nature

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

Accelerated Article Preview

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

Received: 11 August 2020

Accepted: 8 December 2020

Accelerated Article Preview Published online 21 December 2020

Cite this article as: Pullano, G. et al. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* https://doi.org/10.1038/s41586-020-03095-6 (2020).

Giulia Pullano, Laura Di Domenico, Chiara E. Sabbatini, Eugenio Valdano, Clément Turbelin, Marion Debin, Caroline Guerrisi, Charly Kengne-Kuetche, Cécile Souty, Thomas Hanslik, Thierry Blanchon, Pierre-Yves Boëlle, Julie Figoni, Sophie Vaux, Christine Campèse, Sibylle Bernard-Stoecklin & Vittoria Colizza

Article

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

Giulia Pullano[1,2,7], Laura Di Domenico[1,7], Chiara E. Sabbatini[1], Eugenio Valdano[1], Clément Turbelin[1], Marion Debin[1], Caroline Guerrisi[1], Charly Kengne-Kuetche[1], Cécile Souty[1], Thomas Hanslik[1,3,4], Thierry Blanchon[1], Pierre-Yves Boëlle[1], Julie Figoni[5], Sophie Vaux[5], Christine Campèse[5], Sibylle Bernard-Stoecklin[5] & Vittoria Colizza[1,6]✉

As countries in Europe gradually relaxed lockdown restrictions after the first wave, test-trace-isolate strategies became critical to maintain COVID-19 viral activity at low levels[1,2]. Reviewing their shortcomings can provide elements to consider in light of the second wave currently underway in Europe. Here we estimate the rate of detection of COVID-19 symptomatic cases in France after lockdown through the use of virological[3] and participatory syndromic[4] surveillance data coupled with mathematical transmission models calibrated to regional hospitalizations[2]. Our findings indicate that around 90,000 incident symptomatic infections, corresponding to 9 out 10 cases, were not ascertained by the surveillance system in the first 7 weeks following lockdown from May 11 to June 28 2020, although the test positivity rate did not exceed WHO recommendations (5%)[5]. The median detection rate increased from 7% [6-8]% to 38% [35-44]% over time, with large regional variations, owing to a strengthening of the system as well as a decrease of 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 COVID-19 cases did not seek medical advice despite recommendations, as confirmed by serological studies[6,7]. Encouraging awareness and same-day healthcare-seeking behavior in suspect cases is critical to improve detection. However, the capacity of the system remained insufficient even at the low levels of viral circulation achieved after lockdown, and was predicted to deteriorate rapidly with increasing epidemic activity. Substantially more aggressive, targeted, and efficient testing with easier access is required to act as a pandemic-fighting tool. Testing strategy will be once again of critical value to lift current restrictive measures in Europe and avoid a third wave.

Surveillance and detection aim to rapidly identify and isolate cases to prevent onward transmission in the community and avoid substantial resurgence of cases. After an initial period where limited testing capacity mainly focused on severely ill patients, a novel testing policy was implemented in France to systematically screen potential COVID-19 infections and allow lifting of lockdown restrictions on May 11[8].

The specific characteristics of COVID-19 epidemic, however, hinder the identification of cases[9–11]. Large proportions of asymptomatic infectious individuals[12], and presence of mild or paucisymptomatic infections that easily go unobserved[9,11] present serious challenges to detection and control[9,10,13]. Missing a substantial portion of the epidemic compromises the control effort, leaving the virus to silently spread[10–12]. Synthesizing evidence from virological[3] and participatory syndromic surveillance[4] with mathematical modeling[2,14] accounting for behavioral 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

COVID-19 epidemic management in France after lockdown in spring 2020 involved the creation of a centralized database collecting all data on virological testing (SI-DEP[3], Information system for testing). All individuals with symptoms compatible with COVID-19[19] were invited to consult their general practitioner and obtain a prescription for a virological test[8]. Contacts of confirmed cases were traced and tested. 20,777 virologically-confirmed cases were notified from May 13 (week 20) to June 28 (week 26) in mainland France. These include positive

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

individuals with or without symptoms at the time of testing, or positive individuals for which information on clinical status at the time of testing was missing (ED Figure 1). Accounting for presymptomatic individuals among those presenting with no symptoms at the time of testing and after imputation of missing data (see Methods), an estimated 16,165 [95%CI16,101-16,261] symptomatic cases were tested in the study period (Figure 1a). The average delay from symptom onset to testing decreased from 12.5 days in week 20 (w20) to 2.8 days in w26 (Figure 1b and ED Figure 1). Accounting for this delay (see Methods and ED Figure 2), we estimated that 14,061 [13,972-14,156] virologically-confirmed symptomatic cases had onset of symptoms in the study period, showing a decreasing trend over time (2,493 in w20, 1,647 in w26). The test positivity rate decreased in the first weeks and stabilized around 1.2% (average over w24-w26).

A digital participatory system was additionally considered for COVID-19 syndromic surveillance in the general population[20], including those who do not consult a doctor. Called COVIDnet.fr, it was adapted from the platform GrippeNet.fr (dedicated to influenza-like-illness surveillance[4]) 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 socio-demographic information. Based on symptoms declared by an average of 7,500 participants each week, the estimated incidence of COVID-19 suspect cases[19] decreased from about 1% to 0.8% over time (Figure 1c). 162 out of 524 suspect cases (31%) consulted a doctor in the study period. Among them, 89 (55%) received a prescription for a test, resulting in screening for 50 individuals (56% of those given the prescription) (Figure 1d).

Projections of COVID-19 epidemic trajectories and estimated detection rates

We used stochastic discrete age-stratified epidemic models[2,14] based on demography, age profile[21], and social contact data[15] of the 12 regions of mainland France, to account for age-specific contact activity and role in COVID-19 transmission. Disease progression is specific to COVID-19[2,14] and parameterized with current knowledge to include presymptomatic transmission[22], asymptomatic[12] and symptomatic infections with different degrees of severity[9,11,23,24]. The model was shown to capture the transmission dynamics of the epidemic in Île-de-France in the first wave and was used to assess the impact of lockdown and exit strategies[2,14]. Full details are reported in the Methods section.

Intervention measures were modeled as mechanistic modifications of the contact matrices, accounting for a reduction of the number of contacts engaged in specific settings, and were informed from empirical data. Lockdown data came from Refs. [2,14]. The exit phase was modeled considering region-specific data of attendance at school based on Ministry of Education's data[16], partial presence at workplaces based on estimates from mobile phones location history data[17] (Figure 1e), reduction in adoption of physical distancing over time and increased risk aversion of seniors based on survey data[18] (Figure 1f), 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 of contacts for the virologically-confirmed COVID-19 cases[2,14]. Region-specific models were fitted to regional hospital admission data (Figure 2) through a maximum likelihood approach. Further details are reported in Methods and Supplementary Information.

Projected number of cases decreased over time in all regions, in agreement with the decreasing tendency reported in hospital admissions in the study period (Figure 2 and ED Figure 3). Overall, 103,907 [95%CI 90,216-116,377] new symptomatic infections were predicted in mainland France in weeks 20-26 (from 35,704 [30,290-40,748] in w20 to 4,319 [3,773-4,760] in w26). Île-de-France was the region with the largest predicted number of symptomatic cases (12,427 [8,104-14,136] to 1,704 [1,258 -2,004] from w20 to w26), followed by Grand Est and Hauts-de-France (Table 1 and ED Table 1).

Projections were substantially higher than virologically-confirmed cases (Figures 2 and 3). The estimated detection rate for symptomatic infections in mainland France in the period w20-w26 was 14% [12-16]%, suggesting that about 9 out of 10 new cases with symptoms were not identified by the surveillance system. Lower detection was found for asymptomatic infections (ED Figure 5). Estimated detection rate increased over time (7% [6-8]% in w20, 38% [35-44]% in w26, Table 1). By the end of June, 5 regions had a median detection above 50%, and 6 regions detected a number of cases within the confidence interval of model projections (Figure 3b–d). All regions except Brittany displayed average increasing trends in the estimated detection rate in June compared to May. We did not find significant associations between detection rate and number of detected cases, or test positivity rate (ED Figure 4). However, the detection rate was negatively associated with model-predicted incidence (Spearman correlation $r = -0.75$, $p < 10^{-15}$; Figure 3f). In addition, the data followed a power-law function, $\pi = 66 \cdot i^{-0.51}$, with π the weekly detection rate of symptomatic cases (expressed in %) and i the projected weekly incidence (cases per 100,000). This function quantifies the relation between detection capacity of test-trace-isolate system and viral circulation in the population. It clearly shows that capacity rapidly drops as epidemic activity increases.

