scientific reports

OPEN

The real seroprevalence of SARS-CoV-2 in France and its consequences for virus dynamics

Chloé Dimeglio[1,2]✉, Jean-Michel Loubes[3], Marcel Miedougé[2], Fabrice Herin[4], Jean-Marc Soulat[4] & Jacques Izopet[1,2]

The SARS-CoV-2 virus has spread world-wide since December 2019, killing more than 2.9 million of people. We have adapted a statistical model from the SIR epidemiological models to predict the spread of SARS-CoV-2 in France. Our model is based on several parameters and assumed a 4.2% seroprevalence in Occitania after the first lockdown. The recent use of serological tests to measure the effective seroprevalence of SARS-CoV-2 in the population of Occitania has led to a seroprevalence around 2.4%. This implies to review the parameters of our model to conclude at a lower than expected virus transmission rate, which may be due to infectivity varying with the patient's symptoms or to a constraint due to an uneven population geographical distribution.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that emerged in Wuhan, China in December 2019 has spread largely by sustained human-to-human transmission[1]. The WHO declared the resulting disease a pandemic[2]. The virus, which causes severe respiratory illness in susceptible individuals, has spread so rapidly that it has led most affected countries to lockdown their populations in order to break the dynamics of the virus's evolution. Many models have used the proportion of SARS-CoV-2 infected people tested by RT-PCR tests to predict the course of the epidemic[3–5]. Our statistical model for predicting the spread of SARS-CoV-2 in France is based on a diffusion and transmission coefficient that varies with an individual's age, the likelihood of contagion, and two administration parameters (lockdown and quarantine)[4]. It indicated that the seroprevalence in France at the end of the first lockdown (from March 17, 2020 to May 11, 2020) would be no greater than 17.5%[4]. Recent results of serological tests measuring the effective seroprevalence of SARS-CoV-2 in the general population have led us to assess the consequences of a much lower seroprevalence on the real virus dynamics in the population. The aim of this work is to determine the reason why there is a difference between the seroprevalence results obtained by the model and those obtained by the serological tests, and to modify the parameters of the model accordingly.

Results

We define $I_{d,R}$ as the real seroprevalence on day d at the end of lockdown and $I_{d,E}$ as the seroprevalence estimated by the model on the same day. More generally, for a given variable V, V_R is the observed variable and V_E is the variable estimated by the model.

We assumed that there were 3 untested SARS-CoV-2-contagious people in France when the first 3 cases were detected on January 24 (50% of asymptomatic cases). We assume that the number of tests carried out (α) has remain fixed, since the virology laboratories have never been saturated, and that the number of days a person is contagious N_T is equal to 20[6].

The quarantine constraint q is considered to be negligible.

Seroprevalence. Lockdown began on March 17 in France and ended on May 11. The predicting model indicated that the seroprevalence in Occitania before March 17 was 0%, and that it was 4.2% after lockdown. From June 10, 2020 to July 10, 2020, 1 to 2 months after the end of lockdown in France, everyone who worked at Toulouse University Hospital (n = 16,571) was invited for screening total serum antibodies by ELISA test. Among the 8758 people who took part in the study, the ELISA test indicated that 276 had SARS-CoV-2 antibodies, corresponding to an overall seroprevalence of 3.2% (95% confidence interval [CI] 2.8–3.5%)[7]. Based on the demography in the Occitania region, the median age was 44 years (IQR 30–57) for 58.2% of women

[1]2 INSERM UMR1291 – CNRS UMR5051, Toulouse Institute for Infectious and Inflammatory Diseases (INFINITy), 31300 Toulouse, France. [2]Virology Laboratory, CHU Toulouse, Hôpital Purpan, 31300 Toulouse, France. [3]Institut de Mathématiques de Toulouse, Université de Toulouse, 31400 Toulouse, France. [4]Occupational Diseases Department, Toulouse-Purpan University Hospital, 31000 Toulouse, France. ✉email: dimeglio.c@chu-toulouse.fr

and 42.8% of men. Corrections were performed using the R software/environment[8] version R-3.5.1. (Copyright 2007 Free Software Foundation, Inc; http://www.gnu.org/licenses/gpl.html). We have used the "sampling" package obtained from the Comprehensive R Archive Network (CRAN) mirror sites[9] to correct the estimated SARS-CoV-2 seroprevalence based on our dataset[10]. By correcting our seroprevalence data to align them with these regional demographic data and the geographical distribution of postcodes, we obtained a SARS-CoV-2 seroprevalence of 2.4%.

