Challenges in control of Covid-19: short doubling time and long delay to effect of interventions†

Lorenzo Pellis[1]*, Francesca Scarabel[2,3], Helena B. Stage[1], Christopher E. Overton[1], Lauren H. K. Chappell[4], Katrina A. Lythgoe[5], Elizabeth Fearon[6], Emma Bennett[7], Jacob Curran-Sebastian[1], Rajenki Das[1], Martyn Fyles[1], Hugo Lewkowicz[1], Xiaoxi Pang[1], Bindu Vekaria[1], Luke Webb[1], Thomas House[1,8], Ian Hall[1,7]

[1]Department of Mathematics, The University of Manchester, UK.

[2]LIAM - Laboratory of Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto Ontario, Canada.

[3]CDLab - Computational Dynamics Laboratory, Department of Mathematics, Computer Science and Physics, University of Udine, Italy.

[4]Department of Plant Sciences, University of Oxford, UK.

[5]Big Data Institute, University of Oxford, UK.

[6]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, UK.

[7]Emergency Response Department, Public Health England, UK.

[8]IBM Research, Hartree Centre, SciTech Daresbury, Warrington, UK.

*Correspondence to: lorenzo.pellis@manchester.ac.uk.

†This preprint was first submitted to arXiv on 31st March 2020 at: arXiv:2004.00117v1

Abstract

Early assessments of the spreading rate of COVID-19 were subject to significant uncertainty, as expected with limited data and difficulties in case ascertainment, but more reliable inferences can now be made. Here, we estimate from European data that COVID-19 cases are expected to double initially every three days, until social distancing interventions slow this growth, and that the impact of such measures is typically only seen nine days - i.e. three doubling times - after their implementation. We argue that such temporal patterns are more critical than precise estimates of the basic reproduction number for initiating interventions. This observation has particular implications for the low- and middle-income countries currently in the early stages of their local epidemics.

1. Introduction

In December 2019, a cluster of unexplained pneumonia cases in Wuhan, the capital of Hubei province in the People's Republic of China, rapidly progressed into a large-scale outbreak, and a global pandemic by 11 March 2020, as declared by the World Health Organisation [*1*]. The disease caused by this highly contagious infection has since been named COVID-19, and is caused by a single-stranded RNA coronavirus (SARS-CoV-2) similar to the pathogen responsible for severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) [*2*]. As of 29 March 2020, 657,140 confirmed cases and 29,957 deaths have been reported in nearly 200 countries and territories globally [*3*].

Various control measures have been implemented worldwide, including isolation of confirmed and suspected cases, contact tracing, and physical distancing. However, only the most aggressive measures have resulted in epidemic suppression. In Hubei, a regional lockdown was implemented on 23-24 January 2020, with a peak in reported cases occurring approximately two weeks later [*4*]. In Italy, a national lockdown was implemented on 9 March 2020 once 7,375 confirmed cases and 366 deaths had been recorded; only in the last few days (as of 29 March) the epidemic seems to be slowing down [*5*]. In comparison, India declared a nationwide lockdown on 24 March 2020, with only 434 confirmed cases and 0 deaths [*6, 7*]. Similarly, South Africa began a 21-day lockdown on 27th March, with 927 known cases and 0 deaths [*8, 9*]. The implementation of such early and aggressive control measures in India and South Africa may substantially increase their chances of successful containment, although the social and personal cost could be substantial [*10*].

Planning of interventions usually relies on estimates of the basic reproduction number R_0, defined as the average number of new infections generated by a single infected person within a susceptible population. Reported estimates of COVID-19 R_0 are highly variable, ranging from 1.4 to 6.49 [*2, 11*], with the differences attributable to the variety of methods, model structures and parameter values (in particular, the estimated or assumed amount of pre-symptomatic transmission), as well as the data sources used. Most official sources settle in the range of 2-3 [*3, 12-15*], but these estimates mostly derive from early studies of the epidemic in Wuhan [*16-18*], or the Diamond Princess Cruise ship [*19*], and so are subject to important limitations: these include small amounts of data, uncertain or biased reporting of early cases in Wuhan, and the uniqueness of the specific settings in which they occurred.

We argue official ranges of R_0 should be continuously updated with more recent estimates coming not only from China [*20, 21*], but also from the many different outbreaks observed worldwide [*22-25*]. Point estimates might not change, but the task remains imperative both because available data is now more numerous and reliable, and because estimates of R_0 in one population do not necessarily translate to another.

However, we emphasise that the speed of epidemic growth and the delay between infection and case detection are more relevant for COVID-19 control implementation than estimates of R_0 [*26, 27*]. Since doubling times can be estimated directly from data and estimates of delays are relatively consistent, sophisticated models are not required to infer when action is urgent.

2. Results

To support our argument, we first estimate the growth rate in multiple countries and with different methods. We then estimate the incubation period and the distribution of times from symptoms onset to hospitalisation in different settings and compare it with other results from the literature. Observation of the outbreak in the UK and Italy provides further support to the

highlighted delays from infection to detection. Finally, we discuss the limitations of relying on R_0 in this context.

For the estimation of the growth rate, we focus on the number of confirmed cases in European countries that have experienced large local epidemics (Figure 1A), as reported by the WHO [*28*]. To avoid relying only on case confirmation, which could be affected by numerous biases, we also estimate the growth rate in hospitalisations, intensive care unit (ICU) admissions and deaths in Italy (Figure 1B; [*5*]). To ensure generalisability of results, we performed another analysis on a larger set of European countries (Figure 2). With the exception of a few countries with a small number of cases and potentially unreliable data, we consistently find doubling times of about 3 days or less, that appear to be sustained before mitigation interventions are put in place (Figure 2). These are significantly shorter than early estimates from China [*16, 29*]. For robustness, we have used two different methods: semi-parametric and generalised linear (details in Materials and Methods). Unsurprisingly, the results differ in terms of their confidence intervals, but the conclusions are similar, and are in agreement with the common exploratory analysis based on visually inspecting data plotted on a logarithmic scale (Figure S3).

Although our results are robust to the method used, they might still be misleading if there are biases in the data, such as errors in reporting, changes in case definition or testing regime, and so on. This issue is particularly critical as lags in reporting of cases can create discrepancies between national and international official sources [*5, 28, 30*] for case counts. However, these are unlikely to affect our conclusions owing to the following considerations:

- The fast growth and high numbers likely make small biases negligible;
- Any multiplicative correction, such as constant underreporting, will not affect the observed trend;
- It is relatively easy for exponential growth to appear slower than reality, for example if reporting rates decline over time, but difficult for growth to appear consistently faster: aggressive swabbing of asymptomatic individuals (e.g. early on in the Italian locked-down towns; [*31*]), as well as changes in case definition [*32*], might explain such a bias in the data, but these factors are unlikely to affect observations for longer than a few days or consistently across different countries;
- Hospitalisation and ICU admission, which in the Italian data appear to grow at similar rates as the number of confirmed cases (Figure 1B), should be much less affected by reporting issues. The even faster increase in death rates we observe, instead, may be explained by rapid outbreaks among vulnerable groups, such as those in care homes, coupled with quicker progression to death among these groups, or possibly local hospital saturation.

We conclude that, although existing data has its limitations, the evidence for fast exponential growth in the absence of intervention is overwhelming.

The delay between infection and case detection is crucial in determining how long cases have been growing unobserved. Early detection of cases during the incubation period, before individuals have become symptomatic, is typically not possible once containment has failed since it relies on full contact tracing and testing of asymptomatic individuals. Detecting cases at symptom onset is potentially more feasible, but in practice depends on the case-finding strategy. For example, in the UK, symptomatic individuals are instructed to self-isolate at home, and are only tested if they subsequently require hospitalisation. Thus the delay between infection and case detection includes the incubation period, the time between symptom onset

and hospitalisation, and the time it takes to receive a positive test result. Similar effects will be visible in other countries where case counts are dominated by hospitalisations.