Validation of the model was performed in two ways. First, we compared our model projections of the percentage of infected population with the results of three independent seroprevalence studies performed after the first wave in France[7,25,26] (see Methods). Modeling results are in agreement with serological estimates at national and regional level (Figure 3e and ED Figure 6). Second, we compared the projected incidence of COVID-19 symptomatic cases in w26 (6.69 [5.84-7.37] per 100,000) with the value obtained from virologically-confirmed cases (2.55 [2.48-2.61] per 100,000) and two estimates based on COVIDnet.fr data (Figure 3g). The first estimate applies the measured test positivity rate to the incidence of self-reported COVID-19 suspect cases (estimate #1, yielding 8.6 [95%CI 6.2-11.5] per 100,000); the second additionally assumes that only 55% would be confirmed as suspect case by a physician and prescribed a test (according to COVIDnet.fr data, estimate #2, yielding 4.7 [3.4-6.3] per 100,000). Our projections are in line with plausible estimates from COVIDnet.fr, and suggest that on average at least 80% of suspect cases should be tested to recover the predicted incidence.

Sensitivity analysis showed that findings were robust against elements of the contact matrices that could not be informed by empirical data (Figure S8, S9). Also, a model selection analysis showed that changes in the contact patterns over time due to restrictions and activities of individuals of different age classes after lockdown (e.g. for partial attendance at school, telework) are needed to accurately capture the transmission dynamics (Table S2, Figure S5).

Discussion

Despite a test positivity rate in mainland France well below WHO recommendations (5%)[5], 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 Switzerland[27]. Surveillance improved substantially over time, leading to half of French regions reporting cases compatible with model projections. The new framework progressively strengthened with increasing resources over time, as signaled by a more rapid detection of cases (78% reduction of 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 of the 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 predicted to have large numbers of weekly infections (Île-de-France detecting only 1 out of 3 symptomatic cases by the end of June, and Grand Est, 1 out of 5). Almost all patients (92%) clinically diagnosed by sentinel general practitioners as suspect cases were prescribed a test[20]. 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 COVID-19 cases were not screened because they did not seek medical advice despite recommendations. This was confirmed by serological studies. In France, only 48% of symptomatic participants with antibodies against SARS-CoV-2 reported consulting a general practitioner[7]; in Spain, between 16% and 20% reported a previous virological screening[6]. By combining estimates from virological and participatory surveillance, we extrapolated an incidence from crowdsourced data that is compatible with model projections, under the hypothesis that the large majority of suspect cases would get tested (>80%). This finding further supports testing for all COVID-19 suspect cases. Large-scale communication campaigns should reinforce recommendations to raise awareness in the population and strongly encourage healthcare-seeking behavior especially in patients with mild symptoms. At the same time, investigations to identify reasons for not consulting could be quickly performed through the participatory surveillance system.

Red tape might have contributed to low testing rates. Prescription for a test was deemed compulsory in the new testing policy to prevent misuse of diagnostic resources[8], however the path involving consultation, prescription, and lab appointment may have discouraged mildly affected individuals not requiring medical assistance. To facilitate access, testing should not require a prescription, as later established by authorities[28]. 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 with the population[29]. The use of antigen tests will further facilitate access. These initiatives are particularly relevant to counteract socio-economic inequalities in access to care in populations vulnerable to COVID-19[30,31]. However, such strategies should not hinder a testing protocol targeting suspect 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 low positivity rate[32]. Without a strong case-based surveillance, the risk is to disperse resources towards random individuals without symptoms who are unlikely to be positive. This might saturate the test-trace-isolate machinery, as observed during summer[33], without slowing down viral circulation required to safeguard the hospital system.

Given pre-symptomatic transmission, notification to contacts should be almost immediate to allow the effective interruption of transmission chains[22]. For testing to be an actionable tool to control COVID-19 transmission, delays should be suppressed and screening rates radically increased but better targeted. Over May-June the average weekly number of tests was 250,000, remaining well below the objective originally fixed by authorities (700,000). Number of tests increased over summer, but proportionally to the increased viral circulation. The capacity of detection of the test-trace-isolate system scaled as the inverse of the square root of the incidence, rapidly deteriorating already at low incidence levels. More aggressive testing targeting suspect index cases should be performed at low viral circulation to avoid case resurgence. The system was predicted to be able to detect more than 2 cases out of 3 (rate>66%) only if the incidence was lower than 1 symptomatic case per 100,000, a figure 50 times smaller than estimated at the exit from lockdown. As detection of at least 50% of cases is needed to control the epidemic avoiding strict social distancing[2], these results indicate that the system was insufficient to perform a comprehensive case-based surveillance, as recommended to phase out restrictions[5]. Current restrictions applied in Europe to curb the second wave offer a second opportunity to improve testing policies and support lifting of these measures in the upcoming weeks. Failing in doing so may lead to rapid and uncontrolled increase of cases[2,34]. Such risk is even stronger in the winter season and with perduring adhesion fatigue[18].

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 lockdown[30], and by the limited mobility increase in May-June, before important inter-regional displacements took place at the start of summer holidays in July. Foreign importations[35] were neglected as France reopened its borders with EU Member States on June 15, and the Schengen area remained closed till July. COVIDnet.fr cohort is not representative of the general population, however prior work on influenza-like-illness showed that adjusted incidence was in good agreement with sentinel estimates[4]. Underdetection may also proceed from the imperfect characteristics of RT-PCR (reverse transcription-polymerase chain reaction) tests used to identify infections[36]. Some cases tested for SARS-CoV-2 could have been falsely negative, e.g. because tested too early after the infection, thus further increasing underdetection. Previous work assessed underdetection rates in 210 countries[32], but it mainly focused on the early global dynamics. Our model gives up geographical extent for higher data quality in a specific country, providing a novel synthesis of data sources characterizing human behavior over time and space with virological and participatory surveillance data to identify the weak links of epidemic response.

Our findings identify critical needs of improvement of the test-trace-isolate response system to control COVID-19 epidemic. Substantially more aggressive and efficient testing targeting COVID-19 suspect cases needs to be achieved to act as a pandemic-fighting tool. Associated communication and logistical needs should not be underestimated. These elements should be considered to allow the lifting of restrictive measures currently applied to curb the second wave 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.

1. 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/#section-d1e1225 (2020).
2. 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).
3. 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).
4. Guerrisi, C. *et al*. The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: the case of France. *Eurosurveillance* **23**, 1700337 (2018).
5. WHO. Public health criteria to adjust public health and social measures in the context of COVID-19. https://www.who.int/publications/i/item/considerations-in-adjusting-public-health-and-social-measures-in-the-context-of-covid-19-interim-guidance (2020).
6. Pollán, M. *et al*. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* **396**, 535–544 (2020).
7. 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://www.medrxiv.org/content/10.1101/2020.09.16.20195693v1 (2020).
8. Ministère des Solidarités et de la Santé. Prise en charge par les médicins 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).
9. Verity, R. *et al*. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis*. **20**, 669–677 (2020).
10. Pinotti, F. *et al*. Tracing and analysis of 288 early SARS-CoV-2 infections outside China: A modeling study. *PLOS Med*. **17**, e1003193 (2020).
11. Li, R. *et al*. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).

