medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Viral load in community SARS-CoV-2 cases varies widely and temporally

A. Sarah Walker[1,2,3,4], Emma Pritchard[1,2], Thomas House[5,6], Julie V Robotham[2,7], Paul J Birrell[7,8], Iain Bell[9], John I Bell[10], John N Newton[11], Jeremy Farrar[12], Ian Diamond[9], Ruth Studley[9], Jodie Hay[13,14], Karina-Doris Vihta[1,2], Tim Peto[1,2,3,15], Nicole Stoesser*[1,2,3,15], Philippa C. Matthews*[1,15], David W. Eyre*[1,2,14,16], Koen B. Pouwels[1,2,17] and the COVID-19 Infection Survey team

[*] contribution considered equal

[1] Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2] The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford, Oxford, UK.

[3] The National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

[4] MRC Clinical Trials Unit at UCL, UCL, London, UK

[5] Department of Mathematics, University of Manchester, Manchester, UK

[6] IBM Research, Hartree Centre, Sci-Tech Daresbury, UK

[7] National Infection Service, Public Health England, London, UK

[8] MRC Biostatistics Unit, University of Cambridge, Cambridge Institute of Public Health, Cambridge, UK

[9] Office for National Statistics, Newport, UK

[10] Office of the Regius Professor of Medicine, University of Oxford, Oxford, UK

[11] Health Improvement Directorate, Public Health England, London, UK

[12] Wellcome Trust, London, UK

[13] University of Glasgow, Glasgow, UK.

[14] Lighthouse Laboratory in Glasgow, Queen Elizabeth University Hospital, Glasgow

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[15] Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK

[16] Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, UK

[17] Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK

See Acknowledgements for the Coronavirus Infection Survey team

Keywords: SARS-CoV-2, viral load, symptoms, clearance, community

Running title: Community SARS-CoV-2 positives

Corresponding author

A. Sarah Walker, sarah.walker@ndm.ox.ac.uk

Alternate corresponding author

Koen Pouwels, koen.pouwels@ndph.ox.ac.uk

SUMMARY (38 words)

Cycle threshold (Ct) values in community SARS-CoV-2 infections from national surveillance vary markedly, including over time and by symptoms (1892 (0.22%) positive nose and throat swabs, 843,851 tested). Ct values could be a useful epidemiological early warning indicator.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT (250 words)

Background: Information on COVID-19 in representative community surveillance is limited, particularly regarding cycle threshold (Ct) values (a proxy for SARS-CoV-2 viral load) and symptoms.

Methods: We included all positive nose and throat swabs between 26 April-11 October 2020 from the UK's national COVID-19 Infection Survey, tested by RT-PCR for the N, S and ORF1ab genes. We investigated predictors of median Ct value using quantile regression.

Results: 1892(0.22%) of 843,851 results were positive, 1362(72%), 185(10%) and 345(18%) for 3, 2 or 1 genes respectively. Ct for different genes were strongly correlated (rho=0.99) with overall median Ct 26.2 (IQR 19.7-31.1; range 10.3-37.6), corresponding to ~2,500 dC/ml (IQR 80-240,000). Ct values were independently lower in those reporting symptoms, with more genes detected, and in first (vs. subsequent) positives per-participant, with no evidence of independent effects of sex, ethnicity, age, deprivation or other test characteristics (p>0.20). Whilst single-gene positives without reported symptoms almost invariably had Ct>30, triple-gene positives without reported symptoms had widely varying Ct. Incorporating pre-test probability and Ct values, 1547(82%) and 112(6%) positives had "higher" or "lower" supporting evidence for genuine infection. Ct values, symptomatic percentages and supporting evidence changed over time. With lower positivity in the summer, there were proportionally more "lower" evidence positives, and "higher" evidence positives had higher Ct values (p<0.0001), suggesting lower viral burden. Declines in mean/median Ct values were apparent throughout August and preceded increases in positivity rates.

Conclusions: Community SARS-CoV-2 infections show marked variation in viral load. Ct values could be a useful epidemiological early-warning indicator.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2969 words (limit 3000)

INTRODUCTION

After initial reductions in SARS-CoV-2 cases in mid-2020, following release of large-scale lockdowns[1], infection rates are re-surging in many countries worldwide. Proposed control strategies include new local or national lockdowns of varying intensity and mass testing, but have major economic and practical limitations. In particular, mass testing of large numbers without symptoms[2], and hence low pre-test probability of positivity, can mean most positives are false-positives depending on test specificity. For example, with 0.1% true prevalence, testing 100,000 individuals with a 99.9% specific test with perfect sensitivity gives 100 true-positives, but also 100 false-positives (positive predictive value (PPV) 50%), whereas specificity of 99.5% increases false-positives to 500 (PPV=17%), and of 99.0% to 999 (PPV=9%), with even lower PPV with imperfect sensitivity[3].

