Estimating effectiveness of frequent PCR testing at different intervals for detection of SARS-CoV-2 infections

Authors: Joel Hellewell[1,†,*], Timothy W. Russell[1,†], The SAFER Investigators and Field Study Team[2], The Crick COVID-19 Consortium[3], CMMID COVID-19 working group[1], Rupert Beale[3], Gavin Kelly[4], Catherine Houlihan[4,5,6], Eleni Nastouli[4,7], Adam J. Kucharski[1]

[1]Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, UK

[2]Department of Clinical Virology, University College London Hospitals, London W1T 4EU, UK

[3]Cell Biology of Infection Laboratory, The Francis Crick Institute; Division of Medicine, UCL, UK

[4]Bioinformatics and Biostatistics, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

[5]Department of Infection and Immunity, University College London, London, UK

[6]Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, UK

[7]Department of Population, Policy and Practice, UCL Great Ormond Street Institute of Child Health, London, WC1N 1EH UK

[†]Authors contributed equally

*Corresponding author: Joel.Hellewell@lshtm.ac.uk

Summary

Using data on twice weekly PCR testing of front-line healthcare workers, we estimated individual infection times and probability of testing PCR positive through time since infection. Our results suggested that PCR-positivity peaked at 4 days, with a peak detection probability of 78% (95% Credible Interval: 55-89%). Using these estimates, we simulated testing strategies and showed that frequent asymptomatic testing can increase the probability of detection early in the infection period.

Main text

Detection of current infection with Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) is a crucial component of targeted policy responses to the COVID-19 pandemic that involve caring for vulnerable groups. For instance, residents and staff in care homes may be tested regularly to minimise outbreaks among elderly populations (1), or healthcare workers (HWCs) may be routinely tested to prevent nosocomial transmission to patients with other comorbidities (2,3). Both of these populations have a substantially higher risk of fatality from COVID-19 infection than the general population (4,5). In the UK, testing commonly uses polymerase chain reaction (PCR) to detect the presence of viral RNA in the nasopharynx of those sampled (6). The sensitivity of these tests depends upon the amount of viral RNA present, which in turn will vary between individuals (7) and with the amount of time that has elapsed between infection and testing (8).

Estimates of temporal variation in PCR sensitivity are crucial for planning effective testing strategies in settings with vulnerable populations. The testing frequency required to detect the majority of infections before they can transmit onwards will depend on both how soon and how long you remain positive by PCR test. Measuring the probability that testing will detect SARS-CoV-2 at a given time-since-infection is challenging for two main reasons. First, it requires knowledge of the timing of infection, which is almost always unobserved. Second, it requires a representative sample of tests done on people with and without symptoms performed at many different times with regards to time since infection. Testing is usually performed on symptomatic infections some time around or just after symptom onset, leading to an unrepresentative sample (9).

To address these challenges, we analyse data that covers the regular testing of healthcare workers (HCWs) in London, UK. We infer their likely time of infection and use the results of the repeated tests performed over the course of their infection to infer the probability of testing positive depending on the time since infection occurred. We therefore overcome the bias towards testing around the time of symptom onset, however we can still only analyse data from symptomatic infections as the timing of symptom onset is used to infer the likely time of infection.

Longitudinal testing of frontline hospital staff in London

We here use data from the SAFER study (10) conducted at University College London Hospitals between 26 March and 5 May 2020, which repeatedly tested 200 patient-facing HCWs by PCR and collected data on COVID-19 symptoms at the time of sampling (10). Samples were tested utilising the pipeline established by the Covid-Crick-Consortium. Individuals were asymptomatic at enrollment and were tested for SARS-CoV-2 antibodies at the beginning and end of the study period. Out of the 200 HCWs enrolled in the study, 46 were seropositive at the first antibody test, 36 seroconverted

over the study period, and 42 returned a positive PCR test at some point during the study (a detailed analysis of the characteristics of this HCW cohort can be found in (10)). Here, we focus on a subset of 27 of these HCWs who seroconverted during the study period, and reported COVID-19 symptoms at one or more sampling times (Figure 1). Combining data on 241 PCR tests performed on self-administered nasopharyngeal samples from these 27 individuals, we estimated the time of infection for each HCW as well as simultaneously estimating the probability of a positive test depending on the time since infection.

