SARS-CoV-2 infection and reinfection in a seroepidemiological workplace cohort in the United States

Emilie Finch†, Rachel Lowe†, Stephanie Fischinger2,3, Michael de St Aubin4, Sameed M. Siddiqui5,6, Diana Dayal7, Michael A. Loesche7,8, Justin Rhee7, Samuel Berger7, Yiyuan Hu7, Matthew J. Gluck7, Benjamin Mormann7, Mohammad A. Hasdianda8, Elon R. Musk7, Galit Alter2,6, Anil S. Menon7*, Eric J. Nilles4,8*, Adam J. Kucharski1* on behalf of the CMMID COVID-19 working group and the SpaceX COVID-19 Cohort Collaborative

1Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine
2Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA
3Institut für HIV Forschung, Universität Duisburg-Essen, Duisburg, Germany
4Harvard Humanitarian Initiative, Cambridge, MA, USA
5Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA, USA
6Broad Institute of MIT and Harvard, Cambridge, MA, USA
7Space Exploration Technologies Corp
8Brigham and Women’s Hospital, Boston

† Author for correspondence at emilie.finch1@lshtm.ac.uk
* Equal contribution

Abstract

Identifying the extent of SARS-CoV-2 reinfection is crucial for understanding possible long-term epidemic dynamics. We analysed longitudinal PCR and serological testing data from a prospective cohort of 4411 US employees in four states between April 2020 and February 2021. We conducted a multivariable logistic regression investigating the association between baseline serological status and subsequent PCR test result in order to calculate an odds ratio for reinfection. We estimated an adjusted odds ratio of 0.09 (95% CI: 0.005 – 0.48) for reinfection, implying that the presence of SARS-CoV-2 antibodies at baseline is associated with around 91% reduced odds of a subsequent PCR positive test. This suggests that primary infection with SARS-CoV-2 provides protection against reinfection in the majority of individuals, at least over a sixth month time period. We also highlight two major sources of bias and uncertainty to be considered when estimating reinfection risk, confounders and the choice of baseline time point, and show how to account for both in our analysis.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Introduction

The rapid global spread of COVID-19 throughout 2020 occurred as a result of the introduction of a highly transmissible virus, SARS-CoV-2, into populations with little pre-existing immunity [1]. Identifying the extent and duration of protective immunity afforded by natural infection (and by vaccination) is therefore of crucial importance for understanding possible long-term epidemic dynamics of SARS-CoV-2 [2].

Studies have estimated that over 95% of symptomatic COVID-19 cases develop antibodies against SARS-CoV-2, with most individuals developing antibodies within three weeks of symptom onset [3,4]. Several serological studies have also characterised individual-level immune dynamics, with some finding evidence for antibody waning and others for sustained antibody responses over several months [5–10]. Antibody kinetics are thought to vary between individuals and are possibly associated with severity of illness, where asymptomatic or mildly symptomatic individuals may develop lower levels of antibodies that wane more rapidly [3,7,11]. While neutralising antibodies are thought to be associated with protection from reinfection, there is currently no established correlate between antibody levels and protective immunity, and the impact of putative antibody waning on reinfection risk is still unclear.

Confirmed cases of reinfection with SARS-CoV-2 have been reported since August 2020 [12]. However, existing large studies examining the relative risk of reinfection in antibody positive individuals have typically involved specific cohorts who may not be representative of the wider community, such as closed communities or healthcare worker cohorts [13–16]. To evaluate the risk of SARS-CoV-2 infection and reinfection over time, we analysed PCR and serological testing data from a prospective cohort of SpaceX employees in the USA between April 2020 and February 2021 [17].

Results

Of 4411 individuals enrolled, 309 individuals tested seropositive during the study period (Figure 1). This resulted in an overall adjusted seroprevalence of 8.2% (95% CI: 7.3–9.1). We defined a possible reinfection as a new positive PCR test more than 30 days after initial seropositive result. This identified 14 possible reinfections with a median time of 66.5 days between initial seropositive test and PCR positive test.
Risk of COVID-19 infection and reinfection

We estimated the odds ratio for SARS-CoV-2 reinfection using multivariable logistic regression. This required us to choose a cut-off week in order to define baseline seroprevalence and the subsequent observation period for PCR testing. To examine how our estimate for the odds ratio for reinfection varied depending on the cut-off week chosen, we repeated the analysis using every possible cut-off week.