We report published estimates of the incubation period and the delay between symptom onset and hospitalisation (Table 1). Since none of these estimates simultaneously account for truncated observations and exponential growth in the number of infected cases, we also include our own estimates (see Materials and Methods) obtained by analysing UK line-list data provided by Public Health England (unfortunately not publicly available) and a publicly available line-list which collates worldwide data [*33*]. Although more robust, our estimates are consistent with the existing literature, and highlight geographical heterogeneity, such as shorter onset-to-hospitalisation intervals in Hong Kong and Singapore compared to the UK. With the exception of Singapore, the sum of the mean incubation period and mean onset-to-hospitalisation interval is never shorter than 9 days, which corresponds to approximately 3 doubling times in an unconstrained epidemic like those observed in Figures 1 and 2.

Our estimate of the delay between infection and detection is consistent with observations of the UK and Italian epidemics (Figure 3). For both countries, we plot the numbers of new cases and notice a visible drop occurring 8-9 days after the first, relatively soft, control measures were implemented (in the UK, recommended self-isolation if symptomatic on 13 March, and in Italy, lockdown of infected towns and school and university closure in Northern Italy, on 22-23 February). After the first control measure in the UK (Figure 3A), cases continued to increase exponentially with an estimated growth rate of ~0.22/day (corresponding to a doubling time of just over 3 days) for 9 days. During this period numbers of daily confirmed cases rose approximately 8-fold. Subsequently, the number of new infections started to tail off. A similar pattern is observed in Italy (Figure 3B). Because non-pharmaceutical interventions, with unknown compliance, cannot be evaluated until their effects emerge in the data, a pattern of introduction of increasingly strong measures has repeated across Europe, with long delays to control. Even with immediate hard interventions halting all community transmission, within-household transmission will continue to occur, creating an additional delay between the beginning of the intervention and its effect. This is consistent with the approximate 2-week delay from lockdown to peak in new cases observed in Hubei [*4*]. Further delays in case-confirmations, hospitalisations, potential ICU admissions and deaths mean the latter figures keep on growing well after transmission control is achieved.

Since R_0 remains the mainstay of most epidemiological analyses, we explored values of R_0 consistent with a range of growth rates and modelling assumptions. For a growth rate of 0.25/day and our delay estimates (Table 1), we obtain values ranging from 2 to 4 or larger (Table S1A), owing to the extreme sensitivity to assumptions, in particular the extent of pre-symptomatic transmission, for which estimates in the literature vary widely [*34-37*].

Although R_0 is commonly used to determine the reduction in person-to-person transmission needed to achieve control, this brings limited insight into how to control the COVID-19 pandemic. First, there is sufficient evidence worldwide that interventions must be draconian, particularly in countries where case numbers are high. Second, given the variability in R_0 estimates obtained from the same growth rate appear predominantly due to the assumed amount of pre-symptomatic transmission, the exact value of R_0 is only poorly correlated with the required aggressiveness of the intervention [*38*]. For example, if R_0 is four and most of the transmission occurs after symptoms onset, 75% of the transmission needs to be stopped and self-isolation when symptomatic can easily achieve this. Conversely, if R_0 is two but most of the transmission is pre-symptomatic, only 50% of the transmission needs to be prevented, but

in practice this can only be achieved through interventions, like quarantining apparently healthy individuals, that are highly socially disruptive and likely enforced rather than spontaneous. Finally, R_0 informs how aggressive interventions should be, but not how quickly they should be implemented.

3. Discussion

The highlighted risks of underestimating the combination of short doubling times and long delays between infection and case detection are consistent with the now-common pattern of countries misjudging the initial small number of observed cases, only to realise the storm has already arrived. At unconstrained growth, even the immense effort of doubling local hospital capacity only buys 3 days of reprieve before bed capacity is breached. Being blind to the extent of an epidemic and the true number of infections at any one time results in intervention strategies based on the number of observed cases and measurements of R_0. For COVID-19, these are insufficient and dangerously underestimate the true degree of intervention required to slow down and bring the epidemic under control. We advocate stronger action from national and international health care communities, with a particular focus on supporting low- and middle-income countries where numbers of cases, at the time of writing, appear to be relatively low. In settings where health care capacity is low and intergenerational mixing common, swift action will save numerous lives.

Acknowledgments: We acknowledge Alex Washburne and his collaborators, who are reaching similar conclusions from different directions, for valuable discussions. The initial motivation for this study comes from the open letter from Italy to the scientific community [*39*], but we thank the numerous members of the general public, especially from hard-hit countries, who have tried to voice these messages with whatever means they had.

Funding: LP, HS and CO are funded by the Wellcome Trust and the Royal Society (grant 202562/Z/16/Z). KL is funded by the Wellcome Trust and Royal Society (107652/Z/15/Z) . FS is funded by the CIHR 2019 Novel Coronavirus (COVID-19) rapid research program. LC is funded by the BBSRC (grant BB/R009236/1). EF is funded by the MRC (grant MR/S020462/1). TH is supported by the Royal Society (Grant Number INF\R2\180067) and Alan Turing Institute for Data Science and Artificial Intelligence. IH is supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emergency Preparedness and Response and the National Institute for Health Research Policy Research Programme in Operational Research (OPERA) and Alan Turing Institute for Data Science and Artificial Intelligence. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Author contributions: EB, IH, and FS collated epidemiological data. IH, TH, CO, and LP carried out the formal data analysis. All authors were involved in the writing of the manuscript.

Competing interests: Authors declare no competing interests.

Ethical approval: No ethical approval was required for this work because the data used is either publicly available or does not allow individual subjects to be identified.

Data and materials availability: All data and code used in this analysis is provided in the Github repository (https://github.com/thomasallanhouse/covid19-growth), with the exception of the UK onset-to-hospitalisation delay data. This data is provided by Public Health England under a data sharing agreement, and we are unable to share this data.

Supplementary Materials: Details regarding the methods used in this analysis are provided in the supplementary material.

References:

1. World Health Organization (WHO), WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020, *WHO* (2020) https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
2. Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, 2 (2020).
3. World Health Organisation (WHO). Coronavirus disease (COVID-2019) situation report 69. *WHO* (2020) https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200329-sitrep-69-covid-19.pdf?sfvrsn=8d6620fa_2 . Accessed March 29, 2020.
4. World Health Organization (WHO) and the Chinese Centre for Disease Control and Prevention, Report of the WHO-China joint mission on Coronavirus disease 2019 (COVID-19), *WHO* (2020) https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.
5. Istituto Superiore di Sanità, Daily Reports, (29.03.2020) http://www.salute.gov.it/portale/news/p3_2_1.jsp?lingua=italiano&menu=notizie&area=nuovoCoronavirus&notizie.page=0 .
6. BBC News, Coronavirus: India enters 'total lockdown' after spike in cases (25.03.2020) https://www.bbc.co.uk/news/world-asia-india-52024239.
7. World Health Organization (WHO) Coronavirus 2019 (COVID-19) Situation report 64, WHO (2020) https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200324-sitrep-64-covid-19.pdf?sfvrsn=723b221e_2.
8. BBC News, South Africa reports first deaths as lockdown begins (27.03.2020) https://www.bbc.co.uk/news/world-africa-52058717.
9. World Health Organization (WHO) Coronavirus 2019 (COVID-19) Situation report - 67, *WHO* (2020) https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200327-sitrep-67-covid-19.pdf?sfvrsn=b65f68eb_4.
10. BBC News, Coronavirus: India's PM Modi seeks 'forgiveness' over lockdown (29.03.2020) https://www.bbc.co.uk/news/world-asia-india-52081396.
11. M. Majumder, K. D. Mandl, Early in the epidemic: impact of preprints on global discourse of 2019-nCOV transmissibility. *SSRN eLibrary*, 1–4 (2020).
12. European Centre for Disease Prevention and Control, Novel coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU / EEA and the UK – sixth update – 12 March 20202. Stockholm: *ECDC*; 2020.
13. Chinese Centre for Disease Control and Prevention, Epidemic update and risk assessment of 2019 Novel Coronavirus. *CCDC*, Published 28.01.2020. http://www.chinacdc.cn/yyrdgz/202001/P020200128523354919292.pdf [easiest accessed via URL].
14. E. Mahase, China coronavirus: what do we know so far? *BMJ*, **368** (2020) doi: https://doi.org/10.1136/bmj.m308.

15. World Health Organization (WHO), Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). *WHO* (2020) https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).