12. Lavezzo, E. *et al*. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429 (2020).
13. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci*. **101**, 6146–6151 (2004).
14. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Can we safely reopen schools during COVID-19 epidemic? Preprint at https://www.medrxiv.org/content/10.1101/2020.05.08.20095521v2 (2020).
15. 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).
16. 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).
17. Google.com. COVID-19 Community Mobility Report. https://www.google.com/covid19/mobility?hl=fr (2020).
18. 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).
19. HCSP. *Signes cliniques d'orientation diagnostique du Covid-19. Rapport de l'HCSP* https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=812 (2020).
20. 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).
21. INSEE.fr. Pyramide des âges 2020 - France et France métropolitaine. https://www.insee.fr/en/statistiques/3696316 (2020).
22. Ferretti, L. *et al*. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
23. Riccardo, F. *et al*. Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.08.20056861v1 (2020).
24. Salje, H. *et al*. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
25. Vu, S. L. *et al*. Prevalence of SARS-CoV-2 antibodies in France: results from nationwide serological surveillance. Preprint at http://medrxiv.org/lookup/doi/10.1101/2020.10.20.20213116 (2020).
26. Ministère des Solidarités et de la Santé: enquête nationale EpiCov. En mai 2020, 4,5% de la populationen France métropolitaine a développédes anticorps contre le SARS-CoV-2. https://drees.solidarites-sante.gouv.fr/IMG/pdf/er1167.pdf (2020).
27. Stringhini, S. *et al*. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *The Lancet* **396**, 313–319 (2020).
28. Ameli.fr. 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).
29. LeParisien.fr. Covid-19: une campagne de tests à grande échelle débute dans 32 communes d'Ile-de-France. https://www.leparisien.fr/societe/covid-19-une-campagne-de-tests-a-grande-echelle-debute-dans-32-communes-d-ile-de-france-29-06-2020-8344329.php (2020).
30. 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).
31. 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://www.medrxiv.org/content/10.1101/2020.11.07.20201335v1 (2020).
32. Russell, T. W. *et al*. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med*. **18**, 332 (2020).
33. LeFigaro.fr. Explosion des délais d'attente des résultats aux tests Covid-19. https://www.lefigaro.fr/actualite-france/explosion-des-delais-d-attente-des-resultats-aux-tests-covid-19-20200922 (2020).
34. 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).
35. Ruktanonchai, N. W. *et al*. Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* **369**, 1465–1470 (2020).
36. Wernike, K. *et al*. Pitfalls in SARS-CoV-2 PCR diagnostics. *Transbound. Emerg. Dis. tbed*.13684 https://doi.org/10.1111/tbed.13684 (2020).

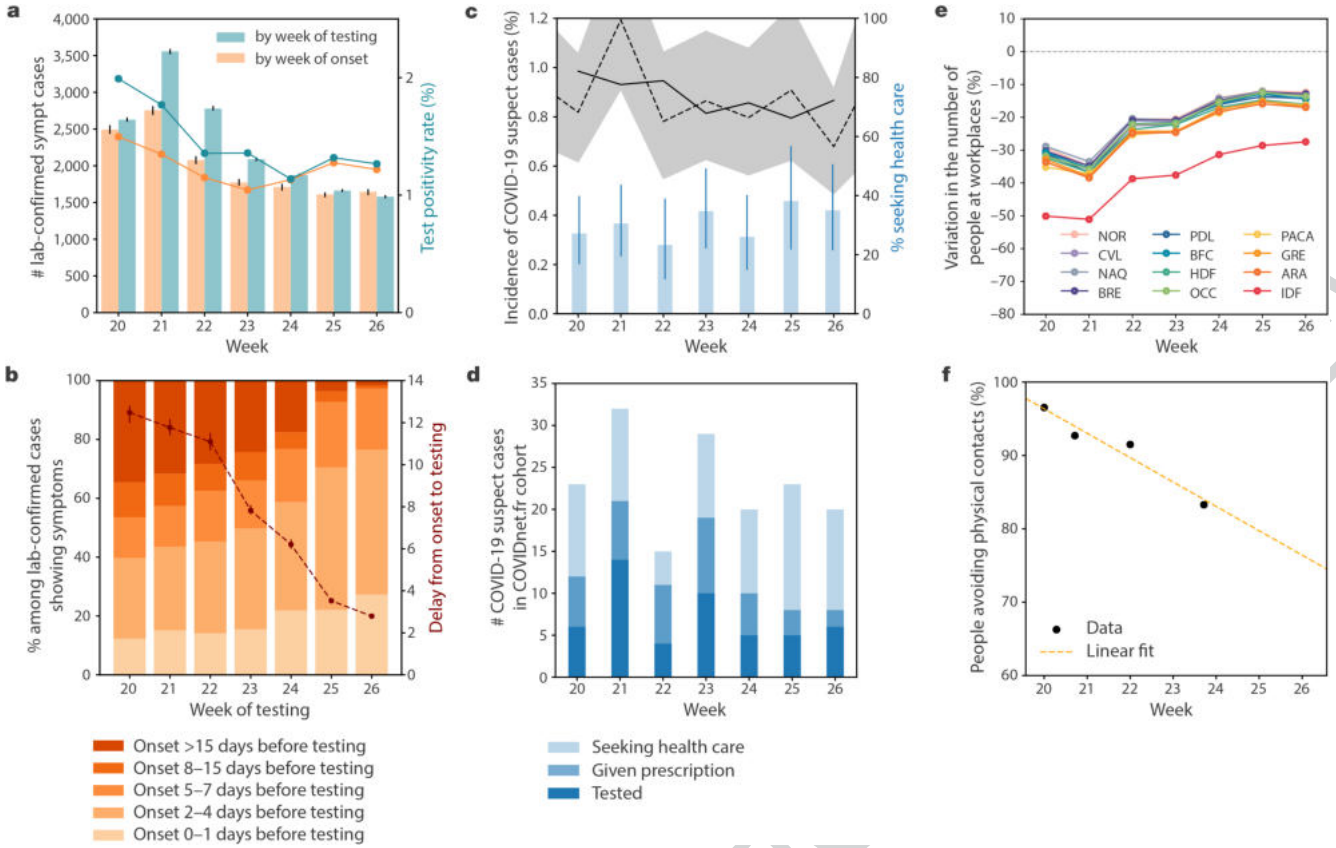


Fig. 1 | Virological surveillance, participatory syndromic surveillance, behavioral data for model parameterization. (a) Estimated number of virologically-confirmed symptomatic cases in mainland France by week of testing and onset, and test positivity rate. Estimates are based on imputation of positive individuals without symptoms 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; imputation of the date of onset of symptoms for presymptomatic and symptomatic cases (see Methods). Imputations were performed n=100 times. Uncertainties (black bars) correspond to 95%CI. Test positivity rate was computed on cases with complete information. Data for w20-26 were consolidated in w30. (b) Breakdown of virologically-confirmed cases with symptoms and complete information in SI-DEP database by week of testing according to declared onset of symptoms (n=5,514). Estimated time from onset to testing is shown (right y axis, median and 95%CI, obtained with n=100 imputations of onset date). (c) Weekly incidence of COVID-19 suspect cases (median (dashed line), 95%CI (shaded area), and 3-week moving average (continuous line)), and percentage of those seeking healthcare (median and 95%CI), estimated from participatory surveillance system COVIDnet.fr (average weekly n=7,481). (d) Number of COVID-19 suspect cases of the participatory cohort seeking healthcare, and among them those receiving a prescription, and performing a virological test given the prescription. (e) Estimated change in presence at workplace locations over time and by region based on Google location history data[17]. Region acronyms are listed in Table 1. (f) Percentage of individuals avoiding physical contacts with respect to lockdown estimated from a large-scale survey conducted by Sante publique France[18].