Mathematical models are powerful tools for evaluating the potential effectiveness of different control strategies, but rely on population-level estimates of infectivity and other parameters. However, there are few unbiased community-based surveillance studies, including individuals both with and without symptoms. Estimates of asymptomatic infection rates vary, being only 17-20% overall in recent reviews[4, 5], but these included many studies of contacts of confirmed cases. Higher prevalence of asymptomatic infection has been reported in screening of defined populations (30%[4]) and community surveillance (e.g. 42%[6], 72%[7]). Studies have generally indicated lower rates of transmission from asymptomatic infection[4, 5]; this may be a proxy for SARS-CoV-2 viral load, as a key determinant of transmission. Finally, most studies rely on “average” estimates of the asymptomatic infection percentage, independent of characteristics and viral load, and have not

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

quantified temporal variation in these key parameters for mathematical models across the community.

Here we therefore characterise variation in SARS-CoV-2 positive tests in the first five months of the UK's national COVID-19 Infection Survey (CIS), which is based on a representative sample of households with longitudinal follow-up. We estimate predictors of RT-PCR cycle threshold (Ct) values (as a proxy for viral load), propose a classification for the strength of evidence supporting positive RT-PCR test results in the community, and demonstrate how this has changed over time.

METHODS

This study included all positive SARS-CoV-2 RT-PCR results between 26 April and 11 October 2020 from nose and throat swabs taken from participants in the Office for National Statistics (ONS) CIS (ISRCTN21086382). The survey randomly selects private households on a continuous basis from address lists and from previous surveys to provide a representative UK sample. If anyone aged 2 years or older currently resident in an invited household agreed verbally to participate, a study worker visited the household to take written informed consent, which was obtained from parents/carers for those 2-15 years; those aged 10-15 years provided written assent. The study protocol is available at https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets.

Individuals were asked about demographics, symptoms, contacts and relevant behaviours (https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/case-record-forms). To reduce transmission risks, self-taken nose and throat swabs were obtained following study worker instructions. Parents/carers took swabs from children under 12 years. At the first visit, participants were asked for (optional) consent for follow-up visits every week for the next month, then monthly

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

for 12 months from enrolment. In a random 10-20% households, those 16 years or older were invited to provide blood monthly for assays of anti-trimeric spike protein IgG using an immunoassay developed by the University of Oxford[8]. The study received ethical approval from the South Central Berkshire B Research Ethics Committee (20/SC/0195).

Swabs were analysed at the UK's national Lighthouse Laboratories at Milton Keynes (National Biocentre) (from 26 April) and Glasgow (from 16 August) using identical methodology, with swabs from specific regions sent consistently to one laboratory. RT-PCR for three SARS-CoV-2 genes (N protein, S protein and ORF1ab) used the Thermo Fisher TaqPath RT-PCR COVID-19 Kit, analysed using UgenTec Fast Finder 3.300.5 (TaqMan 2019-nCoV Assay Kit V2 UK NHS ABI 7500 v2.1). The Assay Plugin contains an Assay specific algorithm and decision mechanism that allows conversion of the qualitative amplification Assay PCR raw data from the ABI 7500 Fast into test results with minimal manual intervention. Samples are called positive in the presence of at least single N gene and/or ORF1ab but may be accompanied with S gene (1, 2 or 3 gene positives). S gene is not considered a reliable single gene positive (as of mid-May 2020).

Twelve specific symptoms were elicited at each visit (cough, fever, myalgia, fatigue, sore throat, shortness of breath, headache, nausea, abdominal pain, diarrhoea, loss of taste, loss of smell), as was whether participants thought they had (unspecified) symptoms compatible with COVID-19. From 26 April through 22 July, questions referred to current symptoms, and from 23 July to the preceding 7 days. Any positive response to any symptom question at the swab-positive visit defined the case as symptomatic "at" the test; we also separately defined any positive response at the swab-positive visit or visits either side (regardless of time between visits) as symptomatic "around" the test.

To investigate the potential increasing contribution of false-positives as population prevalence declines, from 2 August we arbitrarily classified in real-time each positive as:

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

- "Higher" evidence: two or three genes detected (irrespective of Ct).
- "Moderate" evidence: single-gene detected and (i) Ct below the 97.5th percentile of "higher" evidence positives (<34; supporting this threshold, whole genome sequences had been obtained from three single gene positives with Ct 30.8-33.1 by 2 August) or (ii) higher pre-test probability of infection, defined as any symptoms at/around the test or reporting working in a patient-facing healthcare or care/residential home.
- "Lower" evidence: all other positives; by definition single-gene detected at Ct≥34 in individuals not reporting symptoms/working in relevant roles.

