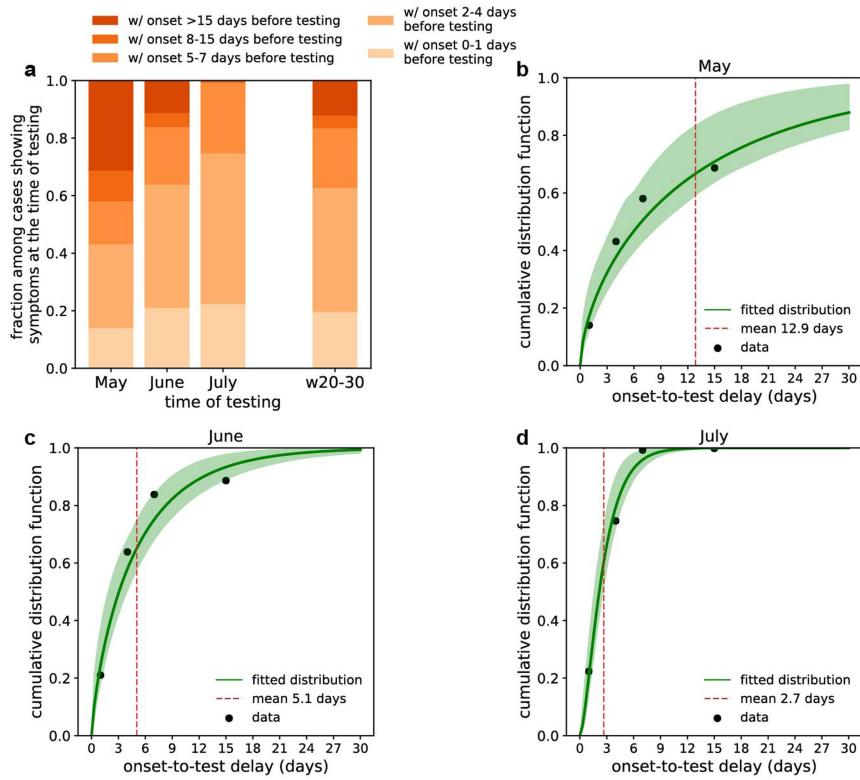
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Article



Extended Data Fig. 1 | Information provided by SI-DEP database and imputation of missing data. **a**, Breakdown of laboratory-confirmed cases of COVID-19 ($n=35,264$) according to information available. Missing data refer to cases for which information regarding absence or presence of symptoms at the time of testing and self-declared onset-to-test delay were not provided. **b**, Breakdown of laboratory-confirmed cases of COVID-19 with complete information ($n=23,210$) according to absence or presence of symptoms on the

