Temporal variation in transmission during the COVID-19 outbreak

S. Abbott (1), J. Hellewell (1), J. D. Munday (1), J. Y. Chun, R. N. Thompson (1), N. Bosse (1), Y. D. Chan (1), T. W. Russell (1), C. I. Jarvis (1), CMMID COVID team (1), S. Flasche (1), A. J. Kucharski (1), R. M. Eggo (1), S. Funk (1).

Correspondence to: sam.abbott@lshtm.ac.uk

1. Center for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, United Kingdom

Last Updated: 2020-03-15

Summary

Aim: To identify changes in the reproduction number, rate of spread, and doubling time during the course of the COVID-19 outbreak whilst accounting for potential biases due to delays in case reporting.

Latest estimates as of the 2020-03-15

Global map

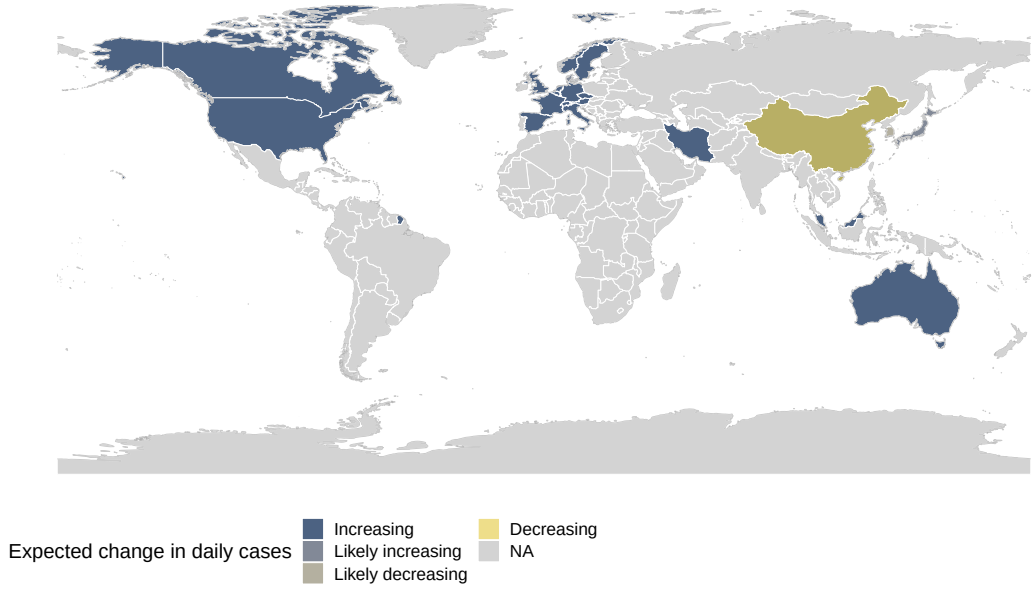

Figure 1: Global map of the expected change in daily cases based on data from the 2020-03-15. Note: only country level estimates are shown.