We assessed independent predictors of Ct values using median (quantile) regression (details in Supplementary Methods).

RESULTS

Number and percentage of positive swabs

From 26 April to 11 October 2020, 232,189 participants from 116,008 households had results from median 3 (IQR 2-5, range 1-13) nose and throat swabs each (53,517(23%) had only one result to date). 21,747(9%), 16,186(7%), 17,348(7%), 37,356(16%), 95,184(41%) and 44,368(19%) were recruited in April/May, June, July, August, September and October respectively. Of 843,851 swab test results, 1892 (0.224%, 95% CI 0.214-0.234%) were positive, in 1,516 individuals from 1,209 households. Of these participants, 625(41%) were positive at their first test in the study and 891 (59%) subsequently, after median 2 negative tests (IQR 1-4, range 1-9).

Viral characteristics

Overall, 1362(72%), 185(10%) and 345(18%) swabs were positive for three, two or one gene(s) respectively (**Table 1**). Where multiple genes were detected, the Cts were highly correlated

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(Spearman rho=0.99, p<0.0001). Taking the per-swab mean Ct across positive genes, the overall median Ct was 26.2 (IQR 19.7-31.1; range 10.3-37.6), varying strongly by number of genes detected (Kruskal-Wallis p=0.0001), but not by their specific pattern after adjusting for number (p=0.58). Based on linearity data (**Supplementary Figure 1**), this corresponds to a median viral load of ~2,500 dC/ml (IQR 80-240,000). Only four Ct values >37 were recorded (one S positive only (May); three N positive only (October)).

Evidence supporting positive results

1547(82%), 233(12%) and 112(6%) positive tests had "higher", "moderate" or "lower" evidence supporting genuine positivity (**Table 2**; definitions in Methods). Even though "higher" evidence was based only on number of genes detected, "higher" evidence positives were more likely to be symptomatic than "moderate" evidence positives (p<0.0001), but were similarly likely to have occupational risk factors (p=0.98). "Higher" evidence positives were more likely to occur in households with other positives (p<0.0001).

Predictors of Ct values

In multivariable regression models, Ct values were independently lower (i.e. viral loads higher) with more genes detected (10.4 lower in triple-gene vs single-gene positives (95% CI 11.3 to 9.5)), if symptoms were reported around the test (1.7 lower (2.4 to 1.0)), and at the first (vs. subsequent) positive per-participant (2.8 lower (3.7 to 2.0) all p<0.0001; **Supplementary Table 1A**), with by far the strongest effect associated with triple-gene positives. There was weak evidence of lower values in those reporting cough/fever/anosmia (2.1 lower (3.0 to 1.3)) vs. other symptoms (1.1 (2.1 to 0.0); heterogeneity p=0.06). Associations were similar but slightly attenuated for symptoms at the positive test. After adjusting for these factors, there was no evidence of independent effects of sex, ethnicity, age, deprivation or whether the positive result was the first test in the study (p>0.20, **Supplementary Table 1A**). Even after adjusting for number of genes detected, symptoms and first (vs. subsequent)

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

positive, Ct values were 1.7 (2.4 to 1.0) lower in individuals where another household member was positive at any point in the study (p<0.0001; other effects similar).

However, number of genes detected and symptoms are both potential mediators of effects of demographic factors (**Supplementary Figure 2**); and there were strong effects of calendar time on positivity in different demographic subgroups[9] and on Ct values (**Figure 1A**). Excluding the potential mediators (number of genes detected, symptoms) and adjusting for visit date, Ct values remained independently lower (i.e. viral loads higher) at the first (vs. subsequent) positive in the study, but were also lower in those in the most deprived quintile (2.6 lower vs least deprived (4.2 to 1.1) p=0.001) and when the positive was not the first test in the study (1.4 lower (2.5 to 0.2), p=0.02) (**Supplementary Table 1B**). However, there was still no evidence of effects of sex (p=0.31), age (p=0.24) or ethnicity (p=0.14).

Of note, whilst single-gene positives without reported symptoms almost invariably had Ct>30, triple-gene positives without reported symptoms had widely varying Ct (**Figure 1A**). Further, whilst the percentage reporting symptoms increased linearly as Ct values dropped from 35 to 25, below 25, the percentages reporting symptoms stayed approximately constant (**Figure 2**).