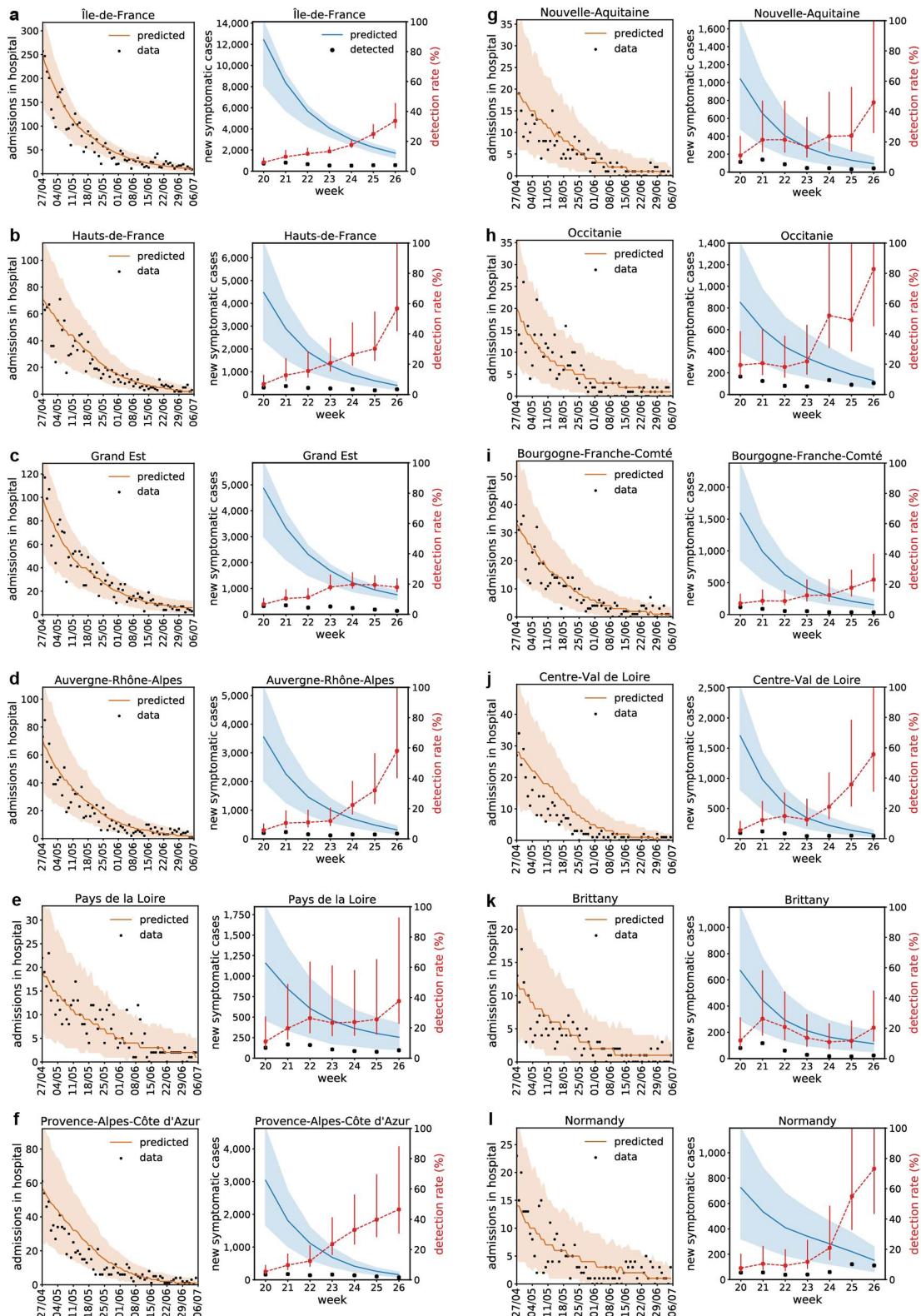
date of testing and onset-to-test delay. **c**, Estimated number of asymptomatic, presymptomatic and symptomatic confirmed cases by week of testing, after imputation of missing data about presence or absence of symptoms. Darker and lighter colours indicate cases with complete information and missing data, respectively. The 95% confidence intervals (black bars) are obtained by applying the imputation procedure 100 times.



Extended Data Fig. 2 | Estimation of onset-to-test delay distribution.

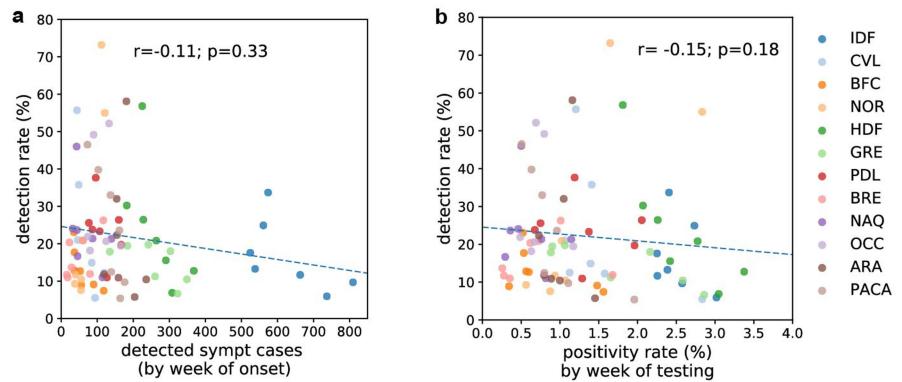
a, Data used for the fit. Breakdown of laboratory-confirmed cases of COVID-19 with symptoms ($n=10,494$) according to month of testing and the declared time interval of onset-to-testing delay. **b**, Gamma distribution fitted through maximum likelihood to data referring to May (shape parameter 0.61, 0.42–0.80 (median estimate and 95% Wald confidence interval); scale parameter 21.08,