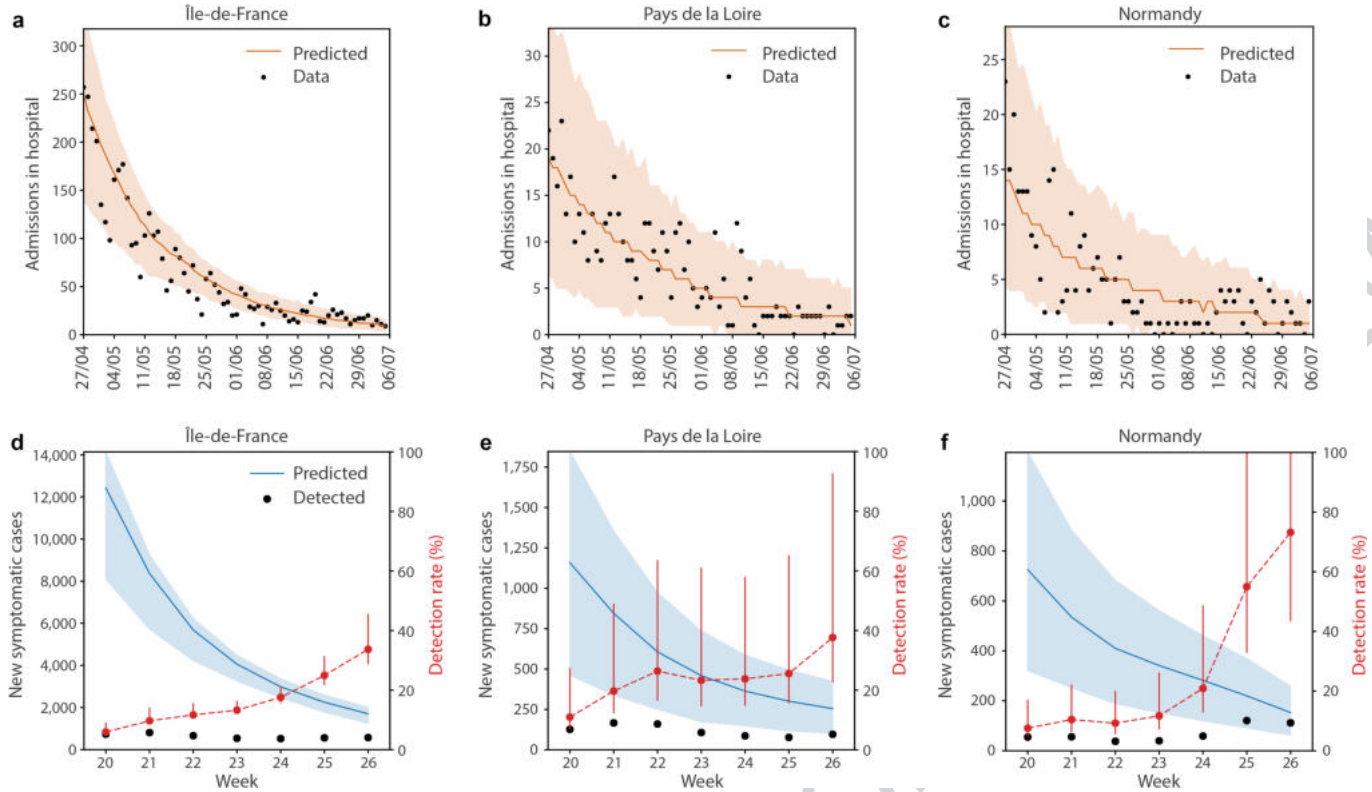


Fig. 2 | Hospital admissions and number of new symptomatic cases. (a-c) Hospital admissions over time, data (points) and simulations (median and 95%CI), for Île-de-France (a), Pays de la Loire (b), Normandy (c). Hospital admission data up to w27 (consolidated in w28) were used to infer parameter values. (d-f) Projected number of new symptomatic cases over time (median and 95%CI) and estimated number of virologically-confirmed symptomatic cases by week of onset (points), for the same regions above. The estimated detection probability of symptomatic cases (%) is also shown (red points, median and 95%CI, right y axis). In all panels, 95%CI are obtained from n=500 independent stochastic runs. Plots for the remaining regions are reported in ED Figure 3.

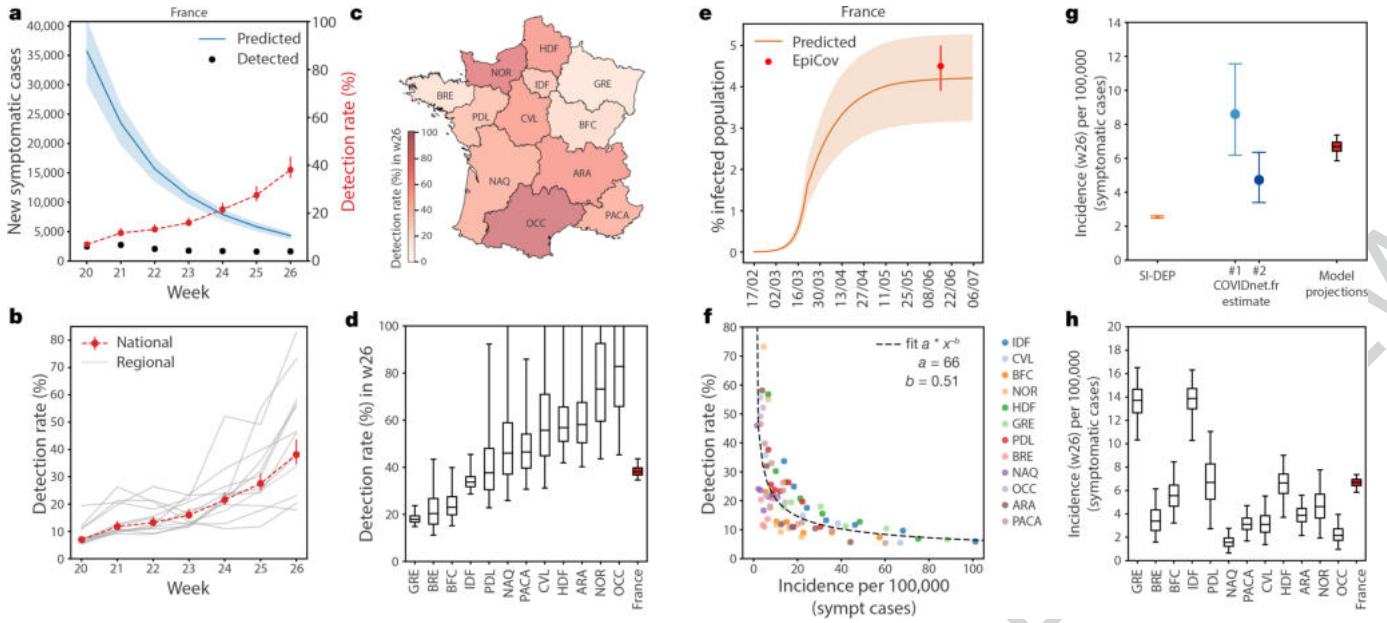


Fig. 3 | Detection rate and incidence. (a) Projected number of new symptomatic cases over time (median and 95%CI) and estimated number of virologically-confirmed symptomatic cases by week of onset (points) in mainland France. The estimated detection rate of symptomatic cases (%) is also shown (red points, median and 95%CI, right y axis). (b) Estimated detection rate of symptomatic cases (%) and 95%CI over time for mainland France (red dots and bars), and for all regions (grey lines, only median values are shown for the sake of visualization). (c) Map of the estimated detection rate (%) by region in w26 (June 22-28, 2020). (d) Estimated detection rate per region compared to national estimate. Regions are ranked by increasing median detection rate. Boxplots 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 population infected (median and 95%CI) compared with estimates from the serological study EpiCov[26] performed on a representative sample of the population in mainland France. (f) Estimated detection rate of symptomatic cases (%) by region and by week vs. projected incidence by region and by week. The curve indicates the result of a least-square fit to the data with a power-law function, $\pi = a \cdot i^{-b}$, with π the detection rate (expressed in %) i the weekly incidence (cases per 100,000), $a = 66$ (95%CI [52, 85]) and $b = 0.51$ [0.41, 0.60]. (g) Estimated incidence of symptomatic cases and 95%CI in mainland France in w26 from different sources: virological surveillance data (SI-DEP), participatory surveillance data (COVIDnet.fr, with two estimates), model projections. (h) Projected incidence per region compared to the national estimate. Regions are ranked as in panel d. Boxplots with whiskers are defined as in panel d. In all panels, median and 95%CI for model projections are obtained from n=500 independent stochastic runs.

