Estimates of nosocomial and community transmission of COVID-19 in the England

Authors: *Sam Abbott, Joel Hellewell, Jonathan Read, Nikos I Bosse, Kath Sherratt, Robin Thompson, James D Munday, and Sebastian Funk on behalf of the LSHTM COVID-19 Modelling Team*

Date: 2020-04-19

Aim

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

Summary

- From the 18th of March hospitalised cases were extracted manually from daily SITREPs. Prior to the 18th data on hospitalisations was not available. We instead estimated hospitalisations using reported cases adjusted using the proportion of reported cases that were hospitalised from the 18th to the 24th of March by region. Data on reported cases was downloaded from a publicly curated dataset (Doughty 2020; Abbott et al. 2020) based on daily confirmed cases reported by Public Health England. Data was only available for England.
- Case onset dates were estimated by transforming the number of new confirmations in hospital using the distribution from onset of symptoms to lab report taken from an anoymised line list of cases. Case infection dates were then estimated by assuming a 5 day incubation period between infection and symptom onset.
- The right-truncation of cases that have had symptom onset but are yet to be confirmed was adjusted for by upscaling the numbers of case onsets close to the present date. This assumes that cases that have already onset are drawn from a binomial distribution with the probability of onset having occurred by a certain point given by the onset-to-confirmation distribution.
- Onset dates were scaled according to the weekly proportion of cases found in the CO-CIN data set to have had symptom onset 5 or more days after their admission date (deemed likely nosocomial), linearly extrapolated between the weekly data points for a daily scaling and extrapolated linearly beyond the most recent data point on 6 April. This resulted in an increase from 1% of cases (2 March) to 13% of cases (12 April) considered likely resulting from nosocomial transmission. The scaling was applied to onset dates and is therefore not reflected in the plots by admission date (bars), but it is reflected in the plots of the number of new cases by likely infection date (lines and shaded ribbons). All other hospitalised cases were deemed likely community cases.
- We assumed that likely nosocomial cases could be caused by all hospitalised cases, whether likely nosocomial or likely community cases.
- Time-varying effective reproduction estimates were generated using *EpiEstim* (Cori 2019; Thompson et al. 2019) 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) (Nishiura, Linton, and Akhmetzhanov 2020). The optimal window for the estimates was selected by comparing estimated cases against observed cases.
- Time-varying estimates of the doubling time were made with a 7-day sliding window by iteratively fitting an exponential regression model.

Limitations

- Estimated onset dates are derived using available data for the delay from symptom onset to lab report from an anoymised line list of cases. This early outbreak data may not be representative of the underlying delay distribution, which may also change over the course of the outbreak.
- The estimate of cases that have been infected but are not yet hospitalised, used to scale up recent confirmed case numbers, is uncertain and relies on the observed delays to lab report remaining constant over the course of the outbreak.
- Trends identified using our approach are robust to under-reporting assuming that the level of reporting remains consistent over time, however absolute values may be biased by reporting rates. Pronounced changes in the availability and use of tests may also impact the trends identified.