9.41–32.78). **c**, Gamma distribution fitted through maximum likelihood to data referring to June (shape parameter 0.75, 0.51–0.99; scale parameter 6.77, 3.77–9.75). **d**, Gamma distribution fitted through maximum likelihood to data referring to July (shape parameter 1.69, 1.12–2.28; scale parameter 1.57, 0.96–2.17). Shaded areas represent 95% confidence interval of the cumulative distribution function.



Extended Data Fig. 3 | Hospital admissions and number of new symptomatic cases. **a–l**, For each region, the pair of panels shows the hospital admissions over time (data (points) and simulations (median and 95% confidence interval)) (left) and the projected number of new symptomatic cases over time (median and 95% confidence interval) and the estimated number of virologically confirmed symptomatic cases by week of onset (points) and detection rate (red points; right yaxis) (right). Medians and 95%

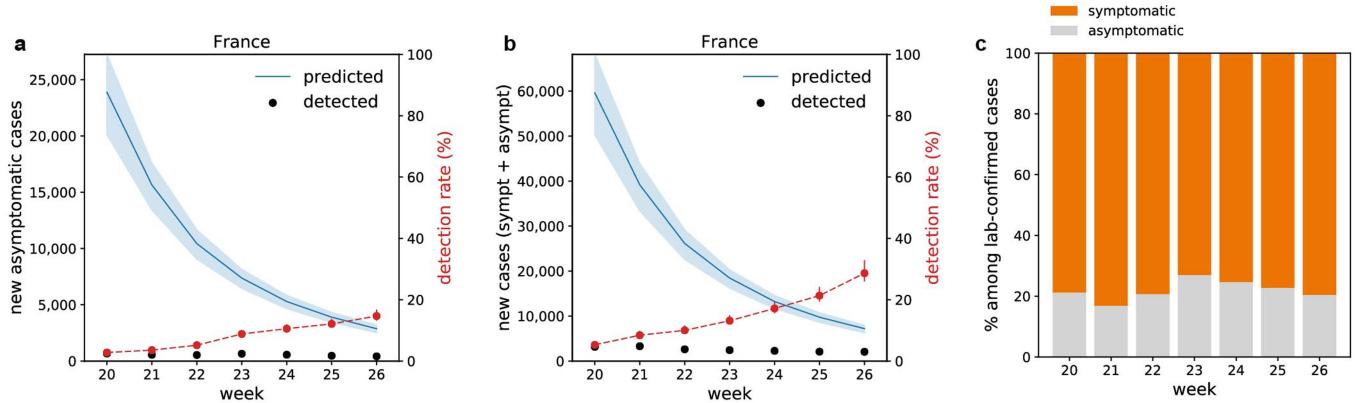
confidence intervals are obtained from $n=500$ independent stochastic runs. Plots are reported for all 12 regions of mainland France: Île-de-France (**a**), Hauts-de-France (**b**), Grand Est (**c**), Auvergne-Rhône-Alpes (**d**), Pays de la Loire (**e**), Provence-Alpes-Côte d'Azur (**f**), Nouvelle Aquitaine (**g**), Occitanie (**h**), Bourgogne-Franche-Comté (**i**), Centre-Val de Loire (**j**), Brittany (**k**) and Normandy (**l**). Hospital admission data up to week 27 (consolidated in week 28) were used to calibrate the models.