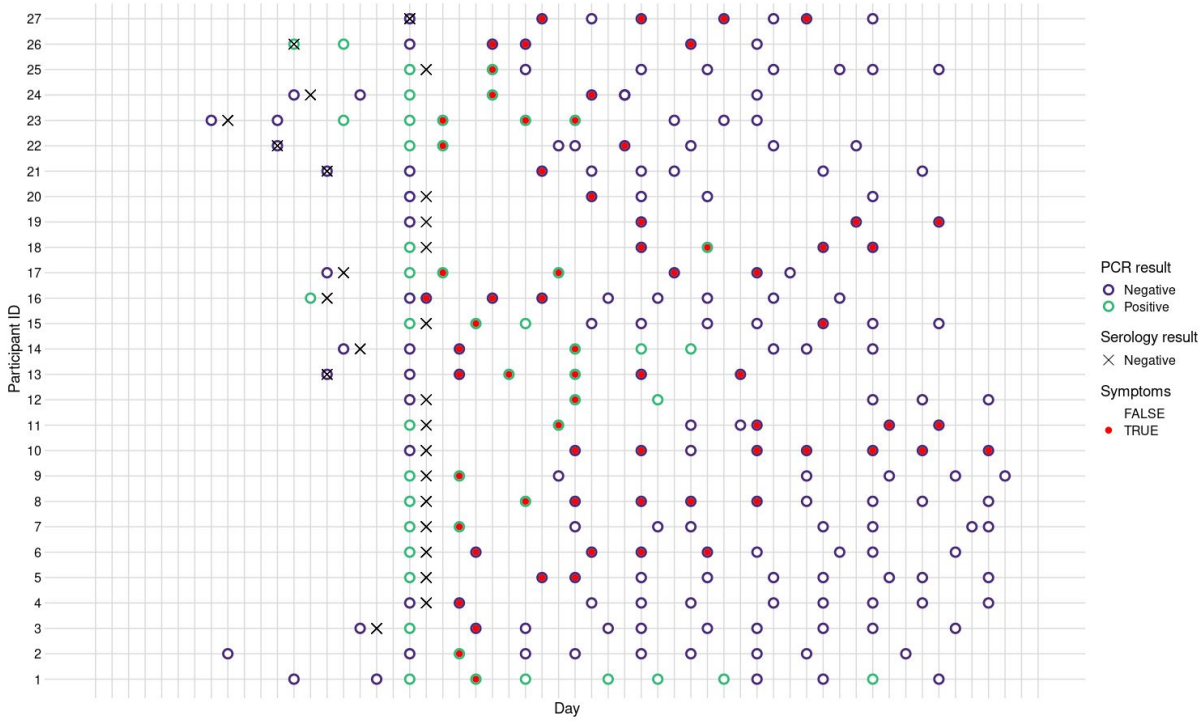

Figure 1: Testing and symptom data for the 27 individuals used in the analysis. Each point represents a symptom report and PCR test result. *Red points indicate a positive PCR result while black points indicate a negative PCR result. If any symptoms were reported, the point is triangular while if no symptoms were reported the point is circular. Green crosses show the date of the initial negative serological test. Points are aligned along the x-axis by the timing of each participant's last asymptomatic report.*

Inference framework

We developed a Bayesian model to jointly infer both the likely infection time for each individual, and the probability of a positive PCR test depending on the time since infection in all individuals. We use a likelihood function specifically for inferring parameters from censored data (11) to derive a posterior distribution of the time of infection. This accounts for the fact that the true onset time is censored, i.e. symptom onset for each individual could have occurred anywhere between their last asymptomatic report and their first symptomatic report, which is necessary as both were reported on a daily timescale. Specifically, individual *i* has their likely infection time, T_i, inferred based on the interval between their last asymptomatic report, t_i^{last} and their first symptomatic report, t_i^{first}. The log-likelihood for the infection time for person i is as follows:

$$\mathcal{L}(T_i \mid t_i^{\text{first}}, t_i^{\text{last}}) = \log(F(t_i^{\text{first}} - T_i) - F(t_i^{\text{last}} - T_i))$$

where F is the cumulative density function of the lognormal distribution that characterises the incubation period of COVID-19 as estimated in (12).

For a given inferred infection time for person i, the relationship between the time since infection and recording a positive n^{th} PCR test on person i, $\mathrm{PCR}^{+}_{n,i}$, administered at time $t_{n,i}$ is given by a piecewise logistic regression model with a single breakpoint:

$$\mathrm{PCR}^{+}_{n,i} \sim \mathrm{Bernoulli}(\mathrm{logit}(\beta_1 + \beta_2 x + \beta_2\beta_3 x I(x))),$$

$$x := t_{n,i} - T_i - C,$$

where C is the time of the breakpoint, x is the amount of time between infection and testing minus the value of the breakpoint, $I(x)$ is a step function that equals 0 if $x < 0$ or equals 1 if $x > 0$, and the β terms define the regression coefficients fit across all tests and people.

To ensure biological plausibility, each individual was assumed to have a negative result at their precise time of infection to constrain the PCR positivity curve to have 0 probability of detection at 0 days since infection. We fitted the model using R 4.0.3 (13) and Stan 2.21.2 (14), the data and the code required to reproduce the figures and results of this study can be found at the public github repository: https://github.com/cmmid/pcr-profile. We ran four MCMC chains for 2000 samples each, discarding the first 1000 samples from each chain as warm-up iterations. Convergence of the chains was assessed using the R-hat statistic being $\hat{\mathrm{R}} < 1.05$ for each model parameter.
We also performed a sensitivity analysis whereby the testing data for one HCW at a time was left out from the model fitting procedure to see if the testing data for any individual HCW had an undue influence on the overall regression fit (results are shown in the Supplementary Material).

Repeated testing model

We looked at two different ways of assessing the performance of different testing frequencies. Firstly, we calculated the probability that a symptomatic case would be detected before symptom onset, this demonstrates the ability of testing to catch infections before people eventually self-isolate due to symptoms (by which point they may already have infected someone).

To calculate the probability that a symptomatic infection is detected prior to symptom onset, let I be the set of the possible testing times for a given test frequency f_x, which given explicitly, can be written as

$$I = \{[0, f_x, 2f_x, \ldots], [1, f_x+1, 2f_x+1, \ldots], \ldots, [f_x-1, 2f_x-1, 3f_x-1 \ldots]\}.$$

The maximum values of $i \in I$ are set at 30 since testing PCR positive 30 days after infection is unlikely.

For the given testing times $i \in I$, if we denote the j_{th} testing time in i as i_j, the number of testing times in i as $|i|$, and d as the delay between test and result, the probability of detecting an infection before symptom onset for testing times i is equal to:

$$P_i = \sum_{j=1}^{|i|} g(i_j)p(i_j - d) \prod_{k=j-1}^{0} \Big((1 - g(i_k))(1 - p(i_k - d)) \Big)$$

Where $g(t)$ is the probability of no onset before time t and $p(t)$ is the probability of a positive test at time t.