Temporal changes in Ct values, evidence and symptomatic percentages

There were also strong effects of calendar time on proportions with any evidence of symptoms, or reporting cough/fever/anosmia (**Figure 1C**), and strength of supporting evidence (**Figure 1D**; all p<0.0001), with markedly fewer positives with Ct <30 (**Figure 1B**), very low percentages with symptoms at/around positive tests, and more "lower" evidence positives in July/August. However, during this period, even "higher" evidence positives had higher Ct vs. earlier and later (p=0.0001; heterogeneity; **Figure 3A**). "Lower" evidence positives also formed a larger percentage of all tests during this period (0.017%; **Figure 3B**) and April/early May (when swabs with only the S gene

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

detected were called positive) (0.039%). However, interestingly, from September, the percentage of "lower" evidence positives increased proportionately with "moderate" and "higher" evidence positives (**Figure 3B**).

Relationship with serostatus

Antibody results were available for 88(6%) participants with positive swabs, but relatively few both before and after the first positive swab (**Supplementary Figure 3**). The majority (45/59, 76%) of those with antibody results before the first swab-positive were antibody-negative (**Supplementary Figure 4**), but seven participants appeared to have become infected despite antecedent high anti-spike antibody titres (**Figure 4**): five single positives without reported symptoms (three "higher", two "lower" evidence), one with three positives with symptoms ("higher" evidence") and one with two positives without reported symptoms ("higher" evidence) more than two months apart and separated by four negative intervening RT-PCR swabs. Plausible seroconversion events were seen in three RT-PCR positive cases (**Figure 4**), with no evidence of seroconversion in 15; eight "higher", two "moderate" and four "lower" evidence positives without reported symptoms, and one "moderate" evidence symptomatic positive (**Figure 4**). The remainder had unclear patterns or insufficient data (**Supplementary Figure 5**).

DISCUSSION

In this large community surveillance study, we found wide variation in Ct values (a proxy for viral load). Whilst Ct values were independently associated with symptoms at/around the test, as previously reported[10, 11], and with the number of genes detected, there was no evidence of association with sex, ethnicity or age (as previously reported for sex and age[12, 13]) and effects of symptoms were small compared with population-level variability. Notably triple-gene positives without reported symptoms had widely varying Ct, including many low levels (**Figure 1A**), potentially

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

explaining variation in dispersion ("k") and super-spreading events, particularly from those without symptoms but with low Ct/high viral loads[14].

Ct values also varied strongly over time, as did symptoms and evidence supporting positives, suggesting changing viral burden in infection cases, with less severe phenotypes during July/early August 2020. This strongly refutes hypotheses that viral fitness has declined. During this time, higher Ct values were also noted in the English point-prevalence surveillance study, REACT[7], and lower virus levels in Lausanne, Switzerland[13]. However, Ct values were higher even in "higher" evidence positives during this period, consistent with shifting viral burden. Such a shift may also explain the preceding shift towards "moderate" evidence positives and the concurrent higher percentage of "lower" evidence positives, since the less virus present, the less likely it is to be detected on multiple genes. Whilst these findings are consistent with lower viral inoculum during this period[15], we cannot assess whether this is predominantly due to behaviour (e.g. increased time outdoors, face mask use[16]) or other reasons (e.g. environmental/climatic factors).

We used laboratory, clinical and demographic evidence to classify our confidence in positive results. Around 80% had 2 or 3 genes detected ("higher" evidence), providing assurance in overall results, and all but four Ct values were under 37. Whilst Ct values are not directly comparable between studies, REACT has also validated a Ct threshold of 37 for single-gene positives for their test performed in Germany[17], and in the Public Health England (PHE) Schools study, only samples with Ct<37 were positive on repeat testing of the same swab at PHE laboratories[18]. However, every diagnostic test has false-positives, and so some of our single gene "lower", or even "moderate", evidence positives are inevitably false. However, the false-positive rate would be expected to be approximately constant over time, since it is either random or driven by external factors. Variation in the percentage of all tests accounted for by "lower" evidence positives, and in particular proportionate increases in "lower" evidence positives as "higher" evidence positives increased during

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

September supports more genuinely lower-level infections occurring during the summer, and an overall false-positive rate for this test of below ~0.005% i.e. at least 99.995% specificity.

Since RT-PCR and antigen assays test for viral presence, it is more relevant to consider limits of detection, rather than "false-positives" per se. Although they were a small minority (6%), one question is whether single-gene positives with high Ct (≥34 in our study) solely represent long-term shedding of non-transmissible virus[19], with, for example, infectious virus recovered from only 8% (95% CI 3-18%) of samples with Ct>35 in a PHE study[20] and studies reporting no growth of virus for Ct thresholds from >24 to >34 or higher[21]. Whilst we have not directly assessed household transmission in this study, it was notable that Ct values were significantly lower in positives where anyone else in the same household was ever positive, supporting a role for greater within-household transmission with lower Ct values. Ct values were 1.4 higher in those positive at their first study test (where long-term shedders would be expected to be overrepresented), but these formed only 33% of the positives.