Extended Data Fig. 4 | Detection rate versus indicators linked to testing policy and effectiveness. **a**, Detection rate versus the number of detected symptomatic cases, by week and by region. **b**, Detection rate versus

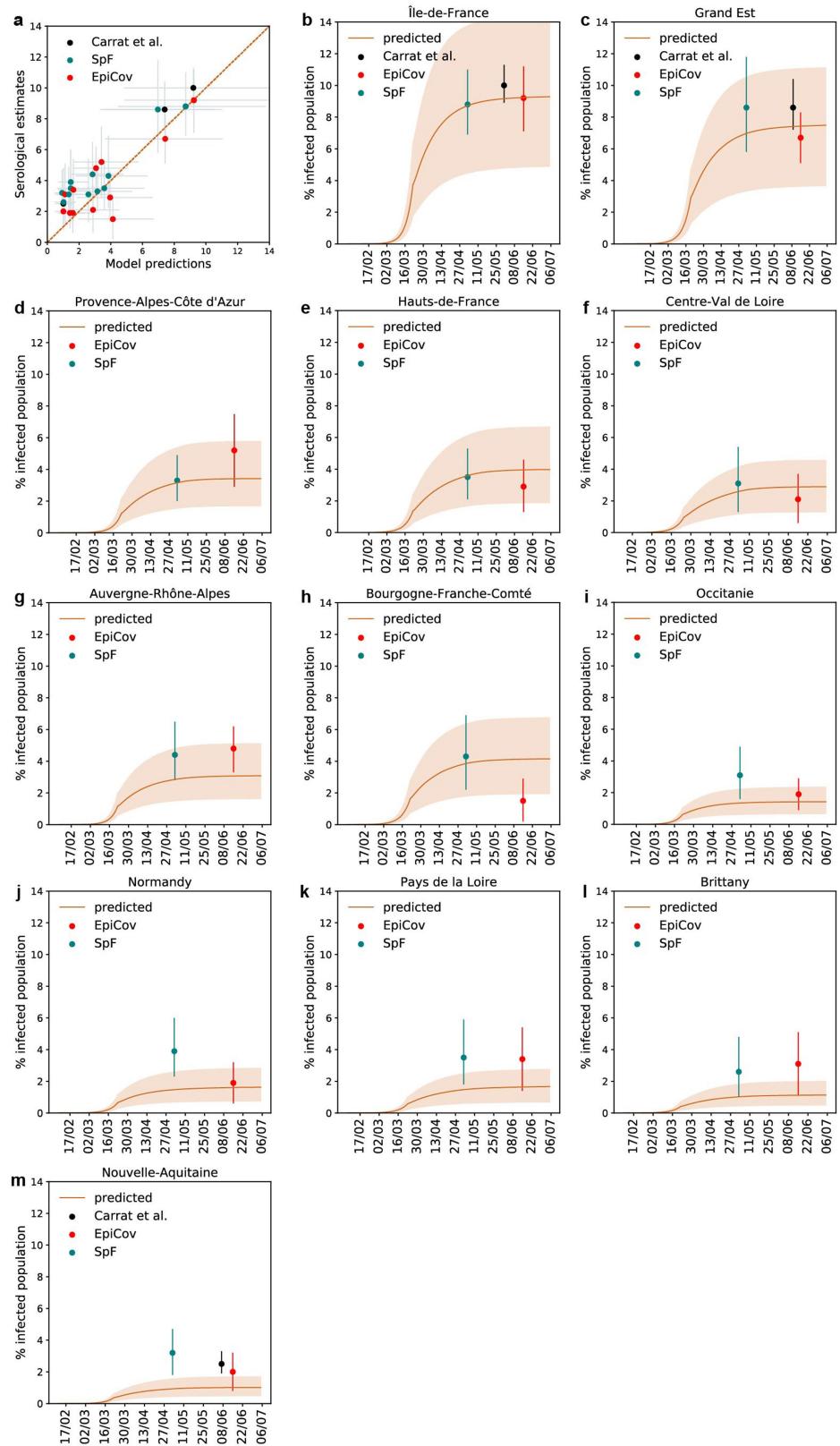
positivity rate by week of testing. Results of a Spearman correlation test are provided (r , Spearman correlation coefficient; P , P -value).