Cases by symptom onset

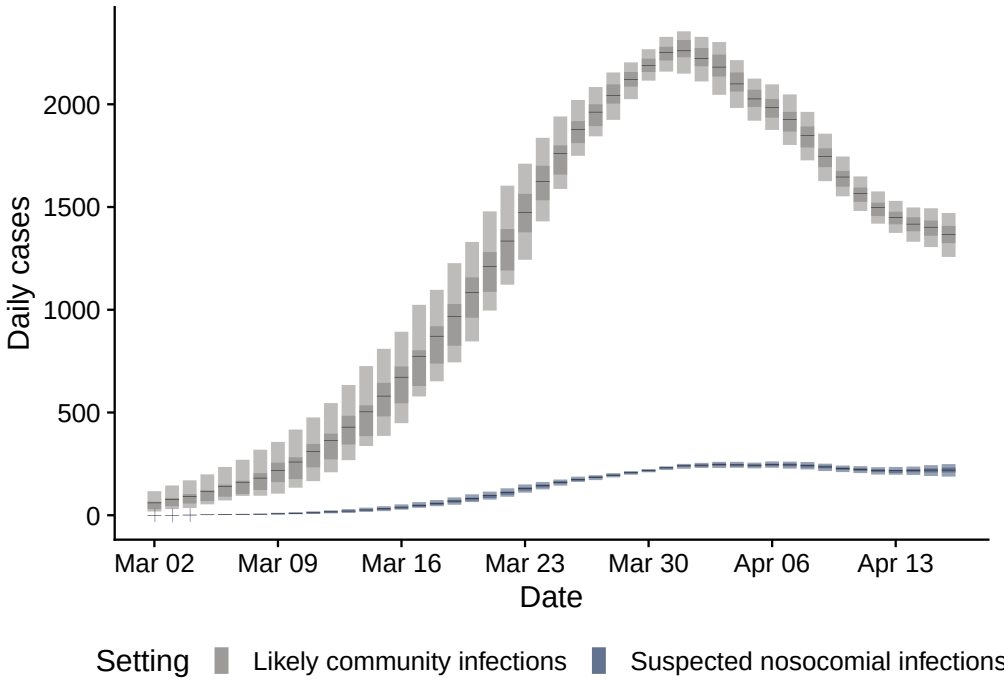

Figure 1: Estimated cases by symptom onset, stratified by suspected nosocomial infections and likely community infections, in England. Light bars = 90% credible interval; dark bars = the 50% credible intervals; black lines indicate the median.

Suspected nosocomial infections

Summary (estimates as of the 2020-04-11)

Table 1: Latest estimates (as of the 2020-04-11) of the number of confirmed cases by date of infection, the expected change in daily confirmed cases, the effective reproduction number, the doubling time, and the adjusted R-squared of the exponential fit. The mean and 90% credible interval is shown for each numeric estimate.

	Estimate
New confirmed cases by infection date	220 (188 – 249)
Expected change in daily cases	Decreasing
Effective reproduction no.	0.1 (0.1 – 0.2)
Doubling time (days)	-6.8 (Inf – Inf)
Adjusted R-squared	0.98 (0.96 – 0.99)

Reported confirmed cases, their estimated date of infection, and time-varying reproduction number estimates