Noting that $|I| = f_x$, the probability of detecting a symptomatic infection before symptom onset over all possible testing time variations $i \in I$ is therefore

$$\frac{1}{f_x} \sum_{i \in I} P_i$$

Secondly, we calculated the probability that an asymptomatic case is caught within 7 days of infection, this shows how frequently you need to test to detect asymptomatic infections in a timely manner.

For asymptomatic infections, the value of $g(t) = 1 \ \forall t$ because there will never be an onset time. For detection within seven days we consider

$$I^* = \{[0, f_x, 2f_x, \ldots], [1, f_x{+}1, 2f_x{+}1, \ldots], \ldots, [f_x{-}1, 2f_x{-}1, 3f_x{-}1 \ldots]\}$$

with values up to $7 - d$, since a positive test needs to be performed by this point to be returned within 7 days. For the given testing times $i \in I^*$ the probability of detecting an asymptomatic infection within 7 days is:

$$P_i^* = \sum_{j=1}^{|i|} p(i_j) \prod_{k=j-1}^{0} 1 - p(i_k)$$

And the probability of detecting an asymptomatic infection within 7 days over all testing time variations $i \in I^*$ is equal to:

$$\frac{1}{f_x} \sum_{i \in I^*} P_i^*$$

Results

Timing of infections

The model found that individuals included in this analysis were infected around the beginning of the study period in late March. This corresponds with a period of greatly increased hospitalisation in London, which could potentially mean much higher exposure to infectious COVID-19 patients. However, this analysis cannot say for certain where these HCWs were infected (Figure 2).

Time since infection and PCR positivity

We estimated that the peak probability of a positive PCR test is 77% (54 - 88%) and is observed at 4 days after infection. After that it decreases to 50% (38 - 65%) by 10 days after infection and reaches virtually 0% probability by 30 days after infection (Figure 3A). The posterior probabilities of the piecewise logistic regression parameters are shown in Table 1.

Our testing scenarios suggested that the higher the frequency of testing, the higher the probability that a symptomatic case will be detected before symptom onset (Figure 3B) and the higher the probability that an asymptomatic case is detected within 7 days (Figure 3C).

A 2 day delay between testing and notification compared to a 1 day delay led to reduced probability of detection in both testing scenarios (Figures 3B, 3C). This is because a longer delay means that an infection must be caught earlier to allow for a longer period of time between a test being administered and an infected person being notified of the results. An increased delay from testing to notification caused a greater relative reduction in the probability of detecting an asymptomatic case within 7 days of infection when the testing frequency was lower (Figure 3C).

When considering what is an acceptable testing frequency for detecting a desired proportion of symptomatic cases prior to their symptom onset, there is a trade-off between testing frequency and the delay from testing to notification. For example, the probability of detecting a symptomatic case prior to onset is very similar for a 2 day testing frequency with a 2 day notification delay (41%, 23 - 58%) compared to a 4 day testing frequency with a 1 day notification delay (39%, 22 - 56%). This trade-off is depicted graphically in the dashed black box in Figure 3B.