Article



Extended Data Fig. 5 | Estimates of the detection rate for symptomatic and asymptomatic cases. **a**, Projected number of new asymptomatic cases over time (median and 95% confidence interval), estimated number of virologically confirmed symptomatic cases by week of onset (black dots) and detection rate by week (red dots, median and 95% confidence interval) in mainland France.

Medians and 95% confidence intervals are obtained from $n=500$ independent stochastic runs. **b**, As in **a**, but considering both symptomatic and asymptomatic cases. **c**, Percentage of estimated asymptomatic and symptomatic individuals among the estimated number of virologically confirmed cases ($n=17,939$).



Extended Data Fig. 6 | Model predictions versus serological estimates.

a, Serological estimates per region from three seroprevalence studies versus corresponding model projections. Error bars correspond to 95% confidence intervals. **b–m**, For each region, the panel shows the predicted percentage of infected population over time (red curves and shaded areas for median and 95% confidence interval) and serological estimates of EpiCov²⁶ (red dots), Santé publique France (SpF)²⁵ (green dots) and ref.⁷ (black dots in **b**, **c**, **m**). Medians

and 95% confidence intervals for model projections are obtained from $n=500$ independent stochastic runs. Plots are reported for all 12 regions of mainland France: Île-de-France (**b**), Grand Est (**c**), Provence-Alpes-Côte d'Azur (**d**), Hauts-de-France (**e**), Centre-Val de Loire (**f**), Auvergne-Rhône-Alpes (**g**), Bourgogne-Franche-Comté (**h**), Occitanie (**i**), Normandy (**j**), Pays de la Loire (**k**), Brittany (**l**) and Nouvelle Aquitaine (**m**).

Article

Extended Data Table 1 | Population, confirmed and projected symptomatic cases, and estimated detection rate

Region	Pop (millions)	# lab-confirmed symptomatic cases by week of onset	#projected symptomatic cases by week of onset (median and 95% CI)	Estimated detection rate (%) of symptomatic cases (median and 95% CI)
		w20-w26	w20-w26	w20-w26
Île-de-France (IDF)	12.3	4,405	37,476 [27,692 - 40,786]	12 [11 - 16]
Grand Est (GRE)	5.5	1,797	15,182 [9,887 - 17,377]	12 [10 - 18]
Hauts de France (HDF)	6.0	1,868	12,404 [6,749 - 17,463]	15 [11 - 28]
Auvergne-Rhône-Alpes (ARA)	8.0	1,204	9,749 [5,592 - 14,134]	12 [9 - 22]
Occitanie (OCC)	5.9	774	2,793 [1,381 - 4,554]	28 [17 - 56]
Provence-Alpes-Côte d'Azur (PACA)	5.1	952	7,470 [4,250 - 11,041]	13 [9 - 22]
Pays de la Loire (PDL)	3.8	821	4,006 [1,654 - 6,393]	20 [13 - 50]
Bourgogne-Franche-Comté (BFC)	2.8	426	4,264 [2,392 - 6,179]	10 [7 - 18]
Nouvelle Aquitaine (NAQ)	6.0	505	2,789 [1,252 - 4,606]	18 [11 - 40]
Centre-Val de Loire (CVL)	2.6	484	4,066 [1,985 - 5,924]	12 [8 - 24]
Brittany (BRE)	3.3	346	2,037 [954 - 3,478]	17 [10 - 36]
Normandy (NOR)	3.3	480	2,682 [1,245 - 4,340]	18 [11 - 39]
France*	64.6	14,061	103,907 [90,216-116,377]	14 [12 - 16]

Results correspond to the full study period, from 11 May to 28 June 2020.