Consequently, it is quite possible that the national seroprevalence of SARS-CoV-2 is also lower than expected.

Consequences for SARS-CoV-2 dynamics. Since the real seroprevalence is lower than expected on day d at the end of lockdown, we have

$$I_{d,R} < I_{d,E}. \tag{1}$$

According to Eq. (4), the number of immunized people on day d corresponds to the number of people immunized on day $d-1$ to which we add the people who were on their last day of infection on day $d-1$, whether they were tested or not.

We have $I_{d,R} = (I_{d-1} + P_{d-1}^{20} + Q_{d-1}^{20})_R$, $I_{d-1,R} = (I_{d-2} + P_{d-2}^{20} + Q_{d-2}^{20})_R$ and finally, $I_{d,R} = (I_1 + \sum_{n=1}^{d-1} P_n^{20} + \sum_{n=1}^{d-1} Q_n^{20})_R$ with $I_1 = 0$ (nobody immunized on day one).

Equation (1) becomes $(\sum_{n=1}^{d-1} P_n^{20} + \sum_{n=1}^{d-1} Q_n^{20})_R < (\sum_{n=1}^{d-1} P_n^{20} + \sum_{n=1}^{d-1} Q_n^{20})_E$.

Since $\forall\, n < 20$, $P_n^{20} = 0$ and $Q_n^{20} = 0$, (1) becomes

$$\left(\sum_{n=20}^{d-1} P_n^{20} + \sum_{n=20}^{d-1} Q_n^{20}\right)_R < \left(\sum_{n=20}^{d-1} P_n^{20} + \sum_{n=20}^{d-1} Q_n^{20}\right)_E.$$

According to Eq. (2), the number of undetected contagious carriers on day n is the number of undetected carriers who were infected and undetected on day $n-1$ and who were not tested.

We have:

$$P_n^{20} = P_{n-1}^{19}(1-\alpha),$$
$$P_{n-1}^{19} = P_{n-2}^{18}(1-\alpha),$$
$$\vdots$$

Finally, $P_n^{20} = P_{n-19}^1 (1-\alpha)^{19}$ with P_{n-19}^1 defined by Eq. (3).

Similarly, Q_n^{20} is defined by Eq. (5) as the number of detected contagious carriers on day $n-1$ to which we add the number of undetected contagious carriers on day $n-1$ who were tested.

We have:

$$Q_n^{20} = Q_{n-1}^{19} + P_{n-1}^{19} \cdot \alpha \quad \text{with} \quad Q_{n-1}^{19} = Q_{n-2}^{18} + P_{n-2}^{18} \cdot \alpha,$$
$$Q_n^{20} = Q_{n-2}^{18} + P_{n-2}^{18} \cdot \alpha + P_{n-2}^{18} \cdot \alpha \cdot (1-\alpha)$$
$$\vdots$$

Finally, $Q_n^{20} = Q_{n-19}^1 + P_{n-19}^1 \cdot \sum_{m=0}^{18} \alpha \cdot (1-\alpha)^m$ with $Q_{n-19}^1 = 0$ (no quarantine on day one) and P_{n-19}^1 defined by Eq. (3).