Although numbers are small, our evaluation of serological responses is one of few in the community to our knowledge, and highlights that a significant proportion of these RT-PCR-positive cases do not appear to seroconvert. Unfortunately whole genome sequence data was not available to confirm potential re-infections in a small proportion of individuals, but one case had "higher evidence" positive tests spanning four negative swabs with a long sampling interval (>30 days) between positive swabs (number 21, **Figure 4**), and six cases had positive swabs after negative swabs on a background of high anti-spike IgG titres (Figure 4). Presumed re-infections have been reported elsewhere[22], including in individuals without previous functional and/or durable antibody responses[23, 24], and may remain relevant to virus transmission, whether they occur with or without symptoms. Our data suggest that these may occur in the presence of anti-spike antibodies, which correlate with neutralising antibody titres. These antibody titres are unlikely to have been false-positives, given the

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

context, persistence, and known diagnostic and analytical specificity of the assay[8], or to all reflect laboratory identifier errors, but further studies are clearly needed.

A major study strength is its design, being a large-scale community survey. However, this is also a limitation, since we were not able to comprehensively characterise individual positives. We may have underestimated the initial prevalence of symptoms due to originally asking about current symptoms, although this was predominantly at the earliest weekly visits (so only very transient symptoms between visits would have been missed). Similar rates of symptom reporting in the first and last third of the study suggests that this question was likely generously interpreted in any case. We made no attempt to collect additional information on symptoms after positives were identified to minimise recall bias. This may partly explain why we observed higher rates of positive tests without reported symptoms than recent reviews[4, 5]; however, many studies in these reviews tested close contacts of index cases identified through symptoms and therefore might plausibly have higher viral loads.

Ultimately the importance of asymptomatic and low virus level infections depends on their transmissibility and their prevalence; regardless of limitations in symptom ascertainment, infection without recognition has the potential for onward transmission and unascertained infections are likely critical for avoiding resurgence after lifting lockdown[25]. Our findings support the use of Ct values and genes detected more broadly in public testing programmes, predominantly testing symptomatic individuals and case contacts, as an “early warning” system for shifts in potential infectious load and hence transmission, and hence the risks posed by individuals to others. This has recently also been proposed on the basis of theoretical work linking effective reproduction numbers to population level Ct[26]. For example, declines in mean and median Ct values were apparent throughout August (**Figure 1B**), although positivity rates in the survey were only noted to increase in early September[9]. Ct data are widely available within laboratory management systems and could be used alongside

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

available risk factor and symptom information to facilitate more informed and effective individual-level and public health responses to the SARS-CoV-2 pandemic.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

FUNDING

This study is funded by the Department of Health and Social Care. ASW, EP, JVR, TEAP, NS, DE, KBP are supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England (PHE) (NIHR200915). ASW and TEAP are also supported by the NIHR Oxford Biomedical Research Centre. EP and KBP are also supported by the Huo Family Foundation. ASW is also supported by core support from the Medical Research Council UK to the MRC Clinical Trials Unit [MC_UU_12023/22] and is an NIHR Senior Investigator. PCM is funded by Wellcome (intermediate fellowship, grant ref 110110/Z/15/Z) and holds an NIHR BRC Senior Fellowship award. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, Department of Health, or PHE.

CONFLICTS OF INTEREST

DWE declares lecture fees from Gilead, outside the submitted work. No other author has a conflict of interest to declare.

DATA AVAILABILITY

De-identified study data are available for access by accredited researchers in the ONS Secure Research Service (SRS) for accredited research purposes under part 5, chapter 5 of the Digital Economy Act 2017. For further information about accreditation, contact Research.Support@ons.gov.uk or visit the SRS website.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ACKNOWLEDGEMENTS

Office for National Statistics: Iain Bell, Ian Diamond, Alex Lambert, Pete Benton, Emma Rourke, Stacey Hawkes, Sarah Henry, James Scruton, Peter Stokes, Tina Thomas.

Office for National Statistics, Analysis: John Allen, Russell Black, Heather Bovill, David Braunholtz, Dominic Brown, Sarah Collyer, Megan Crees, Colin Daglish, Byron Davies, Hannah Donnarumma, Julia Douglas-Mann, Antonio Felton, Hannah Finselbach, Eleanor Fordham, Alberta Ipser, Joe Jenkins, Joel Jones, Katherine Kent, Geeta Kerai, Lina Lloyd, Victoria Masding, Ellie Osborn, Alpi Patel, Elizabeth Pereira, Tristan Pett, Melissa Randall, Donna Reeve, Palvi Shah, Ruth Snook, Ruth Studley, Esther Sutherland, Eliza Swinn, Heledd Thomas, Anna Tudor, Joshua Weston.