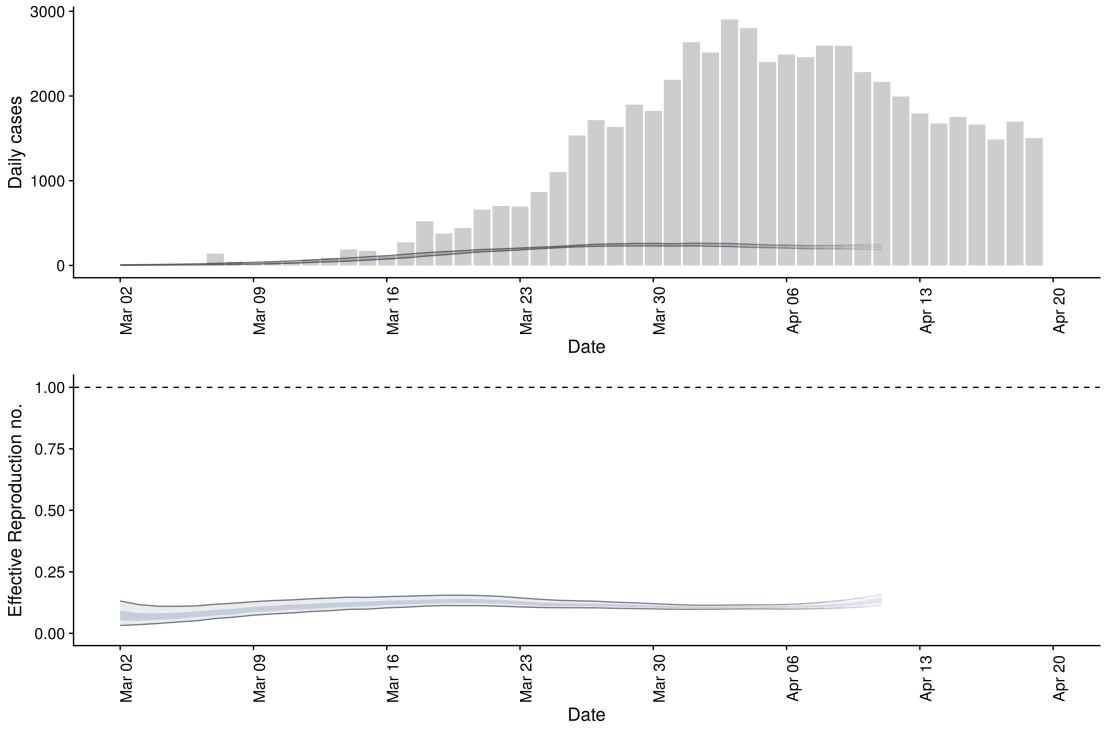

Figure 2: A.) Confirmed cases by date of report (bars) and their estimated date of infection. B.) Time-varying estimate of the effective reproduction number. Light ribbon = 90% credible interval; dark ribbon = the 50% credible interval. Estimates are shown until the 2020-04-11.Dark grey ribbon = 50% credible interval. Confidence in the estimated values is indicated by translucency with increased translucency corresponding to reduced confidence.