Summary of latest reproduction number and case count estimates

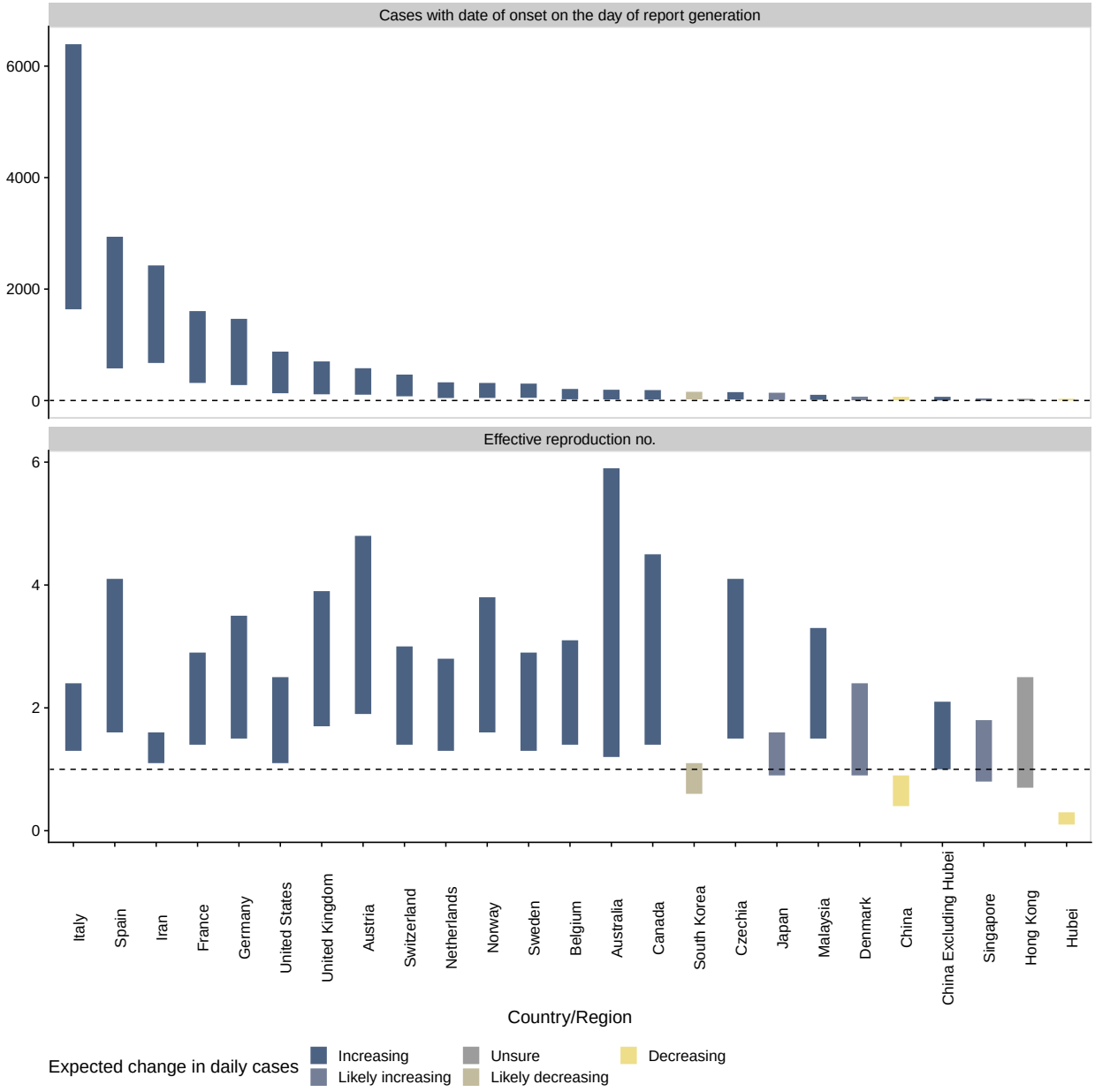

Figure 2: Cases with date of onset on the day of report generation and the time-varying estimate of the effective reproduction number (bar = 95% credible interval) based on data from the 2020-03-15. Countries/Regions are ordered by the number of expected daily cases and shaded based on the expected change in daily cases. The dotted line indicates the target value of 1 for the effective reproduction no. required for control and a single case required fror elimination.

Reproduction numbers over time in the six countries with the most cases currently