Table 1 | Population per region, estimated number of virologically-confirmed symptomatic cases, projected number of symptomatic cases, estimated detection rate, estimated trend in 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 (%) for symptomatic cases (median and 95%CI)		Trend in detection rate (average June vs. average May)	
		w20	*w26*	*w20*	*w26*	*w20*	*w26*		
Île-de-France (IDF)	12.3	737	574	12,427 [8,104 - 14,136]	1,704 [1,258 - 2,004]	6 [5 - 9]	34 [29 - 46]	+146%	+
Grand Est (GRE)	5.5	323	135	4,868 [2,992 - 5,848]	756 [568 - 914]	7 [6 - 11]	18 [15 - 24]	+99%	+
Hauts de France (HDF)	6.0	308	225	4,476 [2,381 - 6,648]	396 [219 - 538]	7 [5 - 13]	57 [42 - 100]	+186%	+
Auvergne-Rhône-Alpes (ARA)	8.0	204	181	3,552 [2,017 - 5,283]	312 [173 - 451]	6 [4 - 10]	58 [40 - 100]	+244%	+
Occitanie (OCC)	5.9	166	106	851 [397 - 1,400]	128 [57 - 235]	19 [12 - 42]	83 [45 - 100]	+165%	+
Provence-Alpes-Côte d'Azur (PACA)	5.1	164	73	3,040 [1,665 - 4,625]	157 [83 - 239]	5 [4 - 10]	46 [31 - 88]	+289%	+
Pays de la Loire (PDL)	3.8	127	96	1,158 [463 - 1,846]	255 [103 - 423]	11 [7 - 27]	38 [23 - 93]	+45%	+
Bourgogne-Franche-Comté (BFC)	2.8	118	36	1,591 [854 - 2,379]	154 [88 - 235]	7 [5 - 14]	23 [15 - 40]	+95%	+
Nouvelle Aquitaine (NAQ)	6.0	115	43	1,040 [482 - 1,691]	94 [38 - 166]	11 [7 - 24]	46 [26 - 100]	+54%	+
Centre-Val de Loire (CVL)	2.6	94	44	1,706 [812- 2,511]	79 [34 - 142]	6 [4 - 12]	56 [31 - 100]	+187%	+
Brittany (BRE)	3.3	80	23	672 [294 - 1,155]	113 [51 - 206]	12 [7 - 27]	20 [11 - 45]	-28%	—
Normandy (NOR)	3.3	55	112	725 [322 - 1,194]	153 [63 - 258]	8 [5 - 17]	73 [43-100]	+342%	+
France*	64.6	2,493	1,647	35,704 [30,290 - 40,748]	4,319 [3,773 - 4,760]	7 [6 - 8]	38 [35 - 44]	+142%	+

Regions are ranked by decreasing number of confirmed cases in w20. The trend is computed comparing the average of the estimated detection rate in the weeks of June (w23-26) with the average in the weeks of May (w20-w22). Median and 95%CI are obtained from n=500 independent stochastic runs. *Mainland France (Corsica and overseas territories excluded).

Methods

Virological surveillance data

The centralized database SI-DEP for virological surveillance[3] collects all tests performed in France for any reason. In the period under study, guidelines recommended to consult a general practitioner at the first sign of COVID-19-like symptoms and obtain a prescription for a virological test (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 (e.g. in some senior homes or long-term healthcare facilities). Data report detailed information on individuals tested in France, including (i) date of test, (ii) result of test (positive or negative), (iii) location (region), (iv) absence or presence of symptoms at the time of testing, (v) self-declared delay between onset to 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 >15 days before. For some tests, information on (iv) and (v) is missing. The SI-DEP database provided complete information for 23,210 (66%) out of 35,264 laboratory-confirmed COVID-19 cases tested between week 20 (May 11-May 17) and week 30 (July 19-July 26), with an increasing trend of complete information over time (from 49% in week 20 to 76% in week 30, ED Figure 1). Among confirmed cases with complete information, 12,716 (55%) showed no symptoms at the time of testing (ED Figure 1). The study referred to the period from w20 to w26. Data up to w30 were used to consolidate the data in the study period accounting for the delays.

Imputation of asymptomatic vs. presymptomatic cases, of onset date, and of 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/absence of symptoms at the time of testing. These three subsets of cases were analyzed to account for presence of presymptomatic individuals among those with no symptoms at time of testing, imputation of missing data, estimation of dates of infection or symptom onset.

• For laboratory-confirmed COVID-19 cases with 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 self-declared by the patients (Figure 1b). In the time period between w20 and w30, 20% of cases had onset-to-test delay ≤1day, 63% had delay ≤4 days, 83% had delay ≤7, and 88% had delay ≤15 days (ED Figure 1). We fitted a Gamma distribution to onset-to-test delay data with a maximum likelihood approach, using three different periods of time (May, June, July), to account for changes in the distribution of self-declared delays over time (i.e. longer delays at the beginning of the study period, shorter delays at its end, ED Figure 2). The estimated average delay in May, June, July was equal to 12.9 (95%CI [7.0, 16.1]), 5.1 (95%CI [3.7, 6.3]), 2.7 (95%CI [2.0, 3.1]) days respectively. July data was 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 March 17, 2020 (week 12), therefore we 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 COVID-19 cases with no symptoms at the time of testing, we assumed that on average 40% of them were asymptomatic[12] (see transmission model subsection), 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, 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 mean 1.5 days, corresponding to the duration of the prodromic phase in our model (Table S1). 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 COVID-19 cases 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 5 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 (Figure 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 www.covidnet.fr. It was adapted from GrippeNet.fr[4] to respond to the COVID-19 health crisis in March 2020. GrippeNet.fr is a participatory system for the surveillance of influenza-like-illness available in France since 2011 through a collaboration between Inserm, Sorbonne Université and Sante 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 enrollment[38]. In addition to tracking influenza-like-illness incidence[4,37], GrippeNet.fr was used to estimate vaccine coverage in specific subgroups[39] individual perceptions toward vaccination[40] and healthcare-seeking behavior[41]. It was also used to assess behaviors and perceptions related to other diseases beyond influenza[42], including COVID-19[43].

Participants are on average older and include a larger proportion of women compared to the general population[38,44]. Participating population is however representative in terms of health indicators such as diabetes and asthma conditions. Despite these discrepancies, trends of estimated influenza-like-illness incidence 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 COVID-19 suspect cases 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 notice[19]:

• (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 (Sudden onset of symptoms OR (sudden onset of fever AND fever)) AND

o (age > 5 years old AND (feeling tired or exhausted OR muscle/joint pain OR headache OR (loss of smell WITHOUT runny/blocked nose) OR loss of taste)
o OR ((Age ≥ 80 years old OR Age < 18 years old) AND diarrhea)
o OR (Age < 3 months old AND (fever WITHOUT other symptoms))).
Figure 3 reports two independent estimates obtained from COVIDnet.fr cohort data for the incidence of symptomatic cases in w26. They are computed as follows:
• *Estimate #1* = (COVIDnet.fr estimated incidence of suspect cases in w26) * (test positivity rate from SI-DEP in w26)
• *Estimate #2* = (COVIDnet.fr estimated incidence of suspect cases in w26) * (estimated proportion screened and confirmed as COVID-19 suspect case by a physician, and prescribed a test; estimates from COVIDnet.fr) * (test positivity rate from SI-DEP in w26)

The two estimates were used to validate model projections and identify the specific surveillance mechanisms needing 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 (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.

Transmission models summary

We used a stochastic discrete age-stratified transmission model for each region based on demographic, contact[15], and age profile data of French regions[21]. 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 reporting a very limited epidemic activity and overseas territories characterized by increasing transmission[20].

Four age classes were considered: [0-11), [11-19), [19-65), and 65+ years old, referred to as children, adolescents, adults, seniors. Transmission dynamics follows a compartmental scheme specific to COVID-19, where individuals were divided into susceptible, exposed, infectious, and hospitalized (Supplementary Information, Figure S1, S2). We did not consider further progression from hospitalization (e.g. admission to ICU, recovery, death[2]) 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 where individuals may remain either asymptomatic (I_a, with probability p_a=40%[12]) 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_β=0.55, as estimated in Ref. [11], 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 modeling works[48–53]. A sensitivity analysis was performed on relative susceptibility and transmissibility of children, and on the proportion of asymptomatic infections (Figure S10-S13). Full details are reported in the Supplementary Information.