16. J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. **395**, 689–697 (2020).

17. J. Riou, C. L. Althaus, Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill.* **25**, 1–5 (2020).

18. N. Imai, A. Cori, M. Baguelin, C. A. Connelly, S. Riley, N. M. Ferguson, Report 3: Transmissibility of 2019-nCoV. *Imperial College COVID-19 Response Team* (2020) Publication 25.01.2020; https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-transmissibility-25-01-2020.pdf [Easiest access via URL].

19. S. Zhang, M. Y. Diao, W. Yu, L. Pei, Z. Lin, D. Chen, Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int. J. Infect. Dis.* **93**, 201–204 (2020).

20. B. Tang, F. Xia, N. L. Bragazzi, X. Wang, S. He, X. Sun, S. Tang, Y. Xiao, J. Wu, Lessons drawn from China and South Korea from managing COVID-19 epidemic: insights from a comparative modelling study. *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.09.20033464 13 March 2020.

21. K. Muniz-Rodriguez, G. Chowell, C.-H. Cheung, D. Jia, P.-Y. Lai, Y. Lee, M. Liu, S. K. Ofori, K. M. Roosa, L. Simonsen, C. Viboud, I. C.-H. Fung, Epidemic doubling time of the COVID-19 epidemic by Chinese province. *medRxiv* [Preprint] https://doi.org/10.1101/2020.02.05.20020750 28 February 2020.

22. A. Remuzzi, G. Remuzzi, COVID-19 and Italy: what next? *Lancet*, in press (2020) doi.org/10.1016/S0140-6736(20)30627-9.

23. Z. Zhuang, S. Zhao, Q. Lin, P. Cao, Y. Lou, L. Yang, S, Yang, D. He, L. Xiao, Preliminary estimating the reproduction number of the coronavirus disease (COVID-19) outbreak in the Republic of Korea and Italy by 5 March 2020. *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.02.20030312. 10 March 2020.

24. L. Russo, C. Anastassopoulou, A. Tsakris, G. N. Bifulco, E. F. Campana, G. Toraldo, C. Siettos, Tracing DAY-ZERO and forecasting the fade out of the COVID-19 outbreak in Lombardy, Italy: a compartmental modelling and numerical optimization approach. *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.17.20037689. 27 March 2020.

25. A. Ahmadi, Y. Fadaei, M. Shirani, F. Rahmani, Modeling and forecasting trend of COVID-19 epidemic in Iran. *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.17.20037671. 27 March 2020.

26. A. Weber, Trend analysis of the COVID-19 pandemic in China and the rest of the world, *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.19.20037192 23 March 2020.

27. F. Scarabel, L. Pellis, N. L. Bragazzi, J. Wu, Canada Needs to Rapidly Escalate Public Health Interventions for Its COVID-19 Mitigation Strategies. *SSRN*, in press, http://dx.doi.org/10.2139/ssrn.3559929.

28. World Health Organization (WHO), Coronavirus disease (COVID-2019) Situation Reports. *WHO* (2020) https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

29. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, **382** (13) 1199-1207 (2020).

30. European Centre for Disease Prevention and Control (ECDC), Data on the geographic distribution of COVID-19 cases worldwide, ECDC (2020) https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.

31. A. Crisanti, A. Cassone, In one Italian town, we showed mass testing could eradicate the coronavirus. *The Guardian*, 20.03.2020. https://www.theguardian.com/commentisfree/2020/mar/20/eradicated-coronavirus-mass-testing-covid-19-italy-vo [Easiest access via URL].

32. T. K. Tsang, P. Wu, Y. L. Y. Lin, E. Lau, G. M. Leung, B. J. Cowling, Impact of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China. *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.23.20041319 27 March 2020.

33. K. Sun., Spreadsheet of patient-level data until 31.01.2020. https://docs.google.com/spreadsheets/d/1Gb5cyg0fjUtsqh3hl_L-C5A23zIOXmWH5veBklfSHzg/edit?usp=sharing. From K. Sun, J. Chen, C. Viboud, Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *The Lancet Digital Health*. **2**, 4 (2020).

34. L. Tindale, M. Coombe, J. E. Stockdale, E. Garlock, W. Y. V. Lau, M. Saraswat, Y.-H. B. Lee, L. Zhang, D. Chen, J. Wallinga, C. Colijn, Transmission interval estimates suggest pre-symptomatic spread of COVID-19. *medRxiv* [Preprint] doi:10.1101/2020.03.03.20029983. 06 March 2020.

35. T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, N. Hens, Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv* [Preprint] https://doi.org/10.1101/2020.03.05.20031815. 8 March 2020.

36. X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *medRxiv* [Preprint], doi:10.1101/2020.03.15.20036707. 18 March 2020.

37. L. Tian, X. Li, F. Qi, Q. Tang, V. Tang, J. Liu, X. Cheng, X. Li, Y. Shi, H. Li, L. Tang, Pre-symptomatic transmission in the evolution of the COVID-19 pandemic. *arXiv* [Preprint] a.rXiv:2003.07353. 27 March 2020.

38. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, Factors that make an infectious disease outbreak controllable. PNAS, 101 (16), 6146-6151 (2020).

39. L'unico giornale di sinistra, Covid_19: Open letter from Italy to the international scientific community. LEFT https://left.it/2020/03/13/covid_19-open-letter-from-italy-to-the-international-scientific-community/. 13 March 2020.

40. T. Liu, J. Hu, J. Xiao, G. He, M. Kang, Z. Rong, L. Lin, H. Zhong, Q. Huang, A. Deng, et al. Time-varying transmission dynamics of novel coronavirus pneumonia in China. *bioRxiv* [Preprint] https://doi.org/10.1101/2020.01.25.919787. 13 February 2020.

41. N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S. Jung, B. Yuan, R. Kinoshita, H. Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of Clinical Medicine*, **9**(2):538 (2020).

42. J. A. Backer, D. Klinkenberg, J. Wallinga. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Eurosurveillance*, **25**(5) (2020).

43. M. U. G. Kraemer, C. Yang, B. Gutierrez, C. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, et al. The effect of human mobility and control measures on the COVID-19 epidemic in china. *Science* (25 Mar 2020). https://doi.org/10.1126/science.abb4218.

44. N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College London COVID-19 Response Team* (2020).

45. S. Wood, mgcv: Mixed GAM Computation Vehicle with automatic smoothness estimation. R package version 1.8-31 (2019) https://CRAN.R-project.org/package=mgcv.

46. S. Wood, *Generalized Additive Models* (New York: Chapman and Hall/CRC, 2017). https://doi.org/10.1201/9781315370279.

47. P. McCullagh, J. A. Nelder, Generalized Linear Models (Routledge, 2018).

48. J. D. Murray, *Asymptotic analysis* (Springer, New York, 1992).

49. J. Sun, Empirical estimation of a distribution function with truncated and double interval censored data and its application to aid studies. *Biometrics,* **51**, 1096-1104 (1995).

50. L. Pellis, N. M. Ferguson, C. Fraser, Epidemic growth rate and household reproduction number in communities of households, schools and workplaces. *J. Math. Biol*. **63**, 691-734 (2010).

51. Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, L. A. Meyers, Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis*. 26, (2020 Jun). https://doi.org/10.3201/eid2606.200357.

52. L. Pellis, S. Cauchemez, N. M. Ferguson, C. Fraser, Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nat. Commun*, **11**, 906 (2020). https://doi.org/10.1038/s41467-019-14229-4.