We estimated an adjusted odds ratio of 0.09 (95% CI: 0.005–0.48) for reinfection, with the week of 26th July 2020 as the optimal baseline time point (Figure 2), as this week fell after the first wave of infection and had the most precise odds ratio estimate. Odds ratio estimates using cut-off weeks in between the two waves of infection (between mid-July and mid-September 2020) ranged from 0.09 (95% CI: 0.005–0.48) to 0.25 (95% CI: 0.037–1.01).
Figure 2: A) PCR positivity in the cohort between 5th April 2020 and 31st January 2021. B) Percentage ever seropositive in the cohort (number ever seropositive/ cumulative number enrolled) between 29th March 2020 and 23rd August 2020. Note that the percentage ever positive decreases initially as participants continue to be enrolled in the study. C) Number of possible reinfections in cohort over time (defined as a new positive PCR test more than 30 days after initial seropositive result). D) Odds ratio estimates comparing odds of reinfection in the seropositive group with odds of primary infection in the seronegative group, estimated using logistic regression and adjusted for race, ethnicity, state, job category and BMI. The estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline seroprevalence on the x-axis.
Unadjusted odds ratio estimates tended to overestimate the odds ratio for reinfection compared with primary infection, particularly when using early cut-off weeks (Figure 3). Notably, with early cut-off weeks the unadjusted analysis estimated a higher odds of reinfection compared to primary infection, albeit with wide confidence intervals. We hypothesise that individuals who are at higher risk of seroconversion (who would be included in analyses at earlier time thresholds) would also be at higher risk of later reinfection, giving a biased estimate of the effect of antibodies on subsequent infection for earlier cut-off weeks.

Figure 3: A) Unadjusted odds ratio estimates comparing odds of reinfection in the seropositive group with odds of primary infection in the seronegative group. The estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline seroprevalence on the x-axis. B) Odds ratio estimates comparing odds of reinfection in the seropositive group with odds of primary infection in the seronegative group, estimated using logistic regression and adjusted for race, ethnicity, state, job category and BMI. The estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline seroprevalence on the x-axis.

Simulation using known underlying risk of infection

To further investigate how the precision of odds ratios estimates for reinfection varies depending on population-level epidemic dynamics, we conducted a simulation analysis using a known underlying infection risk distribution.

We used a probability distribution for risk of infection derived from PCR testing data from the study cohort. This was then scaled so that the overall cumulative risk reflected the level of seropositivity in the cohort by the end of the study period (8%). Considering a sample size of 2000 individuals over a period of 44 weeks, and a pre-set probability of reinfection given seropositivity of 0.15, we then re-estimated the corresponding risk ratios (Figure 4).
When considering this two-wave epidemic scenario, we found that the uncertainty in the estimated risk of reinfection was reduced in the middle of the simulation period (i.e. between the two ‘waves’ of infection risk). This supports our choice of optimal cut-off week for estimating the relative risk of reinfection in our study cohort, where we considered odds ratios given the need for an underlying regression analysis.

![Figure 4: A) Cumulative risk of infection in the cohort used for simulation analysis. B) Risk ratio estimates comparing the risk of reinfection with the risk of primary infection. The estimates are presented with their associated 95% confidence intervals and with the cut-off week used to define baseline infection status on the x-axis. The dashed line represents the pre-set reinfection risk ratio of 0.15.](image)