Equation (1) becomes:

$$\left(\sum_{n=20}^{d-1} \frac{S_n}{N} \cdot \left[\sum_i P_{n-20}^i \cdot R_0^i \cdot c\right] \cdot \left[\sum_{m=0}^{18} \alpha \cdot (1-\alpha)^m + (1-\alpha)^{19}\right]\right)_R$$
$$< \left(\sum_{n=20}^{d-1} \frac{S_n}{N} \cdot \left[\sum_i P_{n-20}^i \cdot R_0^i \cdot c\right] \cdot \left[\sum_{m=0}^{18} \alpha \cdot (1-\alpha)^m + (1-\alpha)^{19}\right]\right)_E,$$

which equals:

$$\left(\sum_{n=20}^{d-1} \frac{S_n}{N} \cdot \left[\sum_i P_{n-20}^i \cdot R_0^i \cdot c\right]\right)_R < \left(\sum_{n=20}^{d-1} \frac{S_n}{N} \cdot \left[\sum_i P_{n-20}^i \cdot R_0^i \cdot c\right]\right)_E.$$

The three possible explanations for why the real seroprevalence is lower than expected are:

1. $c_R < c_E$ i.e. the real lockdown constraint c_R was more stringent than expected c_E.
2. $P_{n,R} < P_{n,E}$ i.e. the fraction of the population initially infected but not detected $P_{n,R}$ was smaller than expected $P_{n,E}$.
3. $(R_0^i)_R < (R_0^i)_E$ i.e. the transmission rate was lower than expected.

Discussion

We used a discretized SIR model to predict the dynamics of SARS-CoV-2 infections that takes into account various factors involved in the virus dynamics. We have assumed that the transmission of SARS-CoV-2 was identical regardless of the age of each individual or any individual characteristic. Although children could be less susceptible to SARS-CoV-2[11,12], there is limited published evidence of age-related differences in infectivity[13–15]. Moreover, there may be errors in ascertaining the direction of transmission, leading to confusing differences in infectiousness with differences in susceptibility. Lockdown began on 17 March in France, and was gradually relaxed from May 11. Under these circumstances, the seroprevalence in Occitania at the end of lockdown was estimated to be 17.5% in France and 4.2% in Occitania[4]. But the serological tests carried out since the end of May indicate that the real seroprevalence is lower which suggests that the national seroprevalence is also lower. This is consistent with the findings of a study indicating that 4.4% (range 2.8% to 7.2%) of the French population (2.8 million people) had been infected[16] and with the first results of the national study EpiCoV[17] that found 4.5% of the French people had been affected by SARS-CoV-2 at the end of the first lockdown (May 11).

This could be because the actual lockdown constraint c_R was more stringent than expected c_E. This seems unlikely since we have set the lockdown constraint at 80% and that it was neither strictly respected nor as restrictive as expected, given that approximately 29% of the French population continued to go to work during the first lockdown[18]. This mainly concerned healthcare workers who were required face-to-face for the management of the Covid-19 crisis and people who could not telework.

The second possibility is that the fraction of the population initially infected but not detected $P_{n,R}$ was smaller than expected, $P_{n,E}$, or that the asymptomatic people were less contagious than those who were symptomatic. The Center for Disease Control and Prevention (CDC) has reported that researchers in Singapore concluded that asymptomatic people were the source of 44% of diagnosed COVID-19 cases[19]. Another recent study reported that 104/128 (81 percent) of people on a cruise ship who tested positive for SARS-CoV-2 were asymptomatic[20]. And yet another study found that 42% of those who tested positive for SARS-CoV-2 were without symptoms[21]. On the contrary, we could also consider that the proportion of people infected but not detected was higher than that actually estimated due to the lack of sensitivity of the RT-PCR tests which can be negative for truly infected participants or early infected participants with waning immunity[22]. The proportion of asymptomatic cases in our model was close to 50% but it is possible that the actual proportion was lower even if the assumed proportion remains within the possible range. The expected proportion of asymptomatic people could be higher than that actually observed if these asymptomatic people were less contagious than expected. Our model assumes that the transmission rate can vary with the day of infection but is independent of the patient's symptoms. However, the proportion of asymptomatic infected people could also be lower than expected if these people were less contagious. This relationship was favored in a recent study looking at the symptoms of 455 cases who were in contact with a confirmed but asymptomatic SARS-CoV-2 patient[23].