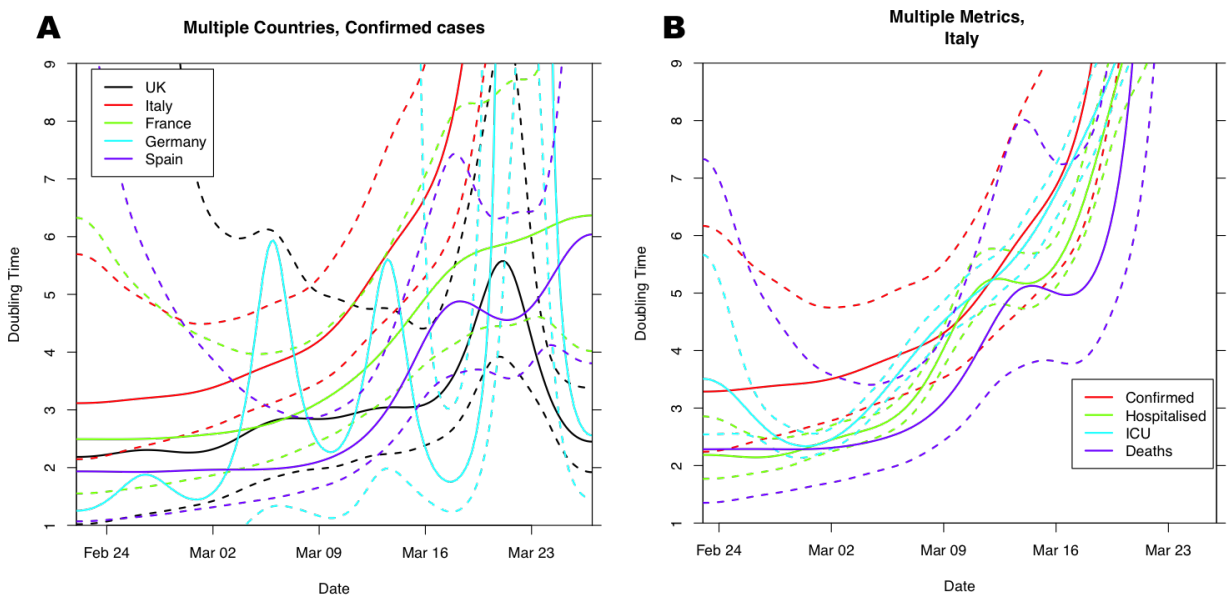

Fig. 1. Fit of a Generalised Additive Model with quasi-Poisson Family and canonical link to data between 23 February to 26 March, thin plate splines were used and the resulting output is then converted to an instantaneous estimate of doubling time (see methods). (A) Instantaneous doubling time in selected European countries (UK, Italy, France, Germany and Spain). (B) Instantaneous doubling time arising from four different daily surveillance metrics (Confirmed Cases, Hospitalised cases, cases in ICU and Deaths) within Italy. Note that the Italian Hospitalised and ICU cases are daily prevalence rather than daily incidence.

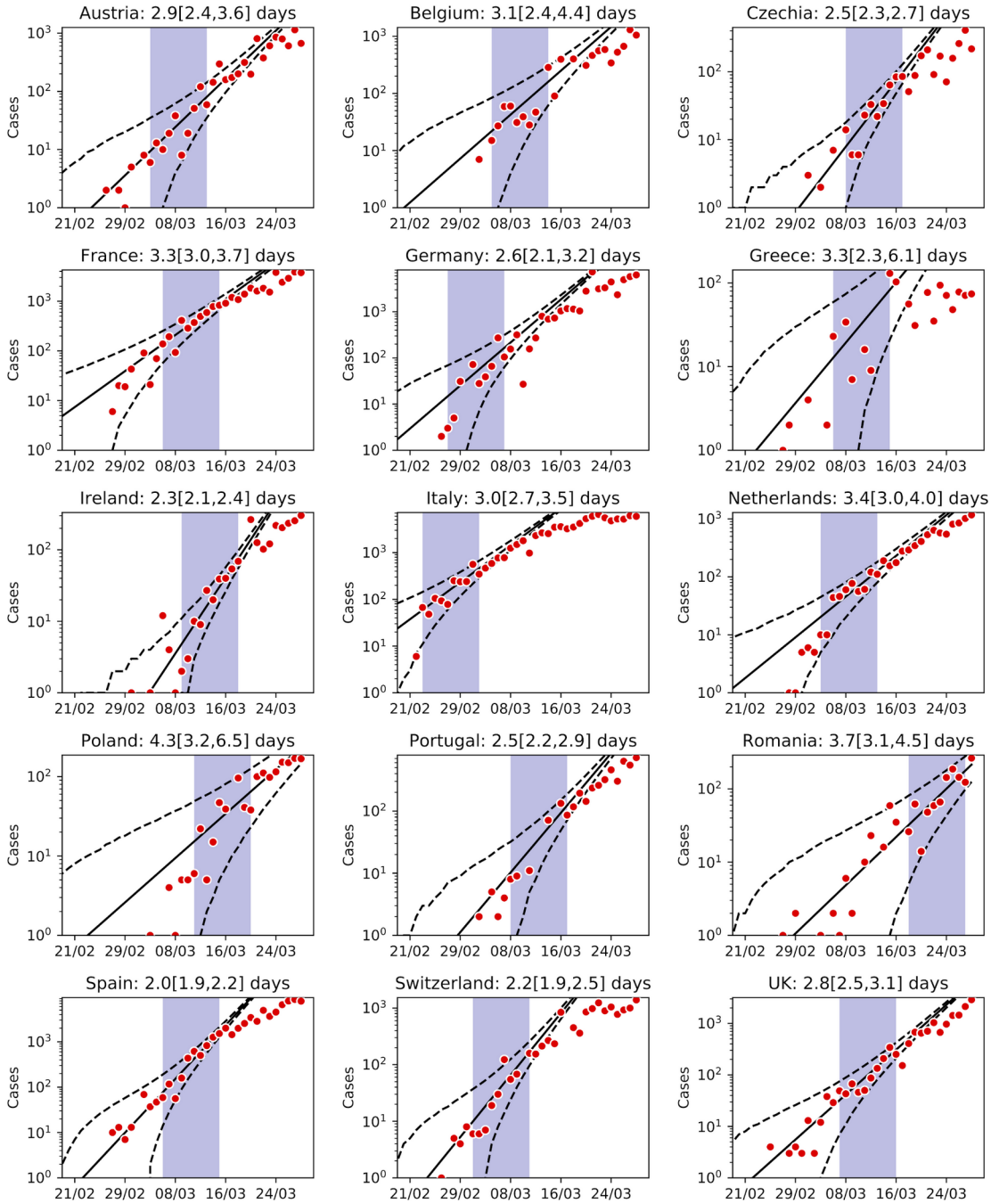

Fig. 2. Log daily confirmed cases and growth rate estimation for numerous European countries. Curves (black lines) are fit using the Generalised Linear Model methodology (see Materials and Methods) to the first 9 days non-zero data after a cumulative incidence of 25 is reached (period shown as a blue rectangle), which for most countries coincided with the start of sustained local transmission, with the exception of the UK and Romania where fitting started an additional 9 days after this criterion was met to reflect the local situation. Daily new cases are shown as red dots.

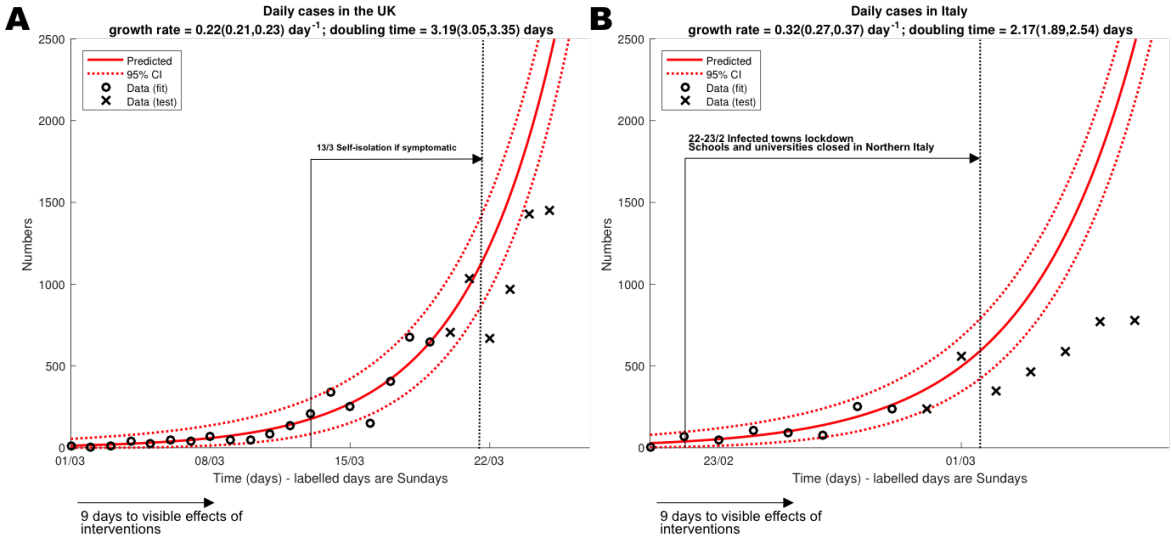

Fig. 3. Observed departure from unconstrained exponential growth approximately 9 days after intervention. Expected numbers of daily confirmed cases in (A) the UK and (B) Italy predicted in the absence of intervention (red line), fit using a Generalised Linear Model (see Materials and Methods). The fit is performed on circles; crosses are data added after the prediction is made. The exponential trend appears reasonably accurate for two more days, after which the effect of the first major intervention becomes visible.