The study was not extended to the summer months, because of (i) the challenge of mechanistically parameterizing the contact matrices during summer, (ii) the increase of movement fluxes across regions weakening our assumption of region-specific models, and (ii) the interruption of COVIDnet.fr surveillance during the summer break preventing 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 France[15]. This refers to the baseline condition, i.e. pre-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 school[16], the percentage of workers going to the workplace[17], the compliance to preventive measures[18], with a higher compliance registered in senior individuals[18]. 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. More in detail:
• ***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 Education[16] (Figure S3). 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%, corresponding for example to the attendance registered in Île-de-France in pre-/primary schools, corresponds to a reduction of 85.5% in the number of contacts established at school by students belonging to that school levels. Contacts on transports were modified accordingly.
• ***Presence at work***. To account for the percentage of individuals at work, given recommendations on telework and activities that were not yet reopened, we used the estimated variation of presence at workplaces based on mobile phone location data and provided by Google Mobility Trends[17]. Contacts at work and on transports were therefore modified according to this percentage, analogously to contacts at school. Household contacts were increased proportionally to each adult staying at home based on statistics comparing weekend vs. weekday contacts[15] and proportion of adults working during the weekend[54], as done previously[2].
• ***Adoption of physical distancing***. Our previous work showed that physical contacts during lockdown were fully avoided[2], in agreement with data collected afterwards[18]. To account for individual adoption of preventive behavior after lockdown, we used the percentage of population avoiding physical contacts estimated from a large-scale survey conducted by Santé Publique France (*CoviPrev*[18]). Data were fitted with a linear regression (Figure 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 seniors***. Data from Santé Publique France survey *CoviPrev*[18] also show that seniors protected themselves further relative to other age classes. On average, they respected physical distancing 28% more than the other age classes (Figure S4). For this reason, we considered a further reduction of 30% of contacts in seniors 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:
• Prior to lockdown, $\{\beta,\ t_0\}$ where β is the transmission rate per contact and t_0 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} 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, $\pi_a(w), \pi_s(w)$ 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 work[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 persons 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 (June 29-July 5) 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, July 6-12, 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

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

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 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 (w20-w26).

We reduced the required computations by making the optimization in 2 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 describing the epidemic trajectory prior to the exit phase to their MLEs. This second step was further simplified through an iterative procedure, and we showed through simulations that the simplified optimization procedure is consistent and well defined. The parameter space was explored using NOMAD software[55]. 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 posterior parameter distribution using Markov Chain Monte Carlo (MCMC) would also have been an alternative to ML 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 (Figure S6, Figure S7, Table S3).

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 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 stochastic fluctuation only, 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 with 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 normalize the contact matrix to its largest eigenvalue and estimate 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 (Table S2 and Figure S5).

Comparison with serological estimates

We compared model projections with serological estimates from three independent studies[7,25,26] (Figure 3e, ED Figure 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 Sante publique France (SpF in the plots)[25] are based on at least 1 positive result in one of the following three tests: ELISA-S, ELISA-NP, and a pseudo-neutralization test which 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 lack of representativeness.

Estimates by EpiCoV[26] ("Enqueˆte Épidémiologie et Conditions de vie liées à la Covid-19") used ELISA-S tests and further validated them 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 the plots the estimates 14 days prior to 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.

Modeling results are in good agreement with the serological estimates at the national level (Figure 3e of the main text) and in the large majority of the regions (ED Figure 6). Projections tend to be systematically smaller than serological estimates in two regions weakly affected by the epidemic (Pays de la Loire, Brittany), though remaining compatible with observations.

Overall differences may be due to the limitations of the methods involved:

• The type of tests, the specificity levels, the samples of the population tested, and the weighting and imputation approaches considered in each serological study, leading to differences across the three investigations. We note for example that larger discrepancies are observed between EpiCov and SpF results in those regions that experienced smaller epidemics. We used EpiCov as the reference study since it is the only one that was conducted on a representative sample of the population.

• The limitations of the dataset of hospital admissions used to calibrate the models: the db infrastructure for data collection became operational in mid-March and filled 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 data[24] 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 models parameterization (contact matrices[15], school attendance[16], presence at workplaces[17], avoidance of physical contacts[18]) are available at the cited sources. Hospitalization data used for model calibration are made available with the code. 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 GrippeNet.fr/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. 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. *Br. J. Gen. Pract.* **1**, bjgpopen17X101253 (2018).
42. Mathieu, P. *et al*. Population perception of mandatory childhood vaccination programme before its implementation, France, 2017. *Eurosurveillance* **24**, 1900053 (2019).
43. Raude, J. *et al*. Are people excessively pessimistic about the risk of coronavirus infection? Preprint at https://psyarxiv.com/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://www.medrxiv.org/content/10.1101/2020.09.01.20135194v2 (2020).
47. Luo, L. *et al*. Modes of contact and risk of transmission in COVID-19 among close contacts. Preprint at https://www.medrxiv.org/content/10.1101/2020.03.24.20042606v1 (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 http://medrxiv.org/lookup/doi/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://www.medrxiv.org/content/10.1101/2020.04.18.20071134v1 (2020).
54. INSEE.fr 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**, 1–15 (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); REACTing COVID-19 modeling and surveillance grants. We thank Pascal Crepey, Camille Pelat, Edouard Chatignoux, Juliette Paireau, Daniel Levy-Bruhl for useful discussions. We also thank all participants of COVIDnet.fr system.

Author contributions VC conceived and designed the study. CT, MD, CG, CKK, CS, TH, TB, SBS, VC developed and implemented the protocol of the participatory platform for COVID-19 surveillance. GP, LDD, CES, CT, JF, SV, CC, SBS collected the data. GP, LDD, CES, CT analyzed the data. GP, LDD, CES, PYB, VC developed the inference framework. GP, LDD developed the code. GP, LDD, CES, PYB performed the numerical simulations, and analyzed the results. GP, LDD, CES, EV, CT, MD, CG, CKK, CS, TH, TB, PYB, JF, SV, CC, SBS, VC interpreted the results. VC drafted the Article. All authors contributed to and approved the final version of the Article.

Competing interests The authors declare no competing interests.