**Discussion**

We identified 14 possible reinfections out of 309 seropositive individuals in this prospective seroepidemiological cohort. We estimated an adjusted odds ratio of 0.09 (95% CI: 0.005–0.48) for reinfection, with the week of 26th July 2020 as the optimal baseline time point in a study period between April 2020 and February 2021. This suggests that the presence of SARS-CoV-2 antibodies at baseline is associated with around 91% reduced odds of a subsequent PCR positive test. This provides further evidence that primary infection with SARS-CoV-2 results in protection against reinfection in the majority of individuals, at least over a sixth month time period. Our findings are consistent with estimates of 0.17 (95% CI 0.13–0.24) odds ratio [13] and 0.11 (0.03–0.44) incidence rate ratio [14] for healthcare workers, 0.18 (0.11–0.28) incidence rate ratio for military recruits [15] and 0.195 (95% CI 0.155–0.246) incidence rate ratio from a Danish population-level study [18].
Our analysis addressed two key sources of bias and uncertainty in estimating reinfection risk. First, confounders may inflate estimates; if a specific subset of the cohort is at higher risk of infection (e.g. due to underlying health conditions or increased risk of exposure), these participants will be more likely to be both initially seropositive and to have a subsequent reinfection. Second, the time period considered could increase uncertainty; defining the baseline seroprevalence at an early time point means few will be seropositive, whereas defining it at a later point means there is less time to observe possible reinfections. We accounted for these two factors by first using logistic regression to identify predictors of baseline seropositivity (i.e. infection risk), to calculate an adjusted odds ratio for reinfection. We then performed a sensitivity analysis to identify the optimum cut-off date to define baseline seroprevalence. We found that for a two-wave epidemic scenario, a cut-off week in the period in between the two waves of infection risk led to the most precise estimates of the odds ratio for reinfection. This demonstrates that the uncertainty surrounding estimates of reinfection risk will be sensitive to the study period chosen, relative to population-level epidemic dynamics.

There are several limitations to the underlying data that should be considered when interpreting these findings. This prospective cohort was recruited opportunistically from employees at one US company and is unlikely to be representative of the general population. However, as we did not identify any workplace outbreaks, transmission in this cohort is likely to be more reflective of community transmission than in health-care worker cohorts or other specialised populations. Additionally, we only considered possible reinfections (as opposed to probable or true reinfections). As possible reinfections did not meet a stringent case definition, such as confirmation through genomic sequencing, they may include cases of prolonged viral shedding following an initial infection. This would result in an overestimation of the odds ratio for reinfection and so our analysis reflects the minimum possible effect of antibodies on future SARS-CoV-2 infection risk.

As well as quantifying reinfection risk over a six-month period among a prospectively followed workplace population, our study highlights the importance of accounting for both individual-level heterogeneity in infection risk and population-level variation in epidemic dynamics when assessing the potential for reinfections.

**Methods**

**Seroepidemiological cohort description**

We used data from a seroepidemiological study of US employees at SpaceX, also described elsewhere [17]. In brief, this study involved 4411 employees from California, Florida, Texas and Washington State, with ages ranging from 18-71. All employees were invited to participate by email and there were no exclusion criteria. Study participants were offered SARS-CoV-2 IgG receptor-binding domain (RBD) antibody testing with an in-house ELISA assay with 82.4% sensitivity and 99.6% specificity [19]. Serological samples were taken during four rounds of testing between April - September 2020. A questionnaire including demographic, symptom and exposure information was conducted at enrolment, and with each round of serological testing.
Individuals continued to be enrolled throughout the study period, and around half of the total participants (48%) were tested at more than one time point. Participants occupied a range of job positions within SpaceX including office-based and factory-based jobs. Additionally, symptomatic and asymptomatic PCR testing were widely available for employees, with data available from April 2020 - January 2021. Both serology and PCR testing data were available for 1800 individuals.

**Statistical Analysis**

To estimate the odds ratio for SARS-CoV-2 reinfection, we conducted multivariable logistic regression analysis investigating the association between baseline serological status and subsequent PCR test result, given a test was sought. Propensity to seek a PCR test did not appear to differ by baseline serostatus.

The choice of cut-off week used to define participants’ baseline seroprevalence and the subsequent observation period for PCR testing has important implications for the precision of the estimated odds ratio for reinfection. For instance, a cut-off week early in the study period will result in few seropositive individuals, while a cut-off week later in the study period leaves less time to observe subsequent PCR testing and detect possible reinfections. To identify the optimal cut-off week and assess how the choice of cut-off week affected estimates of the odds ratio for reinfection, we repeated the multivariable logistic regression for every possible cut-off week.