Parameter	Mean	Standard deviation	Confidence interval for the mean	Sample size	Source
Incubation period	**4.8**	**-**	**2.2 - 7.4**	**16**	**[*40*]**
Incubation period	**5.6**	**3.9**	**4.4 - 7.4**	**52**	**[*41*]**
Incubation period	**6.4**	**2.3**	**5.6 - 7.7**	**88**	**[*42*]**
Incubation period	**4.84**	**2.79**	**-**	**162**	**Materials and Methods (Data – [*33*])**
Onset to confirmation	**4.8**	**3.03**	**-**	**38**	**[*43*]**
Onset to hospitalisation	**5**	**-**	**-**	**-**	**[*44*]**
Onset to hospitalisation (dead)	**6.6**	**-**	**5.2 - 8.8**	**34**	**[*41*]**
Onset to hospitalisation (alive)	**9.7**	**-**	**5.4 - 17**	**155**	**[*41*]**
Onset to hospitalisation (UK)	**5.14**	**4.20**	**-**	**90**	**Materials and Methods (Data - PHE)**
Onset to hospitalisation (Singapore)	**2.62**	**2.38**	**-**	**92**	**Materials and Methods (Data - [*33*])**
Onset to hospitalisation (Hong Kong)	**4.41**	**4.63**	**-**	**52**	**Materials and Methods (Data - [*33*])**

Table 1. Estimates of incubation period and delays from onset of symptoms to confirmation/ hospitalisation. For estimates from the literature, the source is provided. For our estimates, the details can be found in Materials and Methods.

Supplementary Materials for

Challenges in control of Covid-19: short doubling time and long delay to effect of interventions

Lorenzo Pellis[1,*], Francesca Scarabel[2,3], Helena B. Stage[1], Christopher E. Overton[1], Lauren H. K. Chappell[4], Katrina A. Lythgoe[5], Elizabeth Fearon[6], Emma Bennett[7], Jacob Curran-Sebastian[1], Rajenki Das[1], Martyn Fyles[1], Hugo Lewkowicz[1], Xiaoxi Pang[1], Bindu Vekaria[1], Luke Webb[1], Thomas House[1,8], Ian Hall[1,7]

[1]Department of Mathematics, The University of Manchester, Manchester, UK
[2]LIAM - Laboratory of Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto Ontario, Canada
[3]CDLab - Computational Dynamics Laboratory, Department of Mathematics, Computer Science and Physics, University of Udine, Italy
[4]Department of Plant Sciences, University of Oxford, UK
[5]Big Data Institute, University of Oxford, UK
[6]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, UK
[7]Emergency Response Department, Public Health England, UK
[8]IBM Research, Hartree Centre, SciTech Daresbury, Warrington, UK

[*]Correspondence to: lorenzo.pellis@manchester.ac.uk

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S4
Table S1

Other Supplementary Materials for this manuscript include the following:

Data sources and the code used to carry out our data analysis can be found at:
`https://github.com/thomasallanhouse/covid19-growth`

Materials and Methods

Data Sources

We consider four sources for our epidemiological data: the WHO [28], line-list data provided by Public Health England (PHE), line-list data from [33] and the Italian Istituto Superiore di Sanità [5]. Of these data sets, three are publicly available. The line-list from PHE is unfortunately not publicly available. From these sources, we extract epidemiological data concerning case counts, incidence, hospitalisation, and delays between infection and symptom onset, and onset of symptoms and hospitalisation.
These data sources and the code used to carry out our data analysis can be found at:
`https://github.com/thomasallanhouse/covid19-growth`.

Supplementary Text

Fitting the growth rate

Typically, an infection spread from person to person will grow exponentially in the early phase of an epidemic. This exponential growth can be measured through the real time growth rate r so that, loosely speaking, the prevalence of infection is

$$I(t) = I_0 \mathrm{e}^{rt} + \text{noise}. \tag{S1}$$

A natural mathematical model to derive the estimate of r is a Poisson family generalised linear model (GLM) with a log link. Given the over-dispersed noise inherent in both disease dynamics and surveillance data, a quasi-Poisson family is considered here.

The growth rate r is more intuitively reported as a doubling time (the time taken to double case numbers) and so $t_D = \ln(2)/r$. The log-linear analysis from a Poisson GLM defined formally below is restricted to datasets (or time windows) with clear exponential growth, or when additional explanatory variables, which are rarely available in real time, exist.

To allow, in semi-parametric manner, time variation in growth rates we adapt a generalised additive model (GAM) where $I \propto \mathrm{e}^{s(t)}$ for some smoother $s(t)$. In particular, we use a quasi-Poisson family with canonical link and a thin-plate spline as implemented in the R package *mgcv* [45, 46]. The instantaneous local growth rate is then the time derivative of the smoother $\dot{s}(t)$ and an instantaneous doubling time calculated as $t_D = \ln(2)/\dot{s}(t)$. Potential issues with the GAM approach include that extrapolation outside of the data range (and hence forecasting epidemic trend) is not sensible, and that there may be boundary effects from the choice of smoother. However, this approach has the major advantage that it allows for time-varying estimates of doubling time and thereby implicitly allows for missing explanatory information.

As well as the semi-parametric GAM approach, we take a parametric approach based on direct estimation of the exponential growth rate. This lacks the ability to capture time variation, but allows for extrapolation and epidemiological interpretation. To capture over-dispersion we use a quasi-Poisson family for the noise model. Explicitly, the Negative Binomial probability mass function (pmf) is

$$\mathrm{NB}(k|n,p) = \binom{k+n-1}{n-1} p^n (1-p)^k \ . \tag{S2}$$

We will work in the parameterisation where the mean is μ and the variance is $\theta\mu$, i.e.

$$p(\theta) = \frac{1}{\theta} \ , \qquad n(\mu,\theta) = \frac{\mu}{\theta - 1} \ . \tag{S3}$$

Let the number of new cases on day t be $y(t)$. We assume that this is generated by an exponentially growing mean,

$$\mathbb{E}[y(t)] = y_0 \mathrm{e}^{rt} = \exp(\ln(y_0) + rt) \ , \tag{S4}$$

which is then combined with the negative binomial pmf to give a likelihood function for the observations over a set of times $\mathcal{T}$ of

$$L(\mathbf{y}|y_0, r, \theta) = \prod_{t\in\mathcal{T}} \mathrm{NB}(y(t)|n(y_0 \exp(rt), \theta), p(\theta)) \ , \tag{S5}$$

where $\mathbf{y} = (y(t))_{t\in\mathcal{T}}$. This can then be viewed as a generalised linear model (GLM) with time as a continuous covariate, intercept $\ln(y_0)$, slope r, exponential link function and negative binomial noise model [47]. We can perform inference through numerical maximum likelihood estimation (MLE) and calculate confidence intervals using the Laplace approximation [48].

Each page of Figure S1 shows the GAM compared to a simple GLM with $\theta = 1$ and $\mathcal{T}$ taken to be all of the data range, in contrast to the results in the main paper and Figure S2, where we fit θ and let $\mathcal{T}$ correspond to the first nine days of the local epidemic after the cumulative number of cases has reached 25 (the only exceptions are the UK and Romania, where fitting started an additional 9 days later to reflect the local situation). While the simple GLM method is clearly inadequate if the fit is not restricted to a window where exponential growth appears reasonable, it is shown for comparison in these plots. The left panel shows the output of the model fit and the data, the middle panel the instantaneous growth rate from GAM (black) and the growth rate from GLM

(red) with 95% CI, and the right most panel shows these growth rates converted to doubling times. Of the fifteen countries, two (Belgium and Romania) have equivalent fits from GAM as from GLM and show constant growth rates over the time period. The Czech Republic, Greece, Ireland and Poland have the central GLM result within the 95% CI of the GAM suggesting a constant growth rate is a plausible explanation of the data reported. Austria, France, Italy, Portugal, Spain, Switzerland show a fairly smooth transition from short to longer doubling times. Germany, Netherlands and UK show more oscillatory behaviour in doubling times.