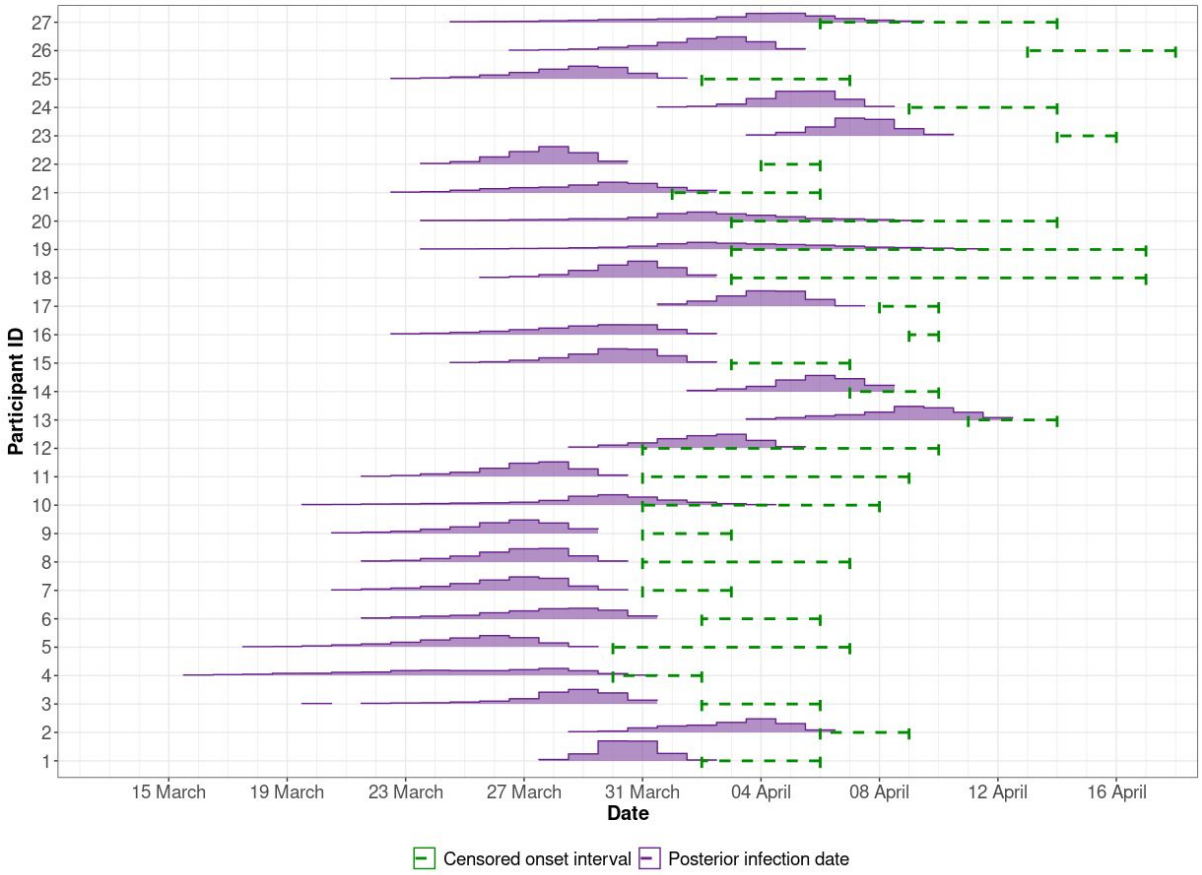

Figure 2: The posterior of the infection time (T_i) of each participant. *The posterior distribution of the infection time for each participant (purple) alongside the censored interval within which their symptom onset occurred (green dashed lines).*

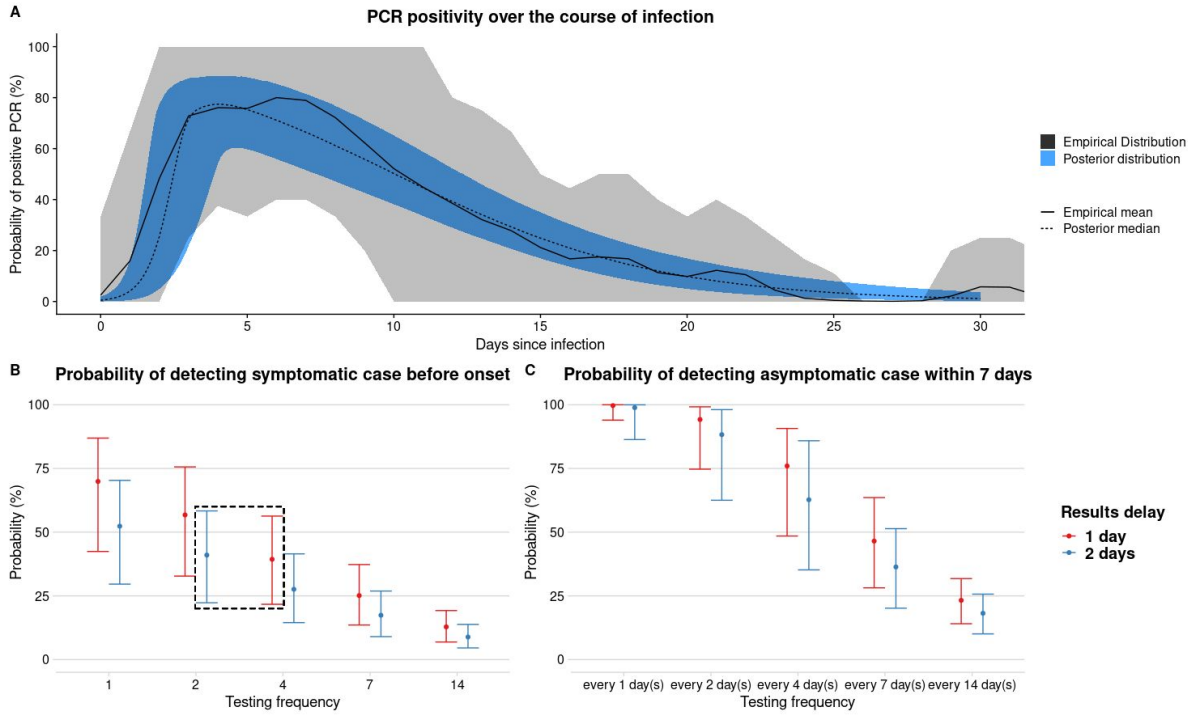

Figure 3: Estimation of positivity over time, and probability that different testing frequencies with PCR would detect virus. *A) Temporal variation in PCR-positivity based on time since infection. The grey interval and solid black line show the 95% uncertainty interval and the mean, respectively, for the empirical distribution calculated from the posterior samples of the times of infection (see Supplementary materials for methodology). The blue interval and dashed black line show the 95% credible interval and median, respectively, of the logistic piecewise regression described above. B) Probability of detecting virus before expected onset of symptoms, based on curve in (A), assuming delay from test to results is either 1 or 2 days. Dashed black box shows a site of possible trade-off between testing frequency and results delay discussed in the text C) Probability of detecting an asymptomatic case within 7 days, based on curve in (A), assuming delay from test to results is either 24 or 48 hours.*

Parameter	Description	Interpretation	Posterior median (95% credible interval)
C	Breakpoint of piecewise regression	The time at which the curve begins to peak	3.18 days post-infection (2.01 - 5.11)
β_1	Intercept of both regression curves	N/A	1.51 (0.80— 2.31)
β_2	Slope of 1st regression curve	The rate of increase in percentage of infections detected after exposure	2.19 (1.26—3.47)
β_3	Slope of 2nd regression curve	The rate of decrease in the percentage of infections detected, after the curve peaks	-1.1 (-1.2—-1.05)

Table 1: Summary of model parameters and the median and 95% credible interval from their fitted posterior distributions.