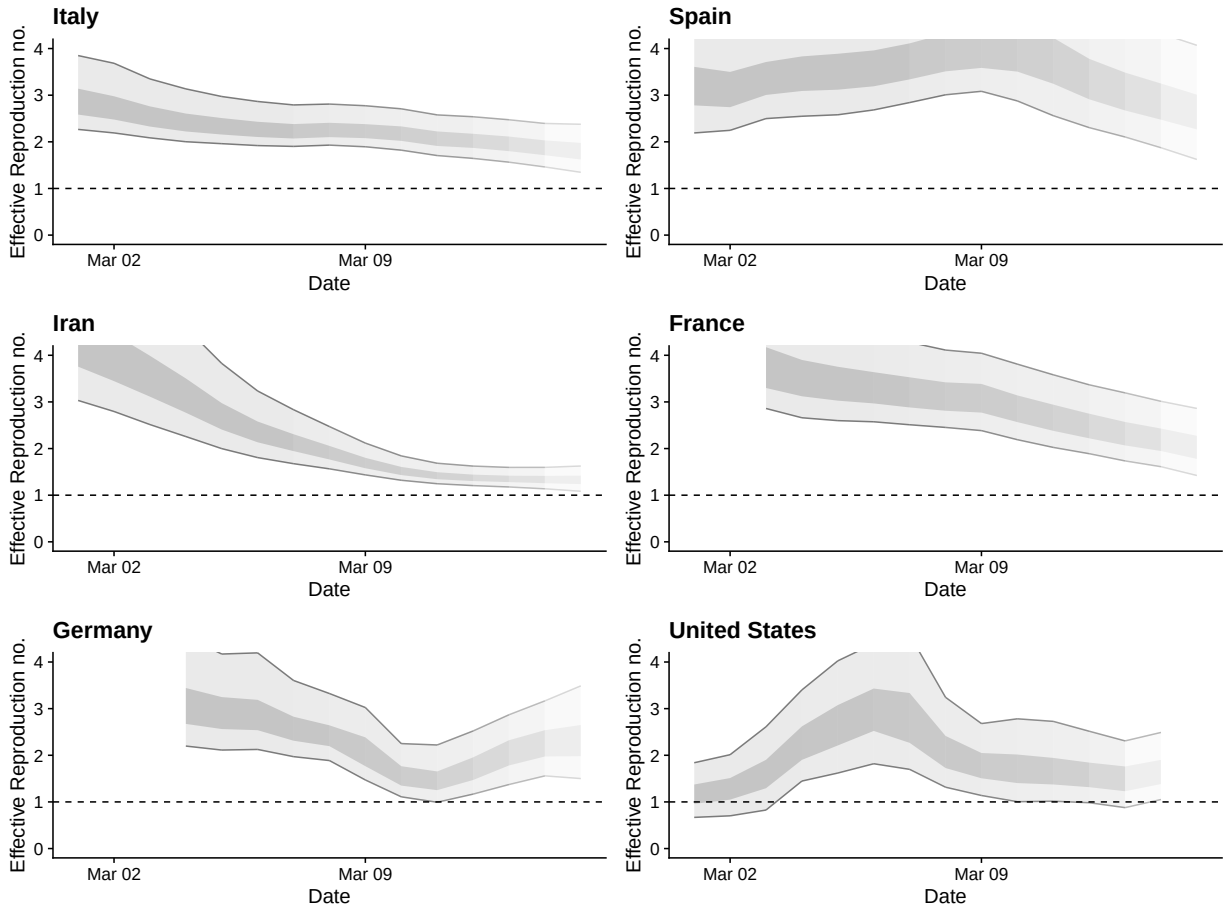

Figure 3: Time-varying estimate of the effective reproduction number (light grey ribbon = 95% credible interval; dark grey ribbon = the interquartile range) based on data from the 2020-03-15 in the countries/regions expected to have the highest number of incident cases. Confidence in the estimated values is indicated by shading with reduced shading corresponding to reduced confidence. The dotted line indicates the target value of 1 for the effective reproduction no. required for control.

Latest estimates summary table

Table 1: Latest estimates of the number of cases by date of onset, the effective reproduction number, and the doubling time for the 2020-03-15 in each region included in the analysis. Based on the last 7 days of data. The 95% credible interval is shown for each numeric estimate. China excludes Hubei.