Our last explanation for a lower than expected seroprevalence assumes that the transmission rate was lower than expected: $(R_0^i)_R < (R_0^i)_E$. The WHO set this rate at an average of 2.2[24], with occasional large variations. A recent review comparing 12 studies that estimated the R_0 for COVID-19 found values of from 1.5 to 6.68[25]. The 2.2 average rate may be less important in reality or it may vary according to an individual's symptoms, as previously discussed. It is also possible that the rate itself does not vary but that its constraints do. Our model assumes that the virus is uniformly distributed throughout the population, but the first episodes of virus infection occurred in clusters, which assumes concentrations of population in given places without necessarily any communication between the clusters. A recent analysis of the haplotypes of SARS-CoV-2 infections in Iceland concluded that the geographical distribution of clades was highly structured[26]. Thus the rate of virus transmission may be that set by the WHO within clusters but it could vary outside them or be much lower because of the uneven population density.

Serological testing has shown that the seroprevalence of SARS-CoV-2 is much lower than expected. In particular our prediction models assumed a 4.2% seroprevalence in Occitania and it turned out to be only 2.4%. This implies slower than expected virus transmission, perhaps due to variations in infectivity influenced by a patient's symptoms or to geographical constraints on the population distribution. It will probably be necessary in the future to include a population density parameter in the SARS-CoV-2 diffusion models, or to work at equal densities[27], to avoid bias.

Methods

Serological tests.   Total plasma anti-SARS-CoV-2 antibodies (Ab) were detected using an enzyme linked immunosorbent assay (ELISA) kit (Beijing Wantai Biological Pharmacy Entreprise Co., Ltd, China) according to the manufacturer's instructions. This assay is based on a recombinant antigen containing the SARS-CoV-2 spike protein receptor-binding domain (RBD). The sensitivity using samples collected 15 to 45 days post symptom-onset or after contact with a SARS-CoV-2 case, including asymptomatic patients, was 100% (CI 95% 88.2–100%) and the specificity using plasma collected before the outbreak of SARS-CoV-2 was 100% (CI 95% 82.1–100%)[28]. Human samples were taken after obtaining the informed consent for the SARS-CoV-2 surveillance. Biological material and clinical data were obtained only for standard viral surveillance (no specific sampling, no modification of the sampling protocol, no questions in addition to the standardized questionnaire). Data were analysed using an anonymized database. According to the Public Health French law (CSP Art L 1121-1.1), such a protocol does not require written informed consent. This study was entered in the Toulouse University Hospital register of retrospective studies: Rn IPH 2020-33 and is covered by MR-004 in accordance with the French data protection authority (Commission Nationale de l'Informatique et des Libertés—CNIL—number: 2206723 V0). All methods were carried out in accordance with relevant guidelines and regulations.

Statistical model. The model is a discretized version of a Susceptible Infectious and Recovered (SIR)-type model[29]. We have added a diffusion/transmission coefficient R_0 in our model that varies with the likelihood of contagion, and two reduction coefficients c and q to describe the impact of public health measures on virus transmission in France.

We estimated the initial model settings using data published by Johns Hopkins University for France (Johns Hopkins database: https://github.com/CSSEGISandData/COVID-19 verified on 03/28/2020.) and data collected by the Toulouse Virology Laboratory. The total French population is N = 67,000,000 (source INSEE). SARS-CoV-2 screening began on January 24, 2020 and containment on March 17. We assumed that there were 50 untested SARS-CoV-2-contagious people in France on January 25, and 3 tested cases (source Johns Hopkins). Some values of c and the value of q have been estimated in previous works by correcting the values predicted by the model using real data collected by the Toulouse Virology Laboratory[4,27]. The model predicts how the SARS-CoV-2 virus would have evolved and projects the daily number of new positive cases. Its accuracy was evaluated at 86.1%[30]. By cumulative effect, we therefore obtain a projection of the seroprevalence of SARS-CoV-2 in France.

We considered the variables (S_n, P_n, Q_n, I_n).

S_n is the number of healthy people on day n, P_n^i is the number of undetected contagious carriers infected for i days ($1 \leq i \leq N_T$). Similarly, Q_n^i is the number of detected contagious carriers infected for i days ($1 \leq i \leq N_T$) on day n, and I_n is the number of people who were immunized. We assume that the risk of reinfection by SARS-CoV-2 after a first infection is negligible.