Discussion

The ongoing COVID-19 pandemic has led to increasing focus on testing strategies that could prevent sustained transmission in hospitals and other defined settings with at-risk individuals, such as care homes. Using data on repeated testing of healthcare workers, we estimated that peak positivity for PCR tests for SARS-CoV-2 infections occurs 4 days after infection, which is just before the average incubation duration, in agreement with other studies finding that viral load in the respiratory tract is highest at this point (8,15).

Previous work looking at PCR positivity since exposure found that the probability of a false negative test decreases from the time of exposure up to around symptom onset, at which point the probability of a false negative test begins to increase again (16). This broadly agrees with our PCR positivity curve, however our PCR positivity curve estimates a far higher probability of testing positive around 1 - 3 days after infection.

Incorporating our estimates of PCR positivity into a model of testing strategies, we found that there is the potential for a trade-off between turnaround time for test results and testing frequency (Example in dashed black box, Figure 3B). This could be particularly relevant for settings that do not have the resources or capacity for very high frequency testing, but could ensure prompt results. Although our analysis focuses on sensitivity of testing, any potential testing and isolation strategy would also need to consider the potential for false positives, particularly at low prevalence (17).

The maximum probability of detection of 77% shown by the curve in Figure 3A refers to the whole population and does not imply that an individual person's peak probability of being detected by a PCR test is 77%. The curve is fitted to combined test results for many individuals, each of whom will

have had variation in the timing of their particular peak probability of detection. This variation is smoothed out over all individuals to lead to the curve shown in Figure 3A.

We assumed that symptoms reported during the study were due to clinical episodes of COVID-19 infection, and not due to other respiratory infections with similar symptoms, but all individuals seroconverted over the course of the study, suggesting that such symptoms were likely to be associated with SARS-CoV-2 infections.

Our analysis is also limited by excluding asymptomatic HCWs that seroconverted over the course of the study. Symptomatic infections may have higher viral loads and be more likely to be detected than asymptomatic infections. Our testing model presents results for detecting asymptomatic infections that relies on the assumption that the PCR positivity curve is the same for symptomatic and asymptomatic infections. If asymptomatic infections are instead less likely to be detected then our probability of detection within 7 days of infection will be an overestimate.

Testing is a crucial component of effective targeted control strategies for COVID-19, and our results suggest that frequent testing and fast turnaround times could substantially increase the probability of detecting infections – and hence prevent outbreaks – early in at-risk settings.

Acknowledgements

The following funding sources are acknowledged as providing funding for the named authors. Wellcome Trust (206250/Z/17/Z: AJK, TWR; 210758/Z/18/Z: JH). AK was supported by the NIHR HPRU in Modelling and Health Economics, a partnership between PHE, Imperial College London and LSHTM (grant code NIHR200908). The views expressed are those of the authors and not necessarily those of the United Kingdom (UK) Department of Health and Social Care, the National Health Service, the National Institute for Health Research (NIHR), or Public Health England (PHE). The SAFER study was funded by MRC UKRI (grant MC_PC_19082) and supported by the UCLH/UCL NIHR BRC..

The following funding sources are acknowledged as providing funding for the working group authors. Alan Turing Institute (AE). BBSRC LIDP (BB/M009513/1: DS). This research was partly funded by the Bill & Melinda Gates Foundation (INV-001754: MQ; INV-003174: KP, MJ, YL; NTD Modelling Consortium OPP1184344: CABP, GFM; OPP1180644: SRP; OPP1183986: ESN; OPP1191821: KO'R, MA). BMGF (OPP1157270: KA). DFID/Wellcome Trust (Epidemic Preparedness Coronavirus research programme 221303/Z/20/Z: CABP, KvZ). DTRA (HDTRA1-18-1-0051: JWR). Elrha R2HC/UK DFID/Wellcome Trust/This research was partly funded by the National Institute for Health Research (NIHR) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care (KvZ). ERC Starting Grant (#757699: JCE, MQ, RMGJH). This project has received funding from the European Union's Horizon 2020 research and innovation programme - project EpiPose (101003688: KP, MJ, PK, RCB, WJE, YL). This research was partly funded by the Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1: AG, CIJ, TJ). HDR UK (MR/S003975/1: RME). MRC (MR/N013638/1: NRW). Nakajima Foundation (AE). NIHR (16/136/46: BJQ; 16/137/109: BJQ, CD, FYS, MJ, YL; Health Protection

Research Unit for Immunisation NIHR200929: NGD; Health Protection Research Unit for Modelling Methodology HPRU-2012-10096: TJ; NIHR200908: RME; NIHR200929: FGS, MJ; PR-OD-1017-20002: AR, WJE). Royal Society (Dorothy Hodgkin Fellowship: RL; RP\EA\180004: PK). UK DHSC/UK Aid/NIHR (ITCRZ 03010: HPG). UK MRC (LID DTP MR/N013638/1: GRGL, QJL; MC_PC_19065: AG, NGD, RME, SC, TJ, WJE, YL; MR/P014658/1: GMK). Authors of this research receive funding from UK Public Health Rapid Support Team funded by the United Kingdom Department of Health and Social Care (TJ). Wellcome Trust (206471/Z/17/Z: OJB; 208812/Z/17/Z: SC, SFlasche; 210758/Z/18/Z: JDM, KS, NIB, SA, SFunk, SRM). No funding (AKD, AMF, AS, CJVA, DCT, JW, KEA, SH, YJ, YWDC).