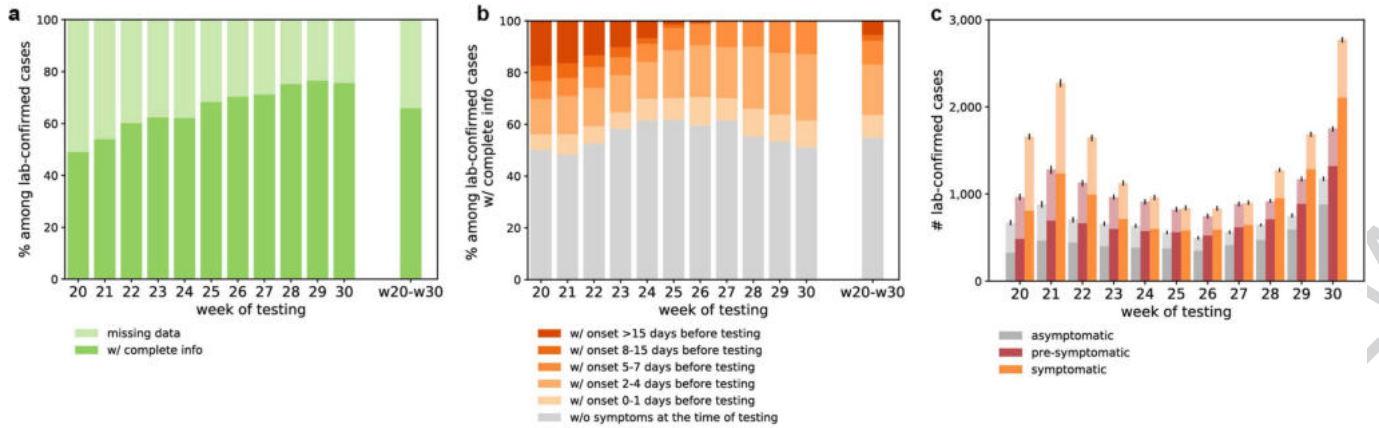
Additional information

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

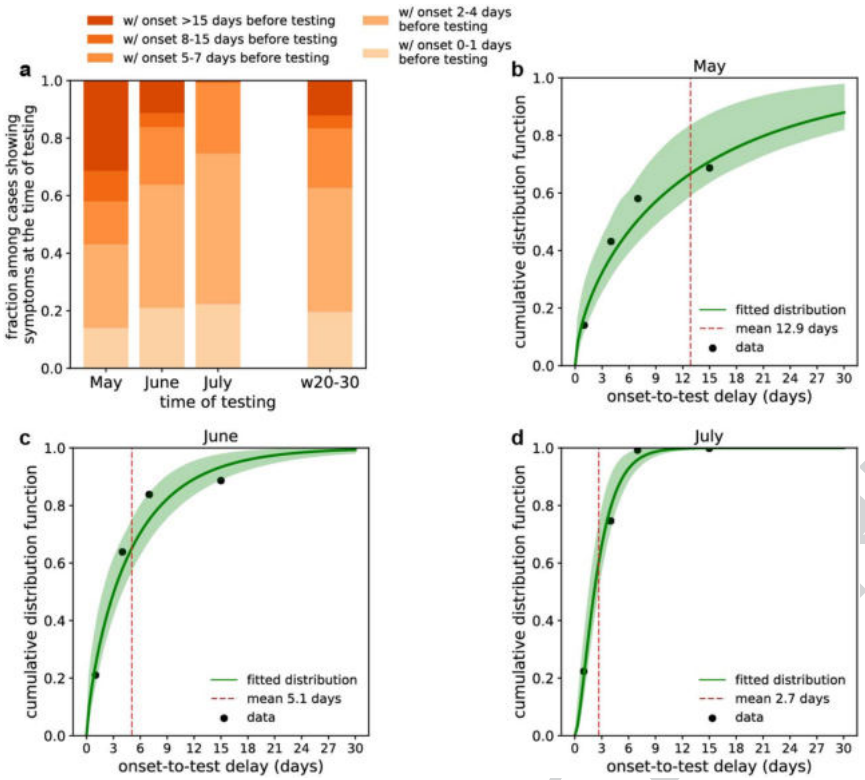
Correspondence and requests for materials should be addressed to V.C.

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.

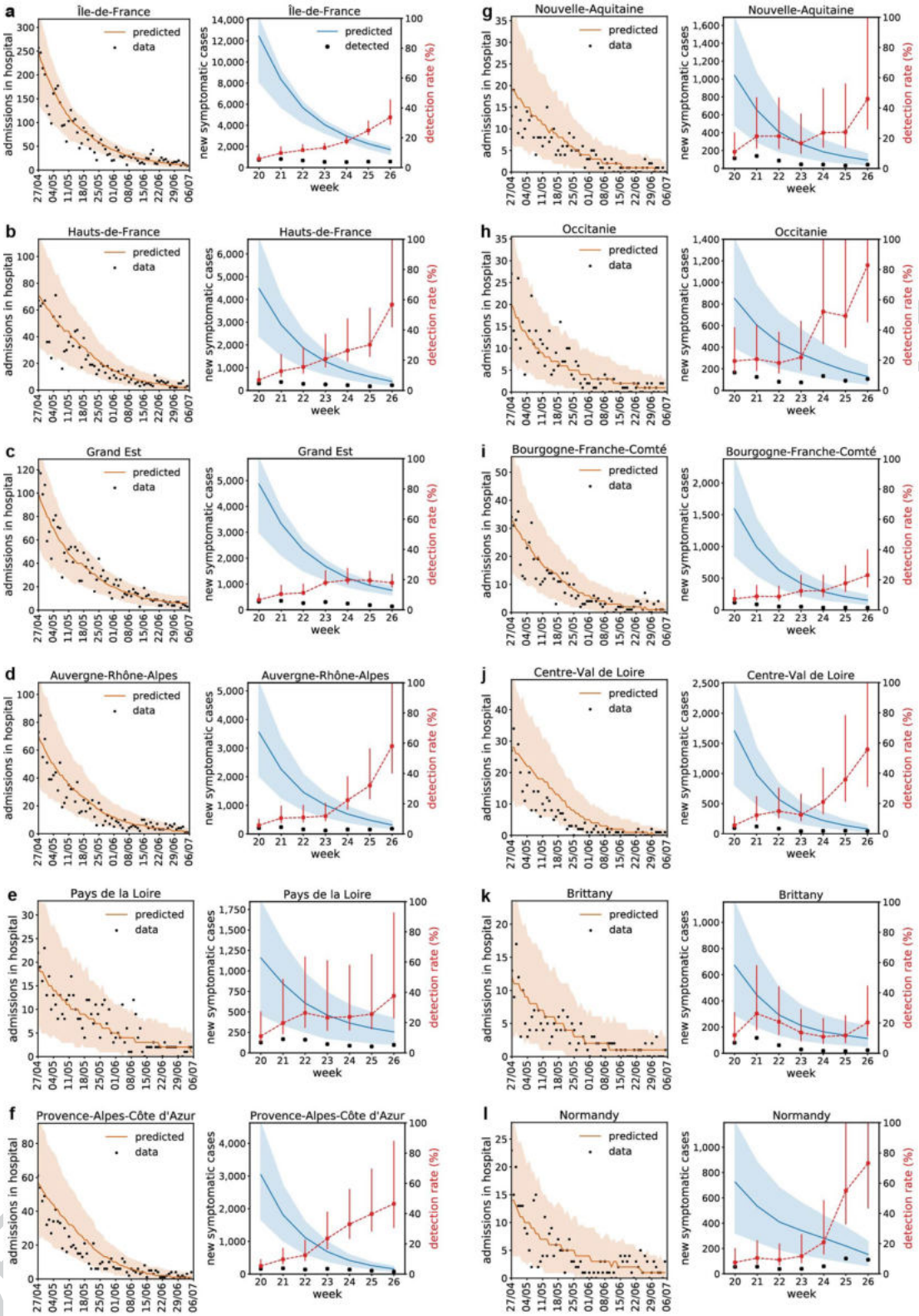
Reprints and permissions information is available at http://www.nature.com/reprints.



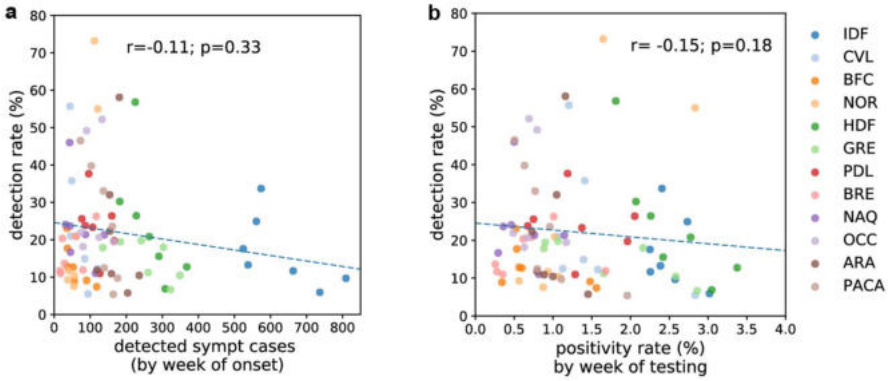
Extended Data Fig. 1 | Information provided by SI-DEP database and imputation of missing data. (a) Breakdown of laboratory-confirmed COVID-19 cases (n=35,264) according to information available. Missing data refer to cases for which information regarding absence/presence of symptoms at the time of testing and self-declared onset-to-test delay were not provided. (b) Breakdown of laboratory-confirmed COVID-19 cases with complete information (n=23,210) according to absence/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 colors indicate cases with complete information and missing data, respectively. 95% confidence intervals (black bars) are obtained by applying the imputation procedure 100 times.