N_T is the number of days a person is contagious and α is the percentage of the population tested on each day. R_0 is the number of healthy people who a contagious person contacts and infects We assume that R_0 varies over time and peaks when the virus load is maximal: 7 days after the start of infection[31,32]. In the absence of a consensus, we assume that the number of days a person is contagious is equal to the time of infection i.e. 20 days[32,33]. For all $1 \leq i \leq N_T$, R_0^i is:

$$R_0^i = R_0 \cdot e^{-\frac{1}{2}\left(\frac{i-7}{\sqrt{20}}\right)^2}.$$

N is the total population at the start of the epidemic phase, c is the multiplier for the pace of the epidemic throughout lockdown ($0 \leq c \leq 1$), and q is the same multiplier during the quarantine period ($0 \leq q \leq 1$). c and q are set at 1 when there is no lockdown or quarantine. The lower the values of c or q, the greater the constraint which applies to the spread of the virus.

N is given by:

$$N = S_n + P_n + Q_n + I_n.$$

On transition from day n to day $n + 1$, we have:

$$\forall\, 1 \leq i \leq N_T - 1,\ P_{n+1}^{i+1} = P_n^i\,(1-\alpha), \tag{2}$$

$$P_{n+1}^1 = \frac{S_n}{N} \cdot \left[\sum_i P_n^i \cdot R_0^i \cdot c + \sum_i Q_n^i \cdot R_0^i \cdot q\right], \tag{3}$$

$$I_{n+1} = I_n + P_n^{N_T} + Q_n^{N_T}. \tag{4}$$

$Q_{n+1}^1 = 0$ (no quarantine on day one, test results needed).

$$\forall\, 1 \leq i \leq N_T - 1,\ Q_{n+1}^{i+1} = Q_n^i + P_n^i \cdot \alpha. \tag{5}$$

We set $R_0 = 2.2$[24] at its peak.

References

1. Perlman, S. Another decade, another coronavirus. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMe2001126 (2020).
2. WHO. *Virtual Press Conference on COVID-19—11 March 2020* (WHO, 2020).
3. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
4. Dimeglio, C. *et al.* The SARS-CoV-2 seroprevalence is the key factor for deconfinement in France. *J. Infect.* **81**, 318 (2020).
5. Wu, J. T. & Leung, K. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: A modelling study. *Lancet* **395**, 689 (2020).
6. Chang, M. G. *et al.* Time kinetics of viral clearance and resolution of symptoms in novel coronavirus infection. *Am. J. Res. Crit. Care Med.* **201**, 1150 (2020).
7. Dimeglio, C. *et al.* Screening for SARS-CoV-2 antibodies among healthcare workers in a university hospital in southern France. *J. Infect.* https://doi.org/10.1016/j.jinf.2020.09.035 (2020).
8. R Development Core Team. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.r-project.org (Accessed 15 April 2021).
9. http://cran.r-project.org/bin/windows/base/ (Accessed 15 April 2021).
10. Tillé, Y. & Matei, A. *R Package 'Sampling' Version 2.8 (Manual)* (2016). R Foundation for Statistical Computing.
11. Dimeglio, C., Mansuy, J. M., Charpentier, S., Claudet, I. & Izopet, J. Children are protected against SARS-CoV-2 infection. *J. Clin. Virol.* **128**, 104451. https://doi.org/10.1016/j.jcv.2020.104451 (2020).