Office for National Statistics, Secure Research Service: Shayla Leib, James Tierney, Gabor Farkas, Raf Cobb, Folkert van Galen, Lewis Compton, James Irving, John Clarke, Rachel Mullis, Lorraine Ireland, Diana Airimitoaie, Charlotte Nash, Danielle Cox, Sarah Fisher, Zoe Moore, James McLean, Matt Kerby.

University of Oxford, Nuffield Department of Medicine: Ann Sarah Walker, Derrick Crook, Philippa C Matthews, Tim Peto, Emma Pritchard, Nicole Stoesser, Karina-Doris Vihta, Alison Howarth, George Doherty, James Kavanagh, Kevin K Chau, Stephanie B Hatch, Daniel Ebner, Lucas Martins Ferreira, Thomas Christott, Brian D Marsden, Wanwisa Dejnirattisai, Juthathip Mongkolsapaya, Sarah Hoosdally, Richard Cornall, David I Stuart, Gavin Screaton.

University of Oxford, Nuffield Department of Population Health: Koen Pouwels.

University of Oxford, Big Data Institute: David W Eyre.

University of Oxford, Radcliffe Department of Medicine: John Bell.

Oxford University Hospitals NHS Foundation Trust: Stuart Cox, Kevin Paddon, Tim James.

University of Manchester: Thomas House.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Public Health England: John Newton, Julie Robotham, Paul Birrell.

IQVIA: Helena Jordan, Tim Sheppard, Graham Athey, Dan Moody, Leigh Curry, Pamela Brereton.

National Biocentre: Ian Jarvis, Kirsty Howell, Bobby Mallick, Phil Eeles.

Glasgow Lighthouse Laboratory: Jodie Hay, Harper Vansteenhouse.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature **2020**; 584(7820): 257-61.

2. Yokota I, Shane PY, Okada K, et al. Mass screening of asymptomatic persons for SARS-CoV-2 using saliva. Clinical Infectious Diseases, **2020**: ciaa1388.

3. Adams ER, Ainsworth M, Anand R, et al. . Antibody testing for COVID-19: A report from the National COVID Scientific Advisory Panel [version 1; peer review: 1 approved]. Wellcome Open Res, **2020**; 5:139.

4. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLoS Med **2020**; 17(9): e1003346.

5. Byambasuren O, Bell K, Clark J, McLaws M, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. MedRxiv **2020**; https://www.medrxiv.org/content/10.1101/2020.05.10.20097543v3.full.pdf.

6. Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature **2020**; 584(7821): 425-9.

7. Riley S, Ainslie KEC, Eales O, et al. Resurgence of SARS-CoV-2 in England: detection by community antigen surveillance. MedRxiv **2020**; https://www.medrxiv.org/content/10.1101/2020.09.11.20192492v2.

8. National Sars-CoV-Serology Assay Evaluation Group. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. Lancet Infect Dis **2020**.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

9. Pouwels KB, House T, Pritchard E, et al. Community prevalence of SARS-CoV-2 in England during April to September 2020: Results from the ONS Coronavirus Infection Survey MedRxiv **2020**.

10. Edwards T, Santos VS, Wilson AL, et al. Variation of SARS-CoV-2 viral loads by sample type, disease severity and time: a systematic review. MedRxiv **2020**: https://www.medrxiv.org/content/10.1101/2020.09.16.20195982v1.full.pdf.

11. Lee S, Kim T, Lee E, et al. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. JAMA Intern Med **2020**.

12. Jones TC, Mühlemann B, Veith T, et al. An analysis of SARS-CoV-2 viral load by patient age. MedRxiv **2020**; https://www.medrxiv.org/content/10.1101/2020.06.08.20125484v1.

13. Jacot D, Greub G, Jaton K, Opota O. Viral load of SARS-CoV-2 across patients and compared to other respiratory viruses. MedrRxiv **2020**: https://www.medrxiv.org/content/10.1101/2020.07.15.20154518v2.full.pdf.

14. Endo A, Centre for the Mathematical Modelling of Infectious Diseases C-WG, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res **2020**; 5: 67.

15. Gandhi M, Beyrer C, Goosby E. Masks Do More Than Protect Others During COVID-19: Reducing the Inoculum of SARS-CoV-2 to Protect the Wearer. J Gen Intern Med **2020**.