Time-varying rate of growth and doubling time

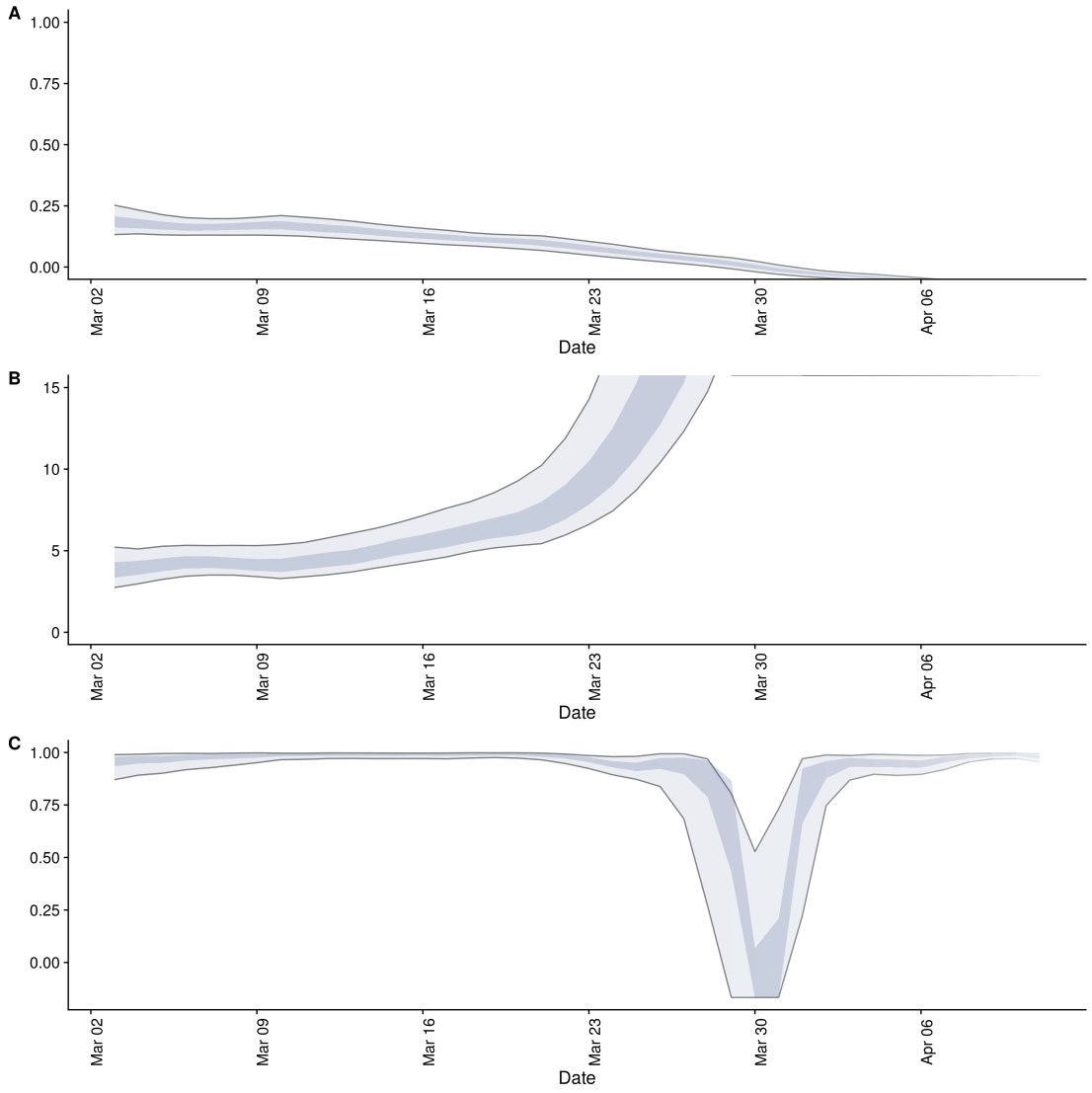

Figure 3: A.) Time-varying estimate of the rate of growth, B.) Time-varying estimate of the doubling time in days (note that when the rate of growth is negative the doubling time is assumed to be infinite), C.) The adjusted R-squared estimates indicating the goodness of fit of the exponential regression model (with values closer to 1 indicating a better fit). Estimates are shown until the 2020-04-11. Light ribbon = 90% credible interval; dark ribbon = the 50% credible interval. Confidence in the estimated values is indicated by translucency with increased translucency corresponding to reduced confidence.

Likely community infections

Summary (estimates as of the 2020-04-11)

Table 2: Latest estimates (as of the 2020-04-11) of the number of confirmed cases by date of infection, the expected change in daily confirmed cases, the effective reproduction number, the doubling time, and the adjusted R-squared of the exponential fit. The mean and 90% credible interval is shown for each numeric estimate.

	Estimate
New confirmed cases by infection date	1368 (1257 – 1471)
Expected change in daily cases	Decreasing
Effective reproduction no.	0.9 (0.8 – 0.9)
Doubling time (days)	-20 (Inf – Inf)
Adjusted R-squared	0.79 (0.56 – 0.98)

Reported confirmed cases, their estimated date of infection, and time-varying reproduction number estimates