Country/Region	Cases with date of onset on the day of report generation	Expected change in daily cases	Effective reproduction no.	Doubling time (days)
Italy	1637 – 6393	Increasing	1.3 – 2.4	3 – Cases decreasing
Spain	577 – 2939	Increasing	1.6 – 4.1	1.9 – 57
Iran	674 – 2425	Increasing	1.1 – 1.6	4 – Cases decreasing
France	316 – 1605	Increasing	1.4 – 2.9	3 – 490
Germany	278 – 1466	Increasing	1.5 – 3.5	1.9 – Cases decreasing
United States	131 – 878	Increasing	1.1 – 2.5	2 – Cases decreasing
United Kingdom	112 – 702	Increasing	1.7 – 3.9	2 – 7.6
Austria	105 – 580	Increasing	1.9 – 4.8	1.7 – 4.3
Switzerland	76 – 466	Increasing	1.4 – 3	2.3 – Cases decreasing
Netherlands	45 – 327	Increasing	1.3 – 2.8	2.9 – Cases decreasing
Norway	47 – 316	Increasing	1.6 – 3.8	1.8 – Cases decreasing
Sweden	49 – 304	Increasing	1.3 – 2.9	2.7 – Cases decreasing
Belgium	23 – 208	Increasing	1.4 – 3.1	1.5 – Cases decreasing
Australia	21 – 194	Increasing	1.2 – 5.9	0.22 – Cases decreasing
Canada	18 – 188	Increasing	1.4 – 4.5	1.4 – Cases decreasing
South Korea	22 – 157	Likely decreasing	0.6 – 1.1	Cases decreasing
Czechia	18 – 151	Increasing	1.5 – 4.1	1.8 – 100
Japan	14 – 140	Likely increasing	0.9 – 1.6	4.2 – Cases decreasing
Malaysia	13 – 103	Increasing	1.5 – 3.3	0.19 – Cases decreasing
Denmark	5 – 70	Likely increasing	0.9 – 2.4	2.5 – Cases decreasing
China	3 – 68	Decreasing	0.4 – 0.9	9.8 – Cases decreasing
China Excluding Hubei	2 – 67	Increasing	1 – 2.1	2.3 – Cases decreasing
Singapore	1 – 38	Likely increasing	0.8 – 1.8	2.9 – Cases decreasing
Hong Kong	1 – 33	Unsure	0.7 – 2.5	0.22 – Cases decreasing
Hubei	1 – 33	Decreasing	0.1 – 0.3	2.7 – Cases decreasing

Methods

Summary

- Case counts by date, stratified by import status (local or imported), were constructed using the World Health Organization (WHO) situation reports and partial line-lists for each region [1,2].
- Case onset dates were estimated using case counts by date of report and a distribution of reporting delays fitted to partial line-lists from each region considered where available.
- Censoring of cases was adjusted for by assuming that the number of cases is drawn from a binomial distribution.
- Time-varying effective reproduction estimates were made with a 7-day sliding window using *EpiEstim* [4,5] adjusted for imported cases and assuming an uncertain serial interval with a mean of 4.7 days (95% CrI: 3.7, 6.0) and a standard deviation of 2.9 days (95% CrI: 1.9, 4.9) [6].
- Time-varying estimates of the doubling time were made with a 7-day sliding window by iteratively fitting an exponential regression model.

Limitations

- The estimated onset dates are based on current data for the delay in reporting. These data may not be representative of the underlying reporting distribution.
- The estimate of not-yet-confirmed cases to scale up recent numbers is uncertain and relies on the observed delays to confirmation to remain constant over the course of the outbreak.
- All data used is at a national/regional level; diagnostic capabilities may vary in different parts of each country/region, adding uncertainty to the reported numbers.
- Trends identified using our approach are robust to under-reporting assuming it is constant but absolute values may be biased by reporting rates. Pronouced changes in reporting rates may also impact the trends identified.
- Data on imported cases was only partially available.
- The reporting delay could not be estimated from line-list data for all regions. Region specific details are given in the individual regional reports.
- Data on imported cases may not be fully complete. This may bias estimates upwards when overall case counts are low.

Detail

Data

We used partial line-lists from each region that contained the date of symptom onset, date of confirmation and import status (imported or local) for each case [3] where available. The region reports give details of the steps taken where this data were not available. Daily case counts by date of report were extracted from the World Health Organization (WHO) situation reports for every location considered [1,2]. The case counts (and partial line-lists where available) were used to assemble the daily number of local and imported cases. Where the partial line-lists and case counts disagreed, it was assumed that the partial line-lists were correct and the WHO case counts were adjusted so that the overall number of cases occurring remained the same but the number of local cases being adjusted as needed.

Adjusting for reporting delays

Reporting delays for each country were estimated using the corresponding partial line-list of cases. The reporting delay could not be estimated from line-list data for all regions. Region specific details are given in the individual regional reports. The estimated reporting delay was assumed to remain constant over time in each location. We fitted an exponential distribution adjusted for censoring [7] to the observed delays using stan [8]. We then took 1000 samples from the posterior distribution of the rate parameter for the

exponential delay distribution and constructed a distribution of possible onset dates for each case based on their reporting date. To prevent spuriously long reporting delays, we re-sampled delays that were greater than the maximum observed delay in the observed data.