16. Gandhi M, Rutherford GW. Facial Masking for Covid-19 - Potential for "Variolation" as We Await a Vaccine. N Engl J Med **2020**.

17. Riley S, Ainslie KEC, Eales O, et al. Transient dynamics of SARS-CoV-2 as England exited national lockdown. MedRxiv **2020**: https://www.medrxiv.org/content/10.1101/2020.08.05.20169078v1.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

18. Ladhani S, Baawuah F, Beckmann J, et al. Prospective active national surveillance of preschools and primary schools for SARS-CoV-2 infection and transmission in England, June 2020 (sKIDs COVID-19 surveillance in school KIDs) Phase 1 Report, **2020**.

19. Moraz M, Jacot D, Papadimitriou-Olivgeris M, et al. Clinical importance of reporting SARS-CoV-2 viral loads across the different stages of the COVID-19 pandemic. MedRxiv **2020**; https://www.medrxiv.org/content/10.1101/2020.07.10.20149773v1.

20. Singanayagam A, Patel M, Charlett A, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Euro Surveill **2020**; 25(32).

21. Jefferson T, Spencer E, Brassey J, Heneghan C. Viral cultures for COVID-19 infectivity assessment. Systematic review MedRxiv **2020**; https://www.medrxiv.org/content/10.1101/2020.08.04.20167932v4.

22. Tomassini S, Kotecha D, Bird PW, Folwell A, Biju S, Tang JW. Setting the criteria for SARS-CoV-2 reinfection - six possible cases. J Infect **2020**.

23. To KK, Hung IF, Ip JD, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis **2020**.

24. Goldman JD, Wang K, Roltgen K, et al. Reinfection with SARS-CoV-2 and Failure of Humoral Immunity: a case report. MedRxiv **2020**; https://www.medrxiv.org/content/10.1101/2020.09.22.20192443v1.full.pdf+html.

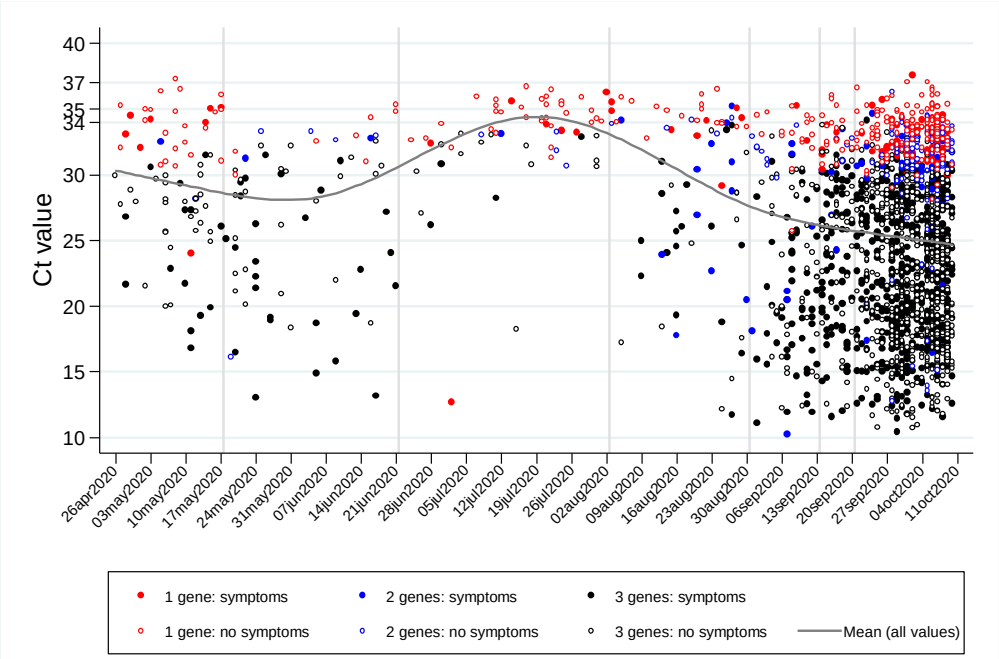
25. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. Nature **2020**; 584(7821): 420-4.

26. Hay JA, Kennedy-Shaffer L, Kanjilal S, Lipsitch M, Mina MJ. Estimating epidemiologic dynamics from single crosssectional viral load distributions. MedRxiv **2020**; https://doi.org/10.1101/2020.10.08.20204222.