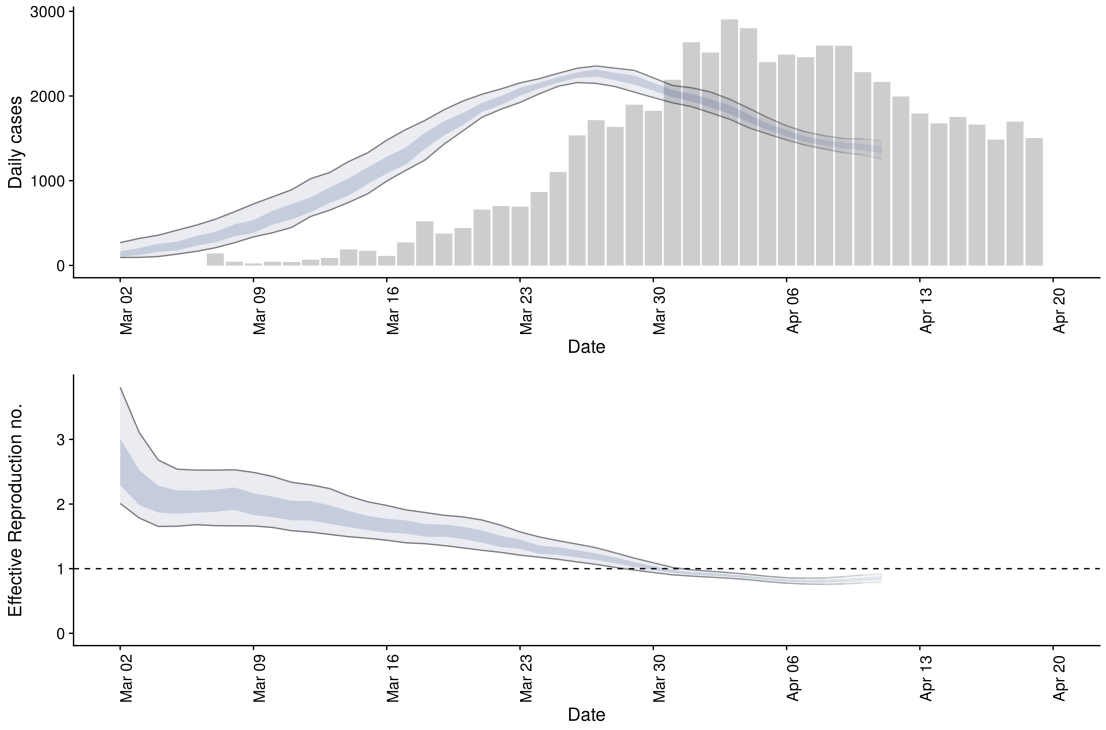

Figure 4: A.) Confirmed cases by date of report (bars) and their estimated date of infection. B.) Time-varying estimate of the effective reproduction number. Light ribbon = 90% credible interval; dark ribbon = the 50% credible interval. Estimates are shown until the 2020-04-11.Dark grey ribbon = 50% credible interval. Confidence in the estimated values is indicated by translucency with increased translucency corresponding to reduced confidence.