* Data are for mainland France; Corsica and overseas territories were excluded.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Computer code was not used to collect data.

Data analysis Analyses were carried out in Python 3.7.1. NOMAD software 3.9.1 (<https://www.gerad.ca/nomad/>) was used for optimizing search in the parameter space during parameter estimation. Code is available at link https://github.com/EPIcx-lab/COVID-19/tree/master/Underdetection_France. The fitting procedure was also independently coded in R 4.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data used for models parameterization (contact matrices[14], school attendance[15], presence at workplaces[16], avoidance of physical contacts[17]) are available at the cited sources. Hospitalization data used for model calibration are made available with the code at https://github.com/EPIcx-lab/COVID-19/tree/master/Underdetection_France. Virological data[3] are available at the cited source; data on onset-to-test delay are made available with the code. COVIDnet.fr individual data cannot be shared due to restrictions imposed by the French national data protection authorities. Requests for custom access to aggregated and post-processed data can be made to GrippeNet.fr/COVIDnet.fr Scientific Committee (<https://covidnet.fr/fr/covidnet/confidentialite-et-securite-des-donnees/>) through the submission of a scientific proposal describing the aims, methods, data format requested, and team proposing the project. Decisions by the COVIDnet.fr Scientific

Committee will be based on pertinence of the scientific proposal with the objectives of COVIDnet.fr and on the constraints on privacy and data treatment imposed by national regulatory authorities. Adjusted data on COVIDnet.fr participants, incidence, and healthcare-seeking behavior are made available with the code.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In the fitting procedure, at each iteration, we averaged 500 stochastic simulations obtained from the same parameter sets to compute the average expected value for the likelihood. For missing data on clinical status at the time of testing and date of onset for virologically confirmed cases, we averaged over 100 independent imputations. For SI-DEP database, we included all positive tests performed in mainland France (no sampling needed). For COVIDnet, we considered all subjects enrolled in the surveillance system. Number of participants are provided with the code at https://github.com/EPIcx-lab/COVID-19/tree/master/Underdetection_France . The COVIDnet cohort is not representative of the French population. Analyses were adjusted to correct for this bias as presented in the methods section and in the references therein cited.
Data exclusions	We focused on mainland France. We excluded Corsica region, reporting a very limited epidemic activity, and overseas territories, characterized by increasing transmission.
Replication	The study contains several stochastic models: the epidemic model, and the imputation of pre-symptomatic vs asymptomatic, missing data and date of onset. Each model replication was independently executed, by assigning a distinct seed to initialize the random number generator (numpy v 1.19). The fitting procedure was independently coded in 2 languages (Python and R) and we used NOMAD and Nelder-Mead for optimization.
Randomization	Our study is observational and population-based in nature, and involves no intervention. Randomization of subjects is therefore not applicable.
Blinding	Our study is observational and population-based in nature, and involves no intervention. Blinding techniques are therefore not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Population of volunteers resident in France who weekly self-declare their symptoms, along with socio-demographic information (age, gender, geographic information, education, occupation, health indicators e.g. vaccination, diet, pregnancy, smoking). Representativeness was previously studied (refs cited in the paper). Analyses were adjusted by age and sex, as in previous works (refs cited in the paper).

Recruitment

Recruitment occurs through communications of supporting institutions, mainstream and social media, word of mouth. Associated biases were studied in previous works (cited in the paper). Analyses accounted for selection bias, as in previous works.

Ethics oversight

GrippeNet.fr/COVIDnet.fr was reviewed and approved by the French Advisory Committee for research on information treatment in the health sector (i.e. CCTIRS, authorization 11.565), and by the French National Commission on Informatics and

Liberty (i.e. CNIL, authorization DR-2012-024) – the authorities ruling on all matters related to ethics, data, and privacy in the country. Informed consent was provided by each participant at enrollment, according to regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.