As a further test for the robustness of our results, we simply visually assess the growth rates of the epidemic in a set of countries by plotting data on a log scale and compare the observed slopes with pure exponential trends (Figure S3). To avoid relying on confirmed cases only, we plot a 'mixed bag' of cumulative cases, daily new confirmations, hospitalisations and deaths. We also plot exponential trends with a lower growth rate ($r = 0.18$ and $r = 0.13$) to aid the visual distinction in slopes between different exponential growths and the visual assessment of the potential effects of interventions in Italy.

Estimating delay distributions

Delay distributions describe the time delay between two events. To understand how long until the impact of an intervention may be observed, we need to understand the delay between infection and symptom onset (the incubation period), and the delay from onset to hospitalisation. A difficulty with estimating delay distributions during an outbreak is that events are only observed if they occur before the final sampling date. Since delay distributions depend on the time between two events, if the first event occurs near to the end of the sampling window, it will only be observed if the delay to the second event is short. This causes an over-expression of short delays towards the end of the sampling window, which is exacerbated by the exponential growth of the epidemic. Therefore, we need to account for this growth and truncation within our model.

To fit the data we use maximum likelihood estimation. However, we do not observe the delay directly, instead observing the timing of the two events. Therefore, we need to construct a likelihood function for observing these events. Following [*49*], we construct the conditional density function for observing the second event given the time of the first event and given that the second event occurs before date T. That is, we are interested in the conditional density function

$$L(X_2 \in x_2 | X_1 \in x_1, X_2 \leq T) = \frac{L(X_2 \in x_2, X_1 \in x_1, X_2 \leq T)}{L(X_1 \in x_1, X_2 \leq T)}, \tag{S6}$$

where x_1 and x_2 can be exactly observed or interval censored.

The delay from onset to hospitalisation for the UK is estimated using FF100 data provided by Public Health England, which contains data on the first few hundred infected individuals in the UK. This data incorporated the time of symptom onset and time of hospitalisation. There were some cases who were hospitalised before their onset date. These cases have been removed from the data set, since they do not provide insight into the delay. Additionally, some cases have no symptom onset, so these have also been removed from the data. For cases where symptom onset and hospitalisation occur on the same day, we add half a day to the hospitalisation day, since the delay is unlikely to be instantaneous. After tidying the data, this left 106 cases from which to infer the onset to hospitalisation delay. The dates in the line list are recorded exactly, so the likelihood function becomes

$$L(H = h | O = o, H \leq T) = \frac{g(o) f(h-o)}{\int_0^{T-o} g(o) f(x) \mathrm{d}x} = \frac{f(h-o)}{\int_0^{T-o} f(x) \mathrm{d}x}, \tag{S7}$$

where f is the density of the onset to hospitalisation delay and g is the density of the onset time. Using this truncation corrected method and a gamma distribution to fit the delay distribution, we get a mean delay of 5.14 with standard deviation 4.20. Unfortunately, we cannot share the FF100 data. To compare different regions, we also use data from Hong Kong and Singapore to estimate the local onset to hospitalisation delays. This data is taken from an open access line-list [*33*], and the filtered data sets used are provided in the supplementary material. Using the method above, for Hong Kong the mean delay is 4.41 days, with standard deviation 4.63, and for Singapore the mean delay is 2.62 with standard deviation 2.38.

For the incubation period, we use data from Wuhan during the early stages of the outbreak. This data was extracted from an open access line-list [*33*], containing dates when individuals were in Wuhan and when they developed symptoms (among other information). Since this data is from the early stages of the epidemic, the majority of cases were in Wuhan. Therefore, it is likely that these individuals were infected in Wuhan, so the time spent in Wuhan provides a potential exposure window during which infection occurred. For individuals with symptom onset date before leaving Wuhan or the same day they left Wuhan, the upper bound on the exposure window was adjusted to half a day before symptom onset. Using the data as of 21/02/2020, we have 162 cases from which to infer the incubation period. This infection date is interval censored, so we obtain the likelihood function

$$L(O = o | a < I < b, O \leq T) = \frac{\int_a^b g(i) f(o-i) \mathrm{d}i}{\int_a^b \int_0^{T-i} g(i) f(x) \mathrm{d}x \mathrm{d}i}, \tag{S8}$$

where f is the incubation period density function and g is the density of the infection date. We assume g is proportional to force of infection of the outbreak, which is assumed to follow exponential growth with rate parameter 0.25 day^{-1}. Using a gamma distribution to describe the incubation period, we get a mean incubation of 4.84 days with standard deviation 2.79. The data for the incubation period is also included as supplementary material, along with MATLAB code to perform the maximum likelihood estimation.

Estimation of R_0

The relationship between the growth rate r and the basic reproduction number R_0 in a simple homogeneously mixing model is provided by the Lotka-Euler equation:

$$\frac{1}{R_0} = \int_0^\infty \omega(\tau) \mathrm{e}^{-r\tau} \mathrm{d}\tau, \tag{S9}$$

where τ represents the time since the infection of an individual and $\omega(\tau)$ is the infectious contact interval distribution, defined as the probability density function (pdf) of the times (since infection) at which an infectious contact is made. An infectious contact is a contact that results in an infection if the contactee is susceptible, and early on in the epidemic any randomly selected contactee is almost surely susceptible.

Equation (S9) assumes all individuals have the same infectious contact interval distribution. However, if we assume random variability between individuals, there will be a set $\mathcal{S}$ of curves $\Omega(\tau)$. However, equation (S9) still applies, with $\omega(\tau)$ being the time-point average of all curves in $\mathcal{S}$ [*50*]; see Figure S4.

The generation time T_g is defined as the mean of the infectious contact interval distribution ω:

$$T_g = \int_0^\infty \tau \omega(\tau) \mathrm{d}\tau. \tag{S10}$$

The same definition extends to a random infectivity profiles of which ω is the time-point average.

For the incubation period we use our estimates from Table 1 (mean 4.84, standard deviation 2.79), which are anyway similar to those estimated by others. However, information about any form of pre-symptomatic transmission is hard to obtain but crucial for R_0 estimates [*34-37*]. Furthermore, there is also limited information concerning how infectivity changes over time. Therefore, Table S1 reports the estimates we obtain assuming the infectious period starts at the onset of symptoms, one, two or three days earlier, and assuming a Gamma-shaped infectivity with mean 2 or 3 days. In both cases, the standard deviation is assumed to be 1.5 and the infectivity is truncated after 7 days (see Figure S4).

We conclude that the estimates of R_0 are highly sensitive to small variations in quantities that are poorly supported by available data, but that for a growth rate of 0.25 day^{-1}, close to what is observed in Italy and the UK, are also generally larger, and possibly much larger, than official estimates [*3*, *12-15*]. Smaller values in this range are associated with significant amounts of pre-symptomatic transmission [*34*], leading to a generation time for example compatible with some of the shortest estimates of the serial interval seen in the literature [*51*], and with a front-loaded infectivity curve (mean 2, rather than 3).

We tested further assumptions. A simple SEIR model, with exponentially distributed incubation and infectious periods (with the same means as above but constant infectivity) leads to much smaller values of R_0 than our estimates, as it favours really short incubation periods (Table S1B, left). Estimates, instead, do not change significantly if high variability in total infectiousness between individuals, in line with what observed for SARS, is assumed (Table S1B, right) or if 50% of cases are assumed to be fully asymptomatic and transmit at half the rate as those with symptoms (not shown).

These simple estimates are obtained under the assumption of mass-action mixing. The explicit presence of a social structure (e.g. age-stratification, household/network structure, etc.), which in principle could affect them, is likely negligible in such a high R_0 and growth rate regime [*52*]. The effect of the social structure on transmission is expected to grow in importance (especially the household structure, since isolation and quarantine facilitate within-household transmission) the closer R_0 is to 1.