Time-varying rate of growth and doubling time

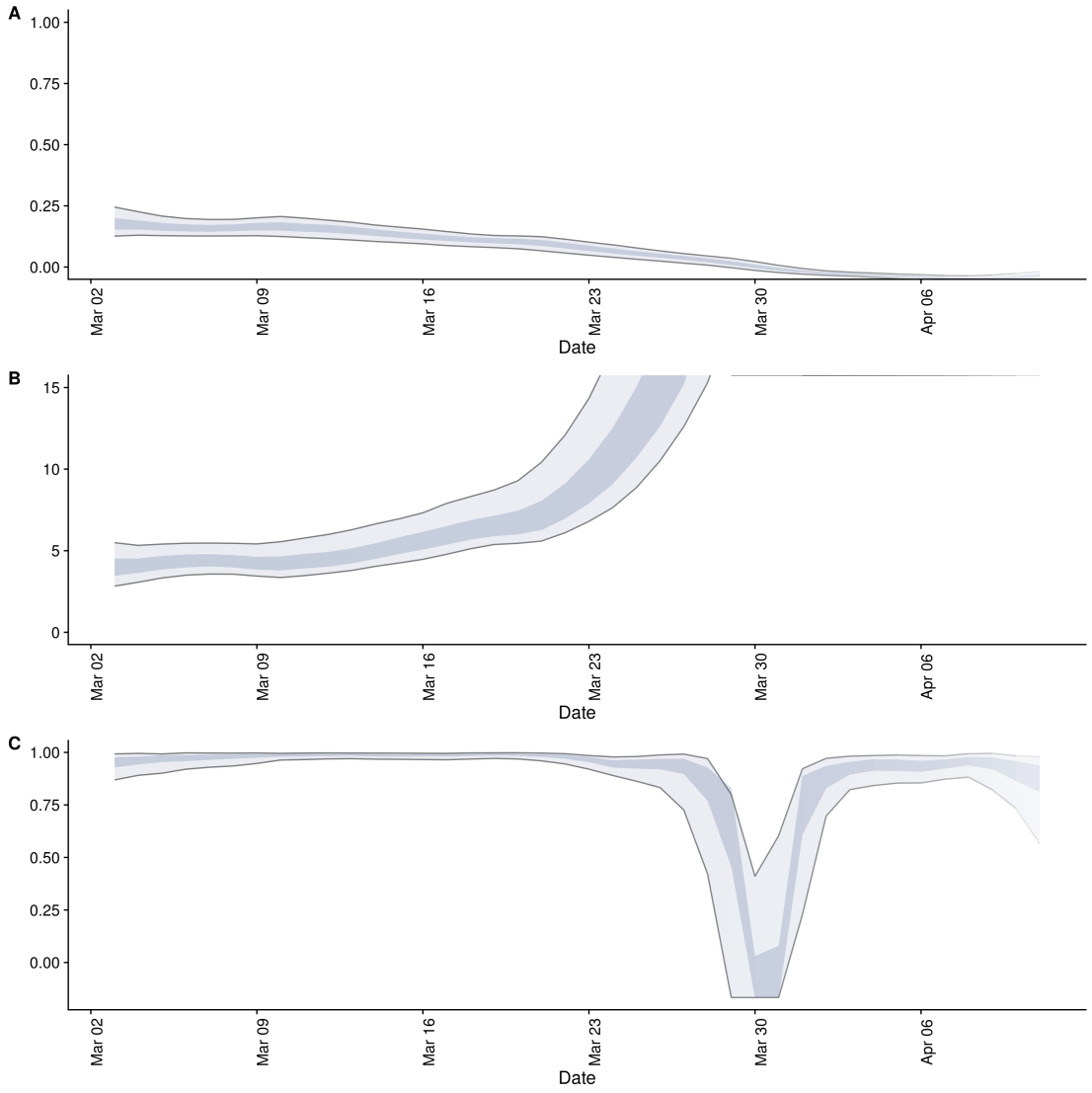

Figure 5: A.) Time-varying estimate of the rate of growth, B.) Time-varying estimate of the doubling time in days (note that when the rate of growth is negative the doubling time is assumed to be infinite), C.) The adjusted R-squared estimates indicating the goodness of fit of the exponential regression model (with values closer to 1 indicating a better fit). Estimates are shown until the 2020-04-11. Light ribbon = 90% credible interval; dark ribbon = the 50% credible interval. Confidence in the estimated values is indicated by translucency with increased translucency corresponding to reduced confidence.

References

Abbott, Sam, Joel Hellewell, James D. Munday, and Sebastian Funk. 2020. "NCoVUtils: Utility Functions for the 2019-Ncov Outbreak." - - (-): –. https://doi.org/10.5281/zenodo.3635417.

Cori, Anne. 2019. *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*. https://CRAN.R-project.org/package=EpiEstim.

Doughty, Emma. 2020. "Visualisation of Covid-19 Official Uk Case Data." Github Repository. https://github.com/emmadoughty/Daily_COVID-19/.

Nishiura, Hiroshi, Natalie M Linton, and Andrei R. Akhmetzhanov. 2020. "Serial Interval of Novel Coronavirus (2019-nCoV) Infections." *medRxiv*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/2020.02.03.20019497.

Thompson, R.N., J.E. Stockwin, R.D. van Gaalen, J.A. Polonsky, Z.N. Kamvar, P.A. Demarsh, E. Dahlqwis, et al. 2019. "Improved Inference of Time-Varying Reproduction Numbers During Infectious Disease Outbreaks." *Epidemics* 29: 100356. https://doi.org/https://doi.org/10.1016/j.epidem.2019.100356.