12. Viner, R. M. *et al.* Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: A systematic review and meta-analysis. *JAMA Pediatr.* https://doi.org/10.1001/jamapediatrics.2020.4573 (2020).
13. Dattner, I. *et al.* The role of children in the spread of COVID-19: Using household data from Bnei Brak, Israel, to estimate the relative susceptibility and infectivity of children. *MedRxiv* **69**, 422. https://doi.org/10.1101/2020.06.03.20121145v1 (2020).
14. National Institute for Public Health and the Environment (RIVM), the Netherlands. *Children and COVID-19* (2020). https://www.rivm.nl/en/coronavirus-covid-19/children-and-covid-19 (Accessed 8 June 2021).
15. Park, Y. J. *et al.* Contact tracing during coronavirus disease outbreak, South Korea, 2020. *Emerg. Infect. Dis.* **26**(10), 2465–2468 (2020).
16. Salje, H. *et al.* Estimating the burden of SARS-CoV-2 in France. *Science* **9**, 538 (2020).
17. INSERM Press Release. (2020). https://presse.inserm.fr/premiers-resultats-des-enquetes-de-sante-publique-de-linserm-sur-la-covid-19-facteurs-de-risque-individuels-et-sociaux/41032/ (Accessed 8 June 2021).
18. Massimo, F. S. Impact of the Covid-19 confinement measures on telework in France. A qualitative survey. In *EUR—Scientific and Technical Research Reports* (2020). https://ec.europa.eu/jrc/sites/jrcsh/files/jrc122669.pdf (Accessed 9 April 2021).
19. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
20. Ing, A. J., Cocks, C. & Green, J. P. COVID-19: In the footsteps of Ernest Shackleton. *Thorax* **75**, 693. https://doi.org/10.1136/thoraxjnl-2020-215091 (2020).
21. Yang, R., Gui, X. & Xiong, Y. Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in Wuhan, China. *JAMA Netw. Open* **3**(5), e2010182 (2020).
22. Carrat, F. *et al.* Evidence of early circulation of SARS-CoV-2 in France: Findings from the population-based "CONSTANCES" cohort. *Eur. J. Epidemiol.* **36**(2), 219–222. https://doi.org/10.1007/s10654-020-00716-2 (2021).
23. Gao, M. *et al.* A study on infectivity of asymptomatic SARS-CoV-2 carriers. *Respir. Med.* **169**, 106026 (2020).
24. WHO. *Report of the WHO-China Joint Mission on Coronavirus Disease (COVID-19)* (WHO, 2019).
25. Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* https://doi.org/10.1093/jtm/taaa021 (2020).
26. Gudbjartsson, D. F. *et al.* Spread of 297 SARS-CoV-2 in the icelandic population. *N. Engl. J. Med.* **382**, 2302 (2020).
27. Dimeglio, C., Loubes, J. M., Mansuy, J. M. & Izopet, J. Quantifying the impact of public health protection measures on the spread of SARS-CoV-2. *J. Infect.* https://doi.org/10.1016/j.jinf.2020.10.026 (2020).
28. Abravanel, F. *et al.* Clinical performance of a rapid test compared to a microplate test to detect total anti SARS-CoV-2 antibodies directed to the spike protein. *J. Clin. Virol.* **130**, 104528 (2020).
29. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721 (1927).
30. Dimeglio, C., Miedougé, M., Loubes, J. M., Mansuy, J. M. & Izopet, J. Estimating the impact of public health strategies on the spread of SARS-CoV-2: Epidemiological modelling for Toulouse, France. *Rev. Med. Virol.* https://doi.org/10.1002/rmv.2224 (2021).
31. Walsh, K. A. *et al.* SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *J. Infect.* **81**(3), 357–371. https://doi.org/10.1016/j.jinf.2020.06.067 (2020).
32. Lescure, F. X., *et al.* Clinical and virological data of the first cases of COVID-19 in Europe: A case series. *Lancet Infect. Dis.* **20**, 697–706. https://doi.org/10.1016/S1473-3099(20)30200-0 (2020). Erratum in: Lancet. Infect. Dis. 2020: Erratum in: Lancet Infect Dis. 2020;20(6):e116.
33. Jaafar, R. *et al.* Correlation between 3790 qPCR positives samples and positive cell cultures including 1941 SARS-CoV-2 isolates. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciaa1491 (1941).

Author contributions
M.M., F.H. and J.M.S. collected the data, C.D. and J.M.L. created the probabilistic model, C.D. and J.I. wrote the manuscript, all authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to C.D.

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

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