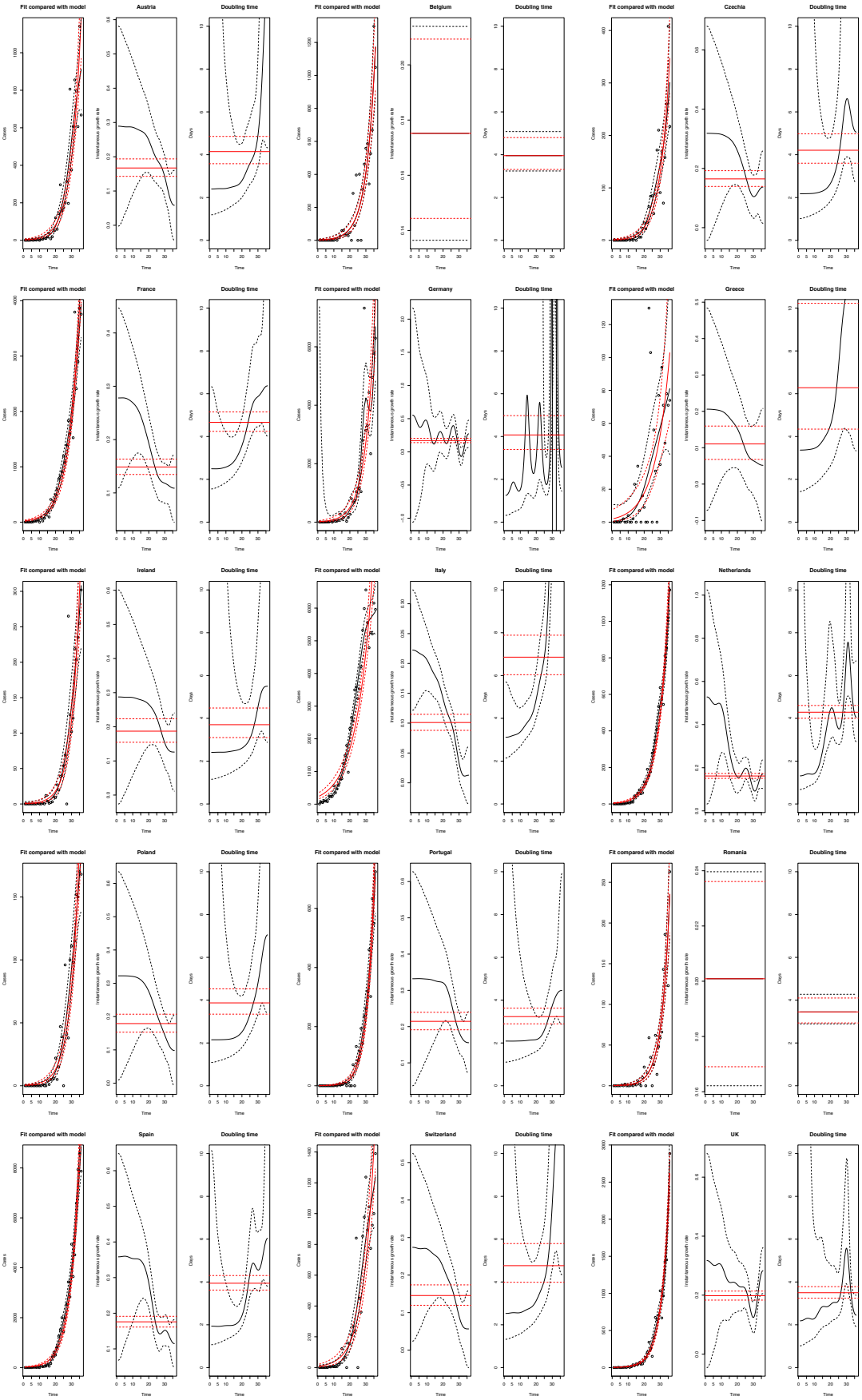

Fig. S1: GAM and GLM model fit for multiple countries. Left panel: model fit and data; middle panel: instantaneous growth rate from the GAM (black) and the (constant) growth rate from the GLM (red) with 95% CI; right panel: doubling time from GAM (black) and GLM (red). The x-axis gives time since 21 February 2020.

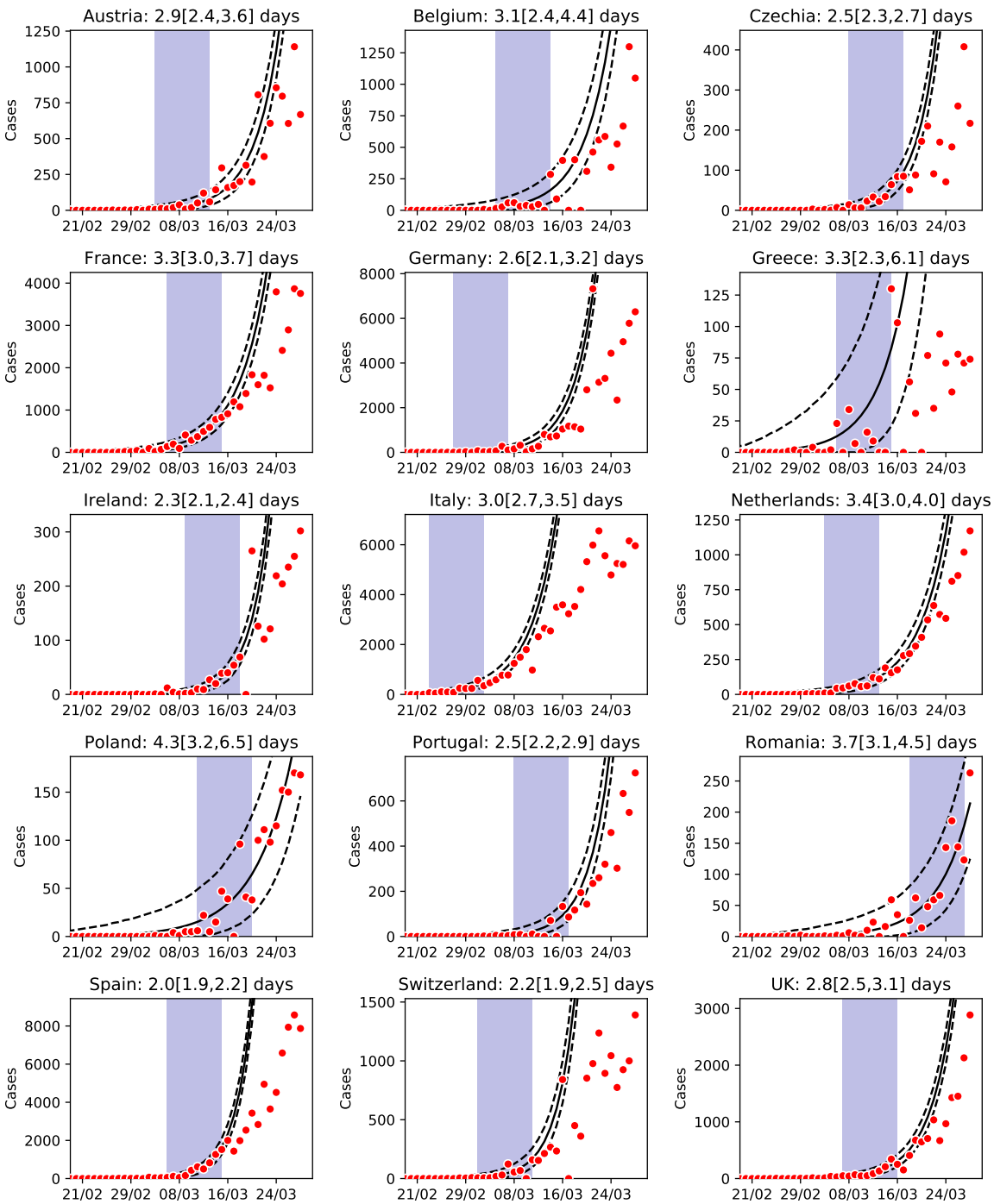

Fig. S2: Linear version of Figure 2 from the main paper.