Potential confounding variables were selected a priori and included; age, sex, race, ethnicity, BMI, state, work location, job category, household size, history of chronic disease or history of smoking. We used logistic regression to identify predictors of baseline seropositivity (i.e. infection risk) using a forwards modelling strategy comparing the Akaike Information Criterion (AIC). Finally, we selected the optimally sized variable set to adjust for by comparing the AIC between models for the same cut-off week for different variable sets. Odds ratio estimates were adjusted for race, ethnicity, state, job category and BMI.

**Simulation analysis**

We then conducted a simulation analysis to further investigate how the precision of odds ratio estimates for reinfection vary depending on the time point chosen to determine participants’ baseline seroprevalence and the subsequent window to monitor PCR infections and possible reinfections.

We simulated a two-wave epidemic scenario using a probability distribution for cumulative risk of infection derived from PCR testing data from the study cohort. This was scaled so the overall cumulative risk of infection reflected the overall level of seropositivity in the cohort (8%). We considered a sample size of 2000 individuals over a period of 44 weeks. For each time point, the number of seropositive and seronegative individuals up to and including week was simulated, according to the background cumulative risk of infection. We then simulated how
many of these seropositive and seronegative individuals would test PCR positive or negative after week, with a pre-set probability of reinfection of 0.15.

Finally, we re-estimated the risk ratio for reinfection for each cut-off time point and assessed the precision by comparing the estimated risk ratio over time with the ‘true’ ratio of 0.15.

Analysis was conducted in R version 4.0.3. Code to reproduce the figures and simulation analysis presented here can be found at https://github.com/EmilieFinch/covid-reinfection.

Acknowledgments

Funding: AJK was supported by Wellcome Trust (grant: 206250/Z/17/Z) and NIHR (NIHR200908). EF was funded by the Medical Research Council (grant number MR/N013638/1), RL was funded by a Royal Society Dorothy Hodgkin Fellowship. The work was also supported by the Translational Research Institute for Space Health through NASA Cooperative Agreement (NNX16AO69A), the Massachusetts Consortium on Pathogen Readiness (MassCPR), the NIH (3R37AI080289-11S1, R01AI146785, U19AI42790-01, U19AI135995-02, U19AI42790-01, 1U01CA260476-01), and the Musk Foundation.

We thank SpaceX employees Lindsay Chapman, Jordan Steinhart, Suzanne Siebert, and Kyle Meade for their valuable support, in addition to the dedication and commitment of the many SpaceX employees that volunteered to participate in this study.

Competing interests: GA is a founder of Seromyx Systems Inc., a company developing platform technology that describes the antibody immune response. GA’s interests were reviewed and are managed by Massachusetts General Hospital in accordance with their conflict-of-interest policies. MJG, SB, DD, YH, JR, EP, BM, ASM, and ERM are employees of Space Exploration Technologies Corp. All other authors have declared that no conflict of interest exists.