To account for censoring, i.e. cases that have not yet been confirmed but will show up in the data at a later time, we randomly sampled the true number of cases (including those not yet confirmed) assuming that the reported number of cases is drawn from a binomial distribution, where each case has independent probability p_i of having been confirmed, i is the number of days of the symptom onset before the report maximum observed report delay, and p_i is the cumulative distribution of cases that are confirmed by day i after they develop symptoms. We did not account for potential reporting biases that might occur due to changes in the growth rate of the outbreak over time.

Statistical analysis

We used the inferred number of cases to estimate the reproduction number on each day using the *EpiEstim* R package [4]. This uses a combination of the serial interval distribution and the number of observed cases to estimate the reproduction number at each time point [10,11], which were then smoothed using a 7-day time window. We assumed that the serial interval was uncertain with a mean of 4.7 days (95% CrI: 3.7, 6.0) and a standard deviation of 2.9 days (95% CrI: 1.9, 4.9) [6]. We used a common prior for the reproduction number with mean 2.6 and a standard deviation of 2 (inflated from 0.5 found in the reference) [12]. Where data was available, we used *EpiEstim* to adjust for imported cases [5]. The expected change in daily cases was defined using the proportion of samples with a reproduction number less than 1 (subcritical). It was assumed that if less than 5% of samples were subcritical then an increase in cases was definite, if less than 20% of samples were subcritical then an increase in cases was likely, if more than 80% of samples were subcritical then a decrease in cases was likely and if more than 95% of samples were subcritical then a decrease in cases was definite. For countries/regions with between 20% and 80% of samples being subcritical we could not make a statement about the likely change in cases (defined as unsure).

We estimated the rate of spread (r) using linear regression with time as the only exposure and logged cases as the outcome for the overall course of the outbreak [13]. The adjusted R^2 value was then used to assess the goodness of fit. In order to account for potential changes in the rate of spread over the course of the outbreak we used a 7-day sliding window to produce time-varying estimates of the rate of spread and the adjusted R^2. The doubling time was then estimated using $\ln(2)\frac{1}{r}$ for each estimate of the rate of spread.

We report the 95% confidence intervals for all measures using the 2.5% and 97.5% quantiles. The analysis was conducted independently for all regions and is updated daily as new data becomes available. Confidence in our estimates is shown using the proportion of data that were derived using binomial upscaling.

References

1 World Health Organization. **Coronavirus disease (COVID-2019) situation reports**. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

2 Brown E. *Data2019nCoV: Data on the covid-19 outbreak.* 2020.

3 Xu B, Gutierrez B, Hill S *et al.* Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. 2020.

4 Cori A. *EpiEstim: Estimate time varying reproduction numbers from epidemic curves.* 2019. https://CRAN.R-project.org/package=EpiEstim

5 Thompson R, Stockwin J, Gaalen R van *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* 2019;**29**:100356. doi:https://doi.org/10.1016/j.epidem.2019.100356

6 Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (2019-nCoV) infections. *medRxiv* Published Online First: 2020. doi:10.1101/2020.02.03.20019497

7 Thompson RN. 2019-20 Wuhan coronavirus outbreak: Intense surveillance is vital for preventing sustained transmission in new locations. *bioRxiv* 2020;1–14.

8 Stan Development Team. RStan: The R interface to Stan. 2020.http://mc-stan.org/

9 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria:: R Foundation for Statistical Computing 2019. https://www.R-project.org/

10 Cori A, Ferguson NM, Fraser C *et al.* A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology* 2013;**178**:1505–12. doi:10.1093/aje/kwt133

11 Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *American Journal of Epidemiology* 2004;**160**:509–16. doi:10.1093/aje/kwh255

12 Imai N, Cori A, Dorigatti I *et al.* **Report 3: Transmissibility of 2019-nCoV**. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf

13 Park SW, Champredon D, Weitz JS *et al.* A practical generation-interval-based approach to inferring the strength of epidemics from their speed. *Epidemics* 2019;**27**:12–8. doi:https://doi.org/10.1016/j.epidem.2018.12.002