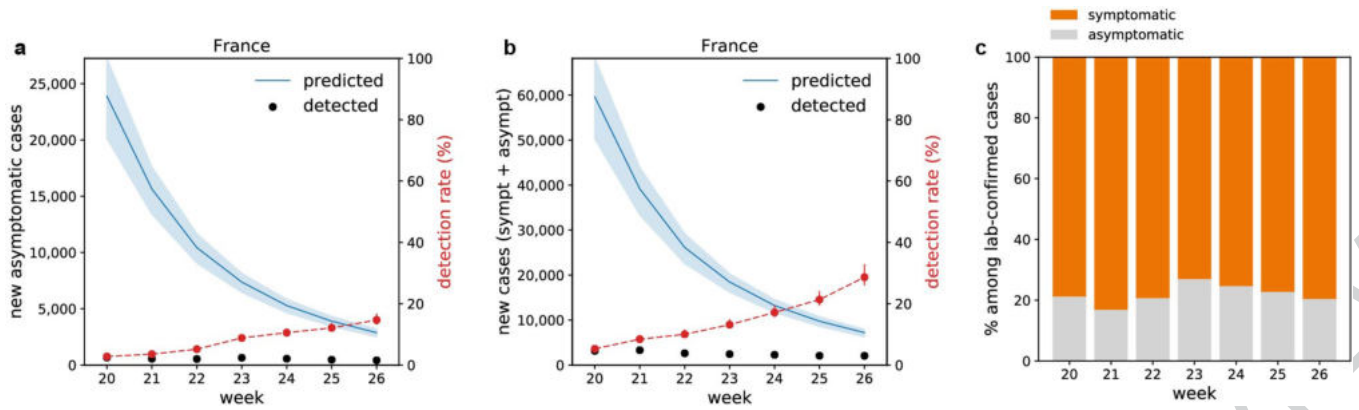
Extended Data Fig. 2 | Estimation of onset-to-test delay distribution. (a) Data used for the fit. Breakdown of laboratory-confirmed COVID-19 cases with symptoms (n=10,494) according to month of testing and the declared time-interval of onset-to-testing delay (b) Gamma distribution fitted through maximum likelihood to data referring to May (shape parameter 0.61 [0.42, 0.80] (median estimate and 95% Wald CI), 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%CI of the cumulative distribution function.



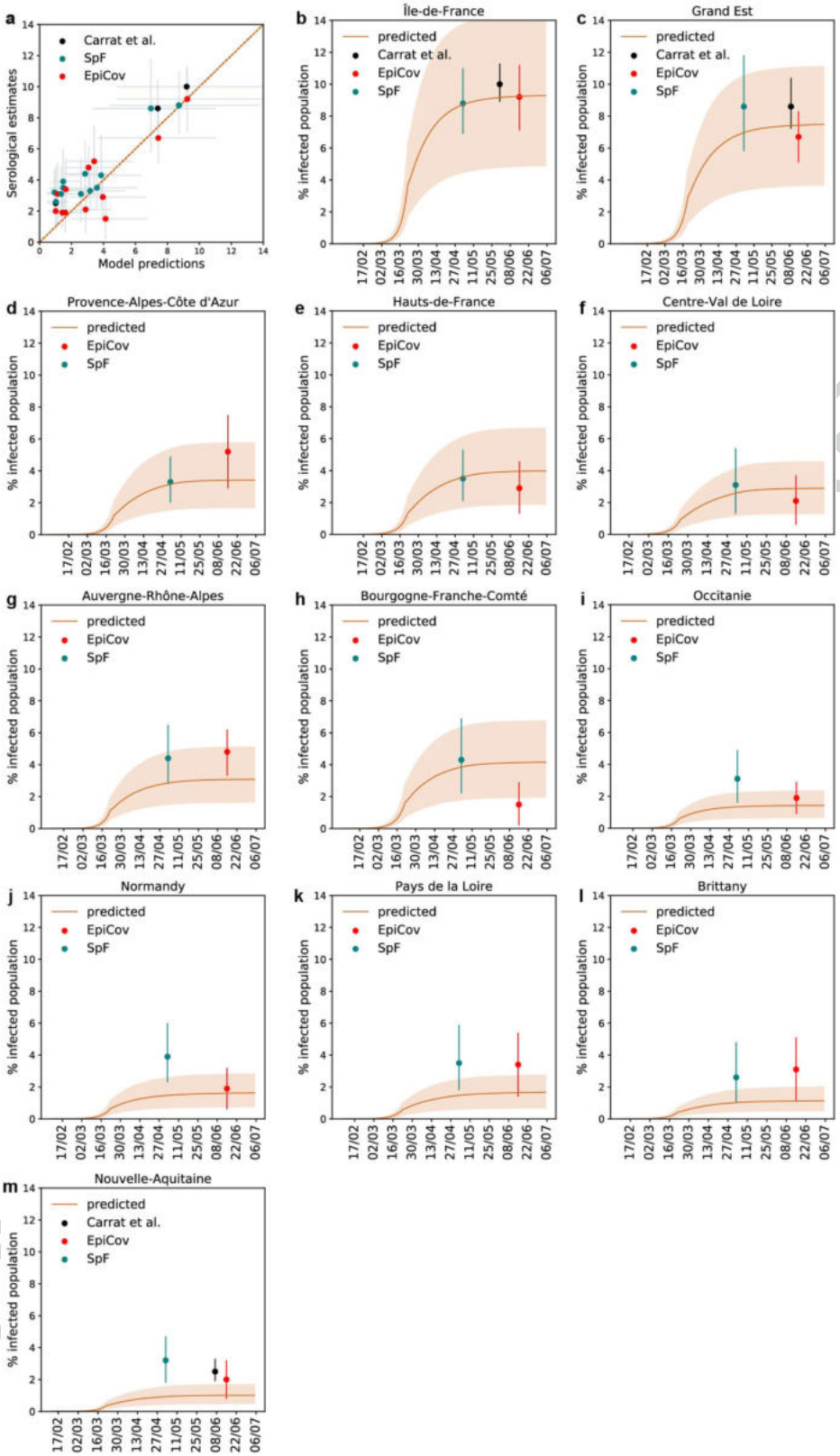
Extended Data Fig. 3 | Hospital admissions and number of new symptomatic cases. For each region, the pair of panels shows: (left) hospital admissions over time (data (points) and simulations-median and 95% confidence interval); (right) projected number of new symptomatic cases over time (median and 95% confidence interval) and estimated number of virologically-confirmed symptomatic cases by week of onset (points), detection rate (red points, right y axis). Median and 95%CI 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-Cote d'Azur (f), Nouvelle-Aquitaine (g), Occitanie (h), Bourgogne-Franche-Comté (i), Centre-Val de Loire (j), Brittany (k), Normandy (l). Hospital admission data up to w27 (consolidated in w28) were used to calibrate the models.



Extended Data Fig. 4 | Detection rate vs. indicators linked to testing policy/effectiveness. (a) Detection rate vs. number of symptomatic detected cases, by week and by region. (b) Detection rate vs. positivity rate by week of testing. Results of Spearman correlation test are provided (r=correlation coefficient, p=p-value).



Extended Data Fig. 5 | Estimates of 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%CI) in mainland France. Median and 95%CI are obtained from n=500 independent stochastic runs. (b) As panel (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 vs. serological estimates. (a) Serological estimates per region from three seroprevalence studies vs. corresponding model projections. Error bars correspond to 95%CI. (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 EpiCov (red dots), Santé Publique France (SpF) (green dots), and Carrat *et al.*[7], for the three regions where the study was conducted (black dots). Median and 95%CI for model projections are obtained from n=500 independent stochastic runs. Plots are reported for all 12 regions of mainland France : Ile-de-France (b), Grand Est (c), Provence-Alpes-Cote d'Azur (d), Hauts-de-France (e), Centre-Val de Loire (f), Auvergne-Rhone-Alpes (g), Bourgogne-Franche-Comte (h), Occitanie (i), Normandy (j), Pays de la Loire (k), Brittany (l), Nouvelle-Aquitaine (m).

Extended Data Table 1 | Population per region, number of virologically-confirmed symptomatic cases (w20-26), number of projected symptomatic cases (w20-26), estimated detection rate (w20-26)

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 May 11 to June 28, 2020. *Mainland France (Corsica and overseas territories excluded)

nature research

Corresponding author(s): Vittoria Colizza

Last updated by author(s): Nov 28, 2020

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ 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.