CMMID COVID-19 working group information
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: Jiayao Lei, Sebastian Funk, Fiona Yueqian Sun, Amy Gimma, Emily S Nightingale, Graham Medley, Sam Abbott, Fabienne Krauer, Nicholas G. Davies, Mark Jit, Akira Endo, Oliver Brady, Anna M Foss, Yung-Wai Desmond Chan, Thibaut Jombart, Kevin van Zandvoort, Rosalind M Eggo, Yang Liu, Gwenaen M Knight, Carl A B Pearson, Kaja Abbas, Katherine E. Atkins, Samuel Clifford, Mihaly Kolta, Yalda Jafari, Damien C Tully, Christopher I Jarvis, Kathleen O'Reilly, Nikos I Bosse, Kiesha Prem, Billy J Quilty, Simon R Procter, Rosanna C Barnard, William Waite, Ciara McCarthy, James D Munday, David Hodgson, W John Edmunds, Alicia Rosello, C Julian Villabona-Arenas, Hamish P Gibbs, Stefan Flasche, Timothy W Russell, Sophie R Meakin, Joel Hellewell, Naomi R Waterlow, Matthew Quaife, Frank G Sandmann.
Funding statements for the CMMID COVID-19 working group are as follows. KvZ: KvZ is supported by the UK Foreign, Commonwealth and Development Office (FCDO)/Wellcome Trust Epidemic Preparedness Coronavirus research programme (ref. 221303/Z/20/Z), and Elrha’s Research for Health in Humanitarian Crises (R2HC) Programme, which aims to improve health outcomes by strengthening the evidence base for public health interventions in humanitarian crises. The R2HC programme is funded by the UK Government (FCDO), the Wellcome Trust, and the UK National Institute for Health Research (NIHR). SC: Wellcome Trust (grant: 208812/Z/17/Z). SFys: NIHR EPIC grant (16/137/109). SFunk: Wellcome Trust (grant: 210758/Z/18/Z), NIHR (NIHR200908). GFM: NTD Modelling Consortium by the Bill and Melinda Gates Foundation (OPP1184344). YJ: LSHTM, DHSC/UKRI COVID-19 Rapid Response Initiative. SRM: Wellcome Trust (grant: 210758/Z/18/Z). WJE: European Commission (EpiPose 101003688), NIHR (NIHR200908). MQ: European Research Council Starting Grant (Action Number #757699); Bill and Melinda Gates Foundation (INV-001754). NRw: Medical Research Council (grant number MR/N013638/1). RME: HDR UK (grant: MR/S003975/1), MRC (grant: MC_PC 19065), NIHR (grant: NIHR200908). NGD: UK Research and Innovation; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK MRC (MC_PC_19065). JYL: Bill & Melinda Gates Foundation (INV-003174). M.K: Foreign, Commonwealth and Development Office / Wellcome Trust. FK: Innovation Fund of the Joint Federal Committee (Grant number 01VSF18015). Wellcome Trust (UNS110424). DCT: No funding declared. AS: No funding declared. AMF: No funding declared. KP: Gates (INV-003174), European Commission (101003688). SFlasche: Wellcome Trust (grant: 208812/Z/17/Z). SA: Wellcome Trust (grant: 210758/Z/18/Z). BJQ: This research was partly funded by the National Institute for Health Research (NIHR) (16/137/109 & 16/136/46) 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. AS: No funding declared. TJQ: This research was partly funded by the National Institute for Health Research (NIHR) (16/137/109 & 16/136/46) 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. AS: No funding declared. TM: No funding declared. TWR: Wellcome Trust (grant: 206250/Z/17/Z). CiJ: Global Challenges Research Fund (GCRF) project ‘RECAP’ managed through RCUK and ESRC (ES/P010873/1). SRP: Bill and Melinda Gates Foundation (INV-016832). AE: The Nakajima Foundation. ESN: Gates (OPP1183986). NIB: Health Protection Research Unit (grant code NIHR200908). CJVA: European Research Council Starting Grant (Action number 757688). FGS: NIHR Health Protection Research Unit in Modelling & Health Economics, and in Immunisation. AG: European Commission (EpiPose 101003688). KA: Bill & Melinda Gates Foundation (OPP1157270, INV-016832). WW: MRC (grant MR/V027956/1). KEA: European Research Council Supporting Grant (Action number 757688). RCB: European Commission (EpiPose 101003688). PK: This research was partly funded by the Royal Society under award RP/EA/180004, European Commission (101003688), Bill & Melinda Gates Foundation (INV-003174). HPG: This research was
produced by CSIGN which is part of the EDCTP2 programme supported by the European Union (grant number RIA2020EF-2983-CSIGN). The views and opinions of authors expressed herein do not necessarily state or reflect those of EDCTP. This research is funded by the Department of Health and Social Care using UK Aid funding and is managed by the NIHR. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health and Social Care (PR-OD-1017-20001). CABP: CABP is supported by the Bill & Melinda Gates Foundation (OPP1184344) and the UK Foreign, Commonwealth and Development Office (FCDO)/Wellcome Trust Epidemic Preparedness Coronavirus research programme (ref. 221303/Z/20/Z). OJB: Wellcome Trust (grant: 206471/Z/17/Z). WJE: European Commission (EpiPose 101003688), NIHR (NIHR200908, MRC: MC_PC_19065). CIJ: Global Challenges Research Fund (GCRF) project ‘RECAP’ managed through RCUK and ESRC (ES/P010873/1). NGD: UKRI Research England; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK MRC (MC_PC_19065).

References