Centre for Mathematical Modelling of Infectious Disease COVID-19 working group

The following authors were part of the Centre for Mathematical Modelling of Infectious Disease COVID-19 working group. Each contributed in processing, cleaning and interpretation of data, interpreted findings, contributed to the manuscript, and approved the work for publication: Amy Gimma, W John Edmunds, Carl A B Pearson, Kiesha Prem, James D Munday, Katharine Sherratt, Naomi R Waterlow, Graham Medley, Billy J Quilty, Kaja Abbas, Akira Endo, Kathleen O'Reilly, Kevin van Zandvoort, C Julian Villabona-Arenas, Stefan Flasche, Rein M G J Houben, Hamish P Gibbs, Yang Liu, Samuel Clifford, Stéphane Hué, Alicia Rosello, Charlie Diamond, Sam Abbott, Quentin J Leclerc, Alicia Showering, Sophie R Meakin, Gwenan M Knight, Yung-Wai Desmond Chan, Sebastian Funk, Rosalind M Eggo, Thibaut Jombart, Nikos I Bosse, Christopher I Jarvis, Fiona Yueqian Sun, Megan Auzenbergs, Katherine E. Atkins, Rosanna C Barnard, Petra Klepac, Oliver Brady, Anna M Foss, Matthew Quaife, Georgia R Gore-Langton, Frank G Sandmann, James W Rudge, Simon R Procter, Jack Williams, Mark Jit, Arminder K Deol, Damien C Tully, David Simons, Rachel Lowe, Yalda Jafari, Nicholas G. Davies, Emily S Nightingale, Jon C Emery.

SAFER Investigators and Field Study Team

Rebecca Matthews, Abigail Severn, Sajida Adam, Louise Enfield, Angela McBride, Kathleen Gärtner, Sarah Edwards, Fabiana Lorencatto, Susan Michie, Ed Manley, Maryam Shahmanesh, Hinal Lukha, Paulina Prymas, Hazel McBain, Robert Shortman, Leigh Wood, Claudia Davies, Bethany Williams, Emilie Sanchez, Daniel Frampton, Matthew Byott, Stavroula M Paraskevopoulou, Elise Crayton, Carly Meyer, Nina Vora, Triantafylia Gkouleli, Andrea Stoltenberg, Veronica Ranieri, Tom Byrne, Dan Lewer, Andrew Hayward, Richard Gilson, Naomi Walker

Crick COVID-19 Consortium

Aaron Ferron, Aaron Sait, Abhinay Ramaprasad, Abigail Perrin, Adam Sateriale, Adrienne E Sullivan, AileenNelson, Akshay Madoo, Alana Burrell, Aleksandra Pajak, Alessandra Gaiba, Alice Rossi, Alida Avola, Alison Dibbs, Alison Taylor-Beadling, Alize Proust, Almaz Huseynova, Amar Pabari, Amelia Edwards, Amy Strange, Ana Cardoso, Ana Agua-Doce, Ana Perez Caballero, Anabel Guedan, Anastacio King Spert Teixeira, Anastasia Moraiti, Andreas Wack, Andrew Riddell, Andrew Buckton, Andrew Levett, Andrew Rowan, Angela Rodgers, Ania Kucharska, Anja Schlott, Annachiara Rosa, Annalisa D'Avola, Anne O'Garra, Anthony Gait, Antony Fearns, Beatriz Montaner, Belen Gomez Dominguez, Berta Terré Torras, Beth Hoskins, Bishara Marzook, Bobbi Clayton, Bruno Frederico,