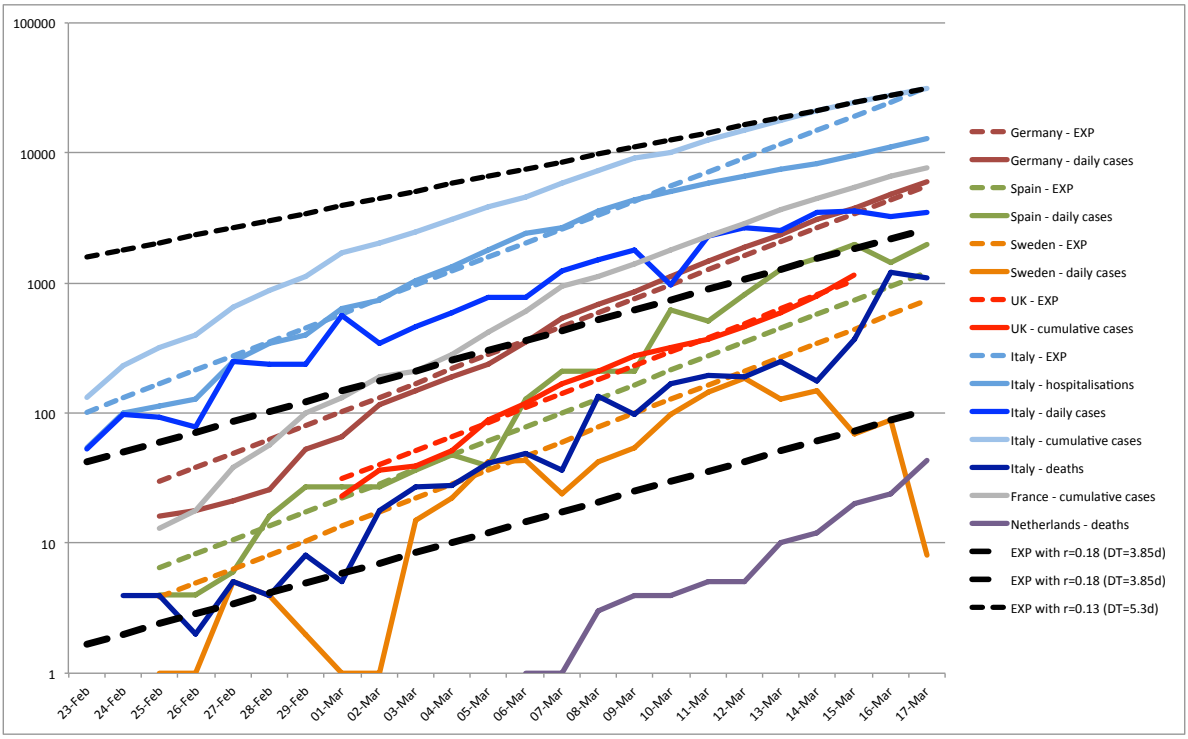

Fig. S3: Growth rates in different EU countries and visual comparison with simple exponential growth. A 'mixed bag' of numbers of daily new confirmed cases, cumulative confirmed cases, hospitalisations and deaths, all showing similar fast growth. Visible exceptions are: Sweden, which has changed testing regime; Italy, where growth in new confirmed and cumulative confirmed cases appears to be slowing down from beginning of March (quarantining infected towns from 22-23 Feb) and growth in hospitalisations around a week later (deaths not slowing down yet); deaths in Netherlands and new cases in Spain, which are growing even faster than the rest. Coloured dashed lines are pure exponential growth with $r = 0.25$ day^{-1} (doubling time of 2.77 days, in the same ballpark as UK and Italy). No statistical fit is performed: lines' slope and intercept are purely eyeballed. Black thick dashed lines show growth at $r = 0.18$ day^{-1} (doubling time of 3.85 days). Black thin dotted line shows growth at $r = 0.13$ day^{-1} (doubling time of 5.33 days). Notice how this is roughly the growth in cases (new and cumulative) and hospitalisations in Italy around 17 March.

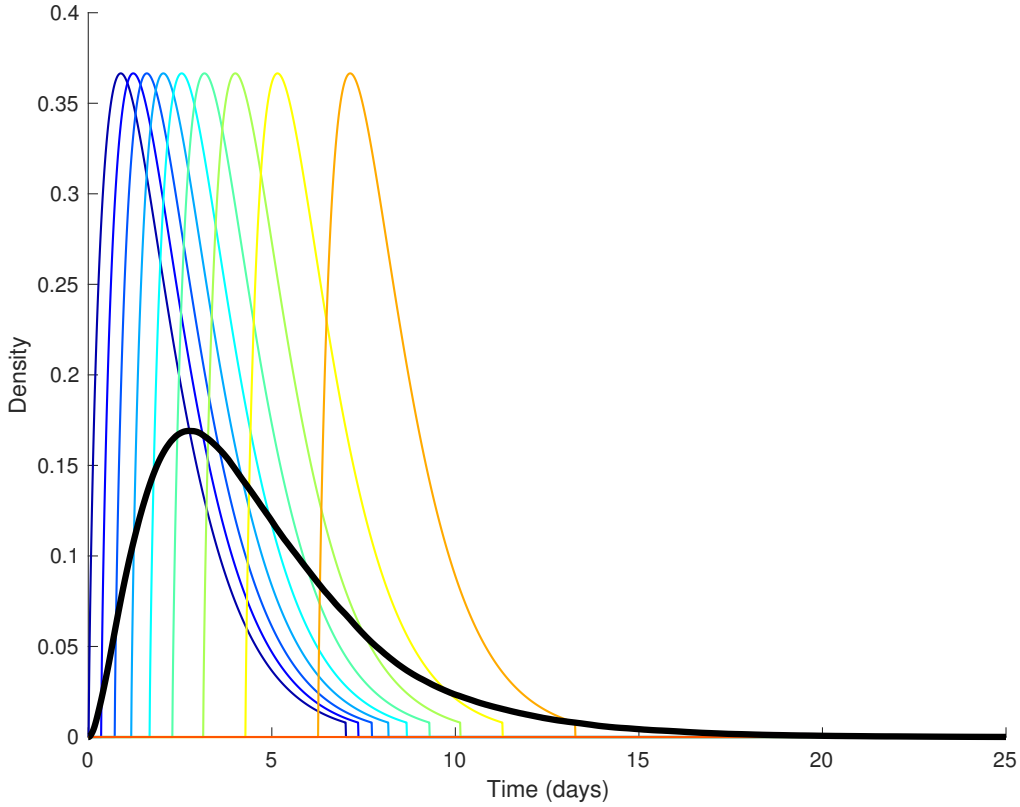

Fig. S4: Random variability in the generation time distribution (thin coloured lines) and time-point average (thick black line).

A		R0, assuming infectivity peaks in 2 days				R0, assuming infectivity peaks in 3 days			
Growth rate (1/day)	Doubling time (day)	Days of pre-symptomatic transmission: 0	1	2	3	Days of pre-symptomatic transmission: 0	1	2	3
0.1	6.93	1.88	1.70	1.55	1.40	2.03	1.84	1.66	1.52
0.15	4.62	2.51	2.16	1.87	1.62	2.81	2.42	2.09	1.83
0.2	3.47	3.29	2.71	2.24	1.85	3.81	3.14	2.58	2.16
0.25	2.77	4.25	3.34	2.64	2.10	5.09	4.00	3.15	2.54
0.3	2.31	5.42	4.07	3.09	2.36	6.70	5.04	3.80	2.95
Generation time:		6.76	5.76	4.78	3.72	7.56	6.55	5.51	4.59

B		R0 for SEIR model				R0 assuming variability in total infectiousness between cases			
Growth rate (1/day)	Doubling time (day)	Days of pre-symptomatic transmission: 0	1	2	3	Days of pre-symptomatic transmission: 0	1	2	3
0.1	6.93	1.63	1.52	1.41	1.31	1.88	1.71	1.55	1.41
0.15	4.62	2.00	1.83	1.66	1.49	2.51	2.17	1.88	1.64
0.2	3.47	2.40	2.16	1.92	1.67	3.29	2.72	2.25	1.88
0.25	2.77	2.81	2.50	2.19	1.87	4.26	3.36	2.66	2.14
0.3	2.31	3.25	2.86	2.47	2.08	5.43	4.10	3.11	2.41
Generation time:		6.20	5.37	4.40	3.39	6.81	5.84	4.78	3.80

Table S1: Values of R_0 derived from different growth rates and different modelling assumptions. A) Gamma-distributed latent period with estimates from Table 1, and Gamma-shaped infectivity profile with mean 2 (left) and 3 (right) and standard deviation 1.5; B) SEIR model (left) and same model as in A) but assuming total infectiousness is randomly drawn from a Gamma-distribution with mean 1 and standard deviation $1/\sqrt{k}$, with $k = 0.25$.