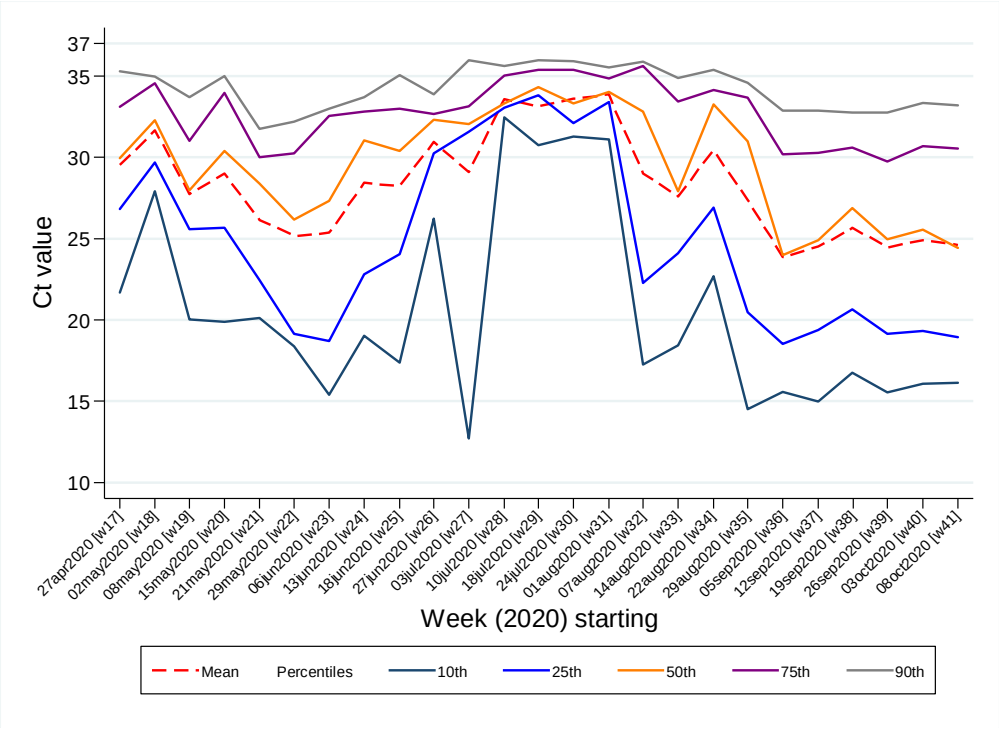
medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1 Variation over calendar time in Ct values (raw values (A) and distribution (B)), reported symptoms (C), and supporting evidence (D) for positive RT-PCR tests from community surveillance

A

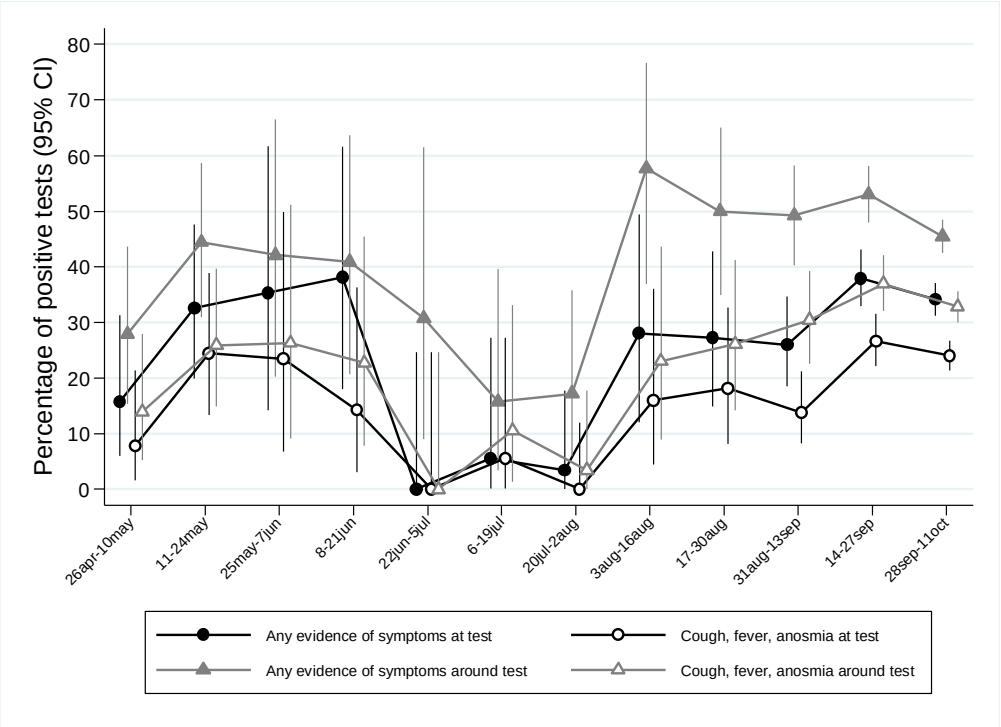


B