Caetano Reis e Sousa, Caitlin Barns-Jenkins, Carlos M Minutti, Caroline Oedekoven, Catharina Wenman, Catherine Lambe, Catherine Moore, Catherine F Houlihan, Charles Swanton, Chelsea Sawyer, Chloe Roustan, Chris Ekin, Christophe Queval, Christopher Earl, Claire Brooks, Claire Walder, Clare Beesley, Claudio Bussi, Clementina Cobolli Gigli, Clinda Puvirajasinghe, Cristina Naceur-Lombardelli, Dan Fitz, Daniel M Snell, Dara Davison, David Moore, Davide Zecchin, Deborah Hughes, Deborah Jackson, Dhruva Biswas, Dimitrios Evangelopoulos , Dominique Bonnet, Edel C McNamara, Edina Schweighoffer, Effie Taylor, Efthymios Fidanis, Eleni Nastouli, Elizabeth Horton, Ellen Knuepfer, Emine Hatipoglu, Emma Russell, Emma Ashton, Enzo ZPoirier, Erik Sahai, Fatima Sardar, Faye Bowker, Fernanda Teixeira Subtil, Fiona McKay, Fiona Byrne, Fiona Hackett, Fiona Roberts, Francesca Torelli, Ganka Bineva-Todd, Gavin Kelly, Gee Yen Shin, Genevieve Barr, George Kassiotis, Georgina H Cornish, Gita Mistry, Graeme Hewitt, Graham Clark, Gunes Taylor, Hadija Trojer, Harriet B Taylor, Hector Huerga Encabo, Heledd Davies, Helen M Golding, Hon-Wing Liu, Hui Gong, Jacki Goldman, Jacqueline Hoyle, James Fleming, James I MacRae, James M Polke, James M A Turner, JaneHughes, Jean D O'Leary, Jernej Ule, Jerome Nicod, Jessica Olsen, Jessica Diring, Jill A Saunders, Jim Aitken, Jimena Perez-Lloret, Joachim M Matz, Joanna Campbell, Joe Shaw, John Matthews, Johnathan Canton, Joshua Hope, Joshua Wright, Joyita Mukherjee, Judith Heaney, Julian AZagalak, Julie A Marczak, Karen Ambrose, Karen Vousden, Katarzyna Sala, Kayleigh Richardson, Kerol Bartolovic, Kevin W Ng, Konstantinos Kousis, Kylie Montgomery, Laura Churchward, Laura Cubitt, Laura Peces-Barba Castano , Laura Reed, Laura E McCoy, Lauren Wynne, Leigh Jones, Liam Gaul, Lisa Levett, Lotte Carr, Louisa Steel, Louise Busby, Louise Howitt, Louise Kiely, Lucia Moreira-Teixeira, Lucia Prieto-Godino, Lucy Jenkins, Luiz Carvalho, Luke Nightingale, Luke Wiliams, Lyn Healy, Magali S Perrault, Malgorzata Broncel, Marc Pollitt, Marcel Levi, Margaret Crawford, Margaux Silvestre, Maria Greco, Mariam Jamal-Hanjani, Mariana Grobler, Mariana Silva Dos Santos, Mark Johnson, Mary Wu, Matthew Singer, Matthew J Williams, Maximiliano G Gutierrez, Melanie Turner, Melvyn Yap, Michael Howell, Michael Hubank, Michael J Blackman, Michael D Buck, Michele S Y Tan, Michelle Lappin, Mimmi Martensson, Ming Jiang, Ming-Han C Tsai, Ming-Shih Hwang, Mint RHtun, Miriam Molina-Arcas, Moira J Spyer, Monica Diaz-Romero, Moritz Treeck, Namita Patel, Natalie Chandler, Neil Osborne, Nick Carter, Nicola O'Reilly, Nigel Peat, Nikhil Faulkner, Nikita Komarov, Nisha Bhardwaj, Nnennaya Kanu, Oana Paun, Ok-Ryul Song, Olga O'Neill, Pablo Romero Clavijo, Patrick Davis, Patrick Toolan-Kerr, Paul Nurse, Paul Kotzampaltiris, Paul C Driscoll, Paul R Grant, Paula Ordonez Suarez, Peter Cherepanov, Peter Ratcliffe, Philip Hobson, Philip A Walker, Pierre Santucci, Qu Chen, Rachael Instrell, Rachel Ambler, Rachel Ulferts, Rachel Zillwood, Raffaella Carzaniga, Rajnika Hirani, Rajvee Shah Punatar, Richard Byrne, Robert Goldstone, Robyn Labrum, Ross Hall, Rowenna Roberts, Roy Poh, Rupert Beale, Rupert Faraway, Sally Cottrell, Sam Barrell, Sam Loughlin, Samuel McCall, Samutheswari Balakrishnan, Sandra Segura-Bayona, Savita Nutan, Selvaraju Veeriah, Shahnaz Bibi, Sharon P Vanloo, Simon Butterworth, Simon Caidan, Solene Debaisieux, Sonia Gandhi, Sophia Ward, Sophie Ridewood, Souradeep Basu, Stacey-Ann Lee, Steinar Halldorsson, Stephanie Nofal, Steve Gamblin, Steve Hindmarsh, Stuart Kirk, Subramanian Venkatesan, Sugera Hashim, Susanne Herbst, Suzanne Harris, Svend Kjaer, Tammy Krylova, Tea Toteva, Theo Sanderson, Theresa Higgins, Thomas Martinez, Timothy Budd, Tom Cullup, Venizelos Papayannopoulos, Vicky Dearing, Vijaya Ramachandran, Wai Keong Wong, Wei-Ting Lu, Yiran Wang, Yogen Patel, Zena Collins, Zheng Xiang, Zoe Allen, Zoe H Tautz-Davis

Conflict of interest

None declared

Authors' contributions

AJK conceived of the study and planned the inference framework with feedback from JH and TWR. JH and TWR ran exploratory data analysis and implemented the model. JH and TWR wrote the manuscript with feedback from all authors.

Code availability.

All of the data and the code required to reproduce the figures and results of this study can be found at the public github repository: https://github.com/cmmid/pcr-profile.

References

1. Vivaldi 1: COVID-19 care homes study report [Internet]. GOV.UK. [cited 2020 Nov 10]. Available from: https://www.gov.uk/government/publications/vivaldi-1-coronavirus-covid-19-care-homes-study-report/vivaldi-1-covid-19-care-homes-study-report
2. Rickman HM, Rampling T, Shaw K, Martinez-Garcia G, Hail L, Coen P, et al. Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital. Clin Infect Dis [Internet]. [cited 2020 Nov 10]; Available from: https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa816/5860253
3. Taylor J, Rangaiah J, Narasimhan S, Clark J, Alexander Z, Manuel R, et al. Nosocomial COVID-19: experience from a large acute NHS Trust in South-West London. J Hosp Infect. 2020 Nov;106(3):621–5.
4. Poletti P, Tirani M, Cereda D, Trentini F, Guzzetta G, Marziano V, et al. Age-specific SARS-CoV-2 infection fatality ratio and associated risk factors, Italy, February to April 2020. Eurosurveillance. 2020 Aug 6;25(31):2001383.
5. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020 Jun 1;20(6):669–77.
6. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med. 2020 Mar 19;382(12):1177–9.
7. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. BMJ [Internet]. 2020 Apr 21 [cited 2020 Nov 10];369. Available from: https://www.bmj.com/content/369/bmj.m1443
8. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020 May;26(5):672–5.
9. COVID-19 testing data: methodology note [Internet]. GOV.UK. [cited 2020 Nov 10]. Available from: https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note
10. Houlihan CF, Vora N, Byrne T, Lewer D, Kelly G, Heaney J, et al. Pandemic peak SARS-CoV-2 infection and seroconversion rates in London frontline health-care workers. The Lancet. 2020 Jul 25;396(10246):e6–7.
11. Delignette-Muller M, Dutang C. fitdistrplus: An R Package for Fitting Distributions. J Stat Softw Artic. 2015;64(4):1--34.
12. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020 Mar 10;172(9):577–82.
13. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; YEAR. Available from:

https://www.R-project.org
14. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. J Stat Softw Artic. 2017;76(1):1–32.
15. Singanayagam A, Patel M, Charlett A, Bernal JL, Saliba V, Ellis J, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Eurosurveillance. 2020 Aug 13;25(32):2001483.
16. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. Ann Intern Med [Internet]. 2020 May 13 [cited 2020 Nov 10]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7240870/
17. Surkova E, Nikolayevskyy V, Drobniewski F. False-positive COVID-19 results: hidden problems and costs. Lancet Respir Med [Internet]. 2020 Sep 29 [cited 2020 Nov 10];0(0). Available from: https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30453-7/abstract

Supplementary Materials

Empirical distribution of PCR positivity

The grey interval in Figure 3A is calculated from the posterior samples of the likely infection time for each individual (T_i). If we let the j_{th} posterior sample of T_i be denoted T_{ij} then we calculate d_{ij}, the time from infection until each test ($t_{n,i}$) performed on individual i, for each sample T_{ij}

$$d_{ij} = t_{n,i} - T_{ij}$$

Each d_{ij} is rounded to the nearest discrete day and for each MCMC iteration (j) we calculate the proportion of tests with $d_{ij} = k$ that were positive for each discrete day k since infection, denoted $p_j(k)$. We then calculate the mean and 95% uncertainty intervals of $p_j(k)$ for each day k over all MCMC samples j. This can be considered a graphical representation of the “data” that the PCR positivity regression is fit to (the precise values rely on the infection time draws at each iteration of the MCMC).

Sensitivity Analysis

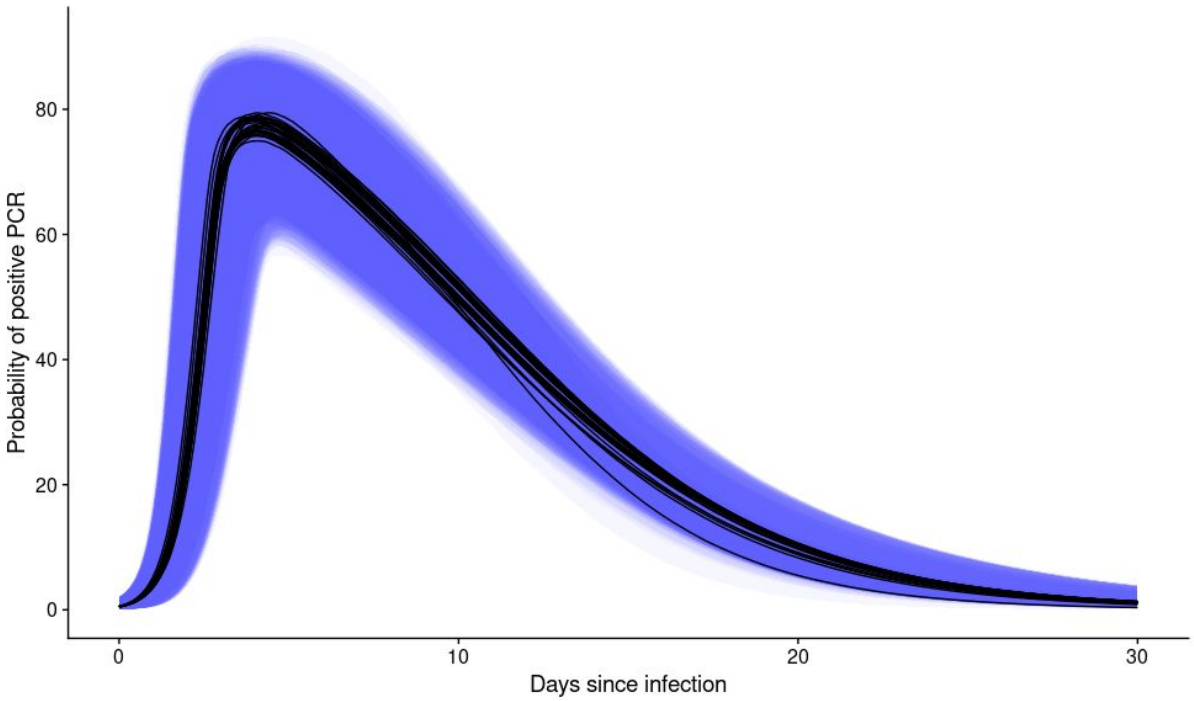

Figure S1: Multiple PCR positivity curves superimposed on top of each other, each curve shows the fitted PCR positivity curve while leaving out data for a different one of the 27 individuals in the data set each time. There is one curve whereby the median posterior probability is around 5% lower from ~12 days after infection onwards if data for an individual is excluded. This suggests that one individual out of the 27 HCWs continued to test positive for a long time after their inferred infection date, which could possibly bias our PCR positivity upwards slightly towards the tail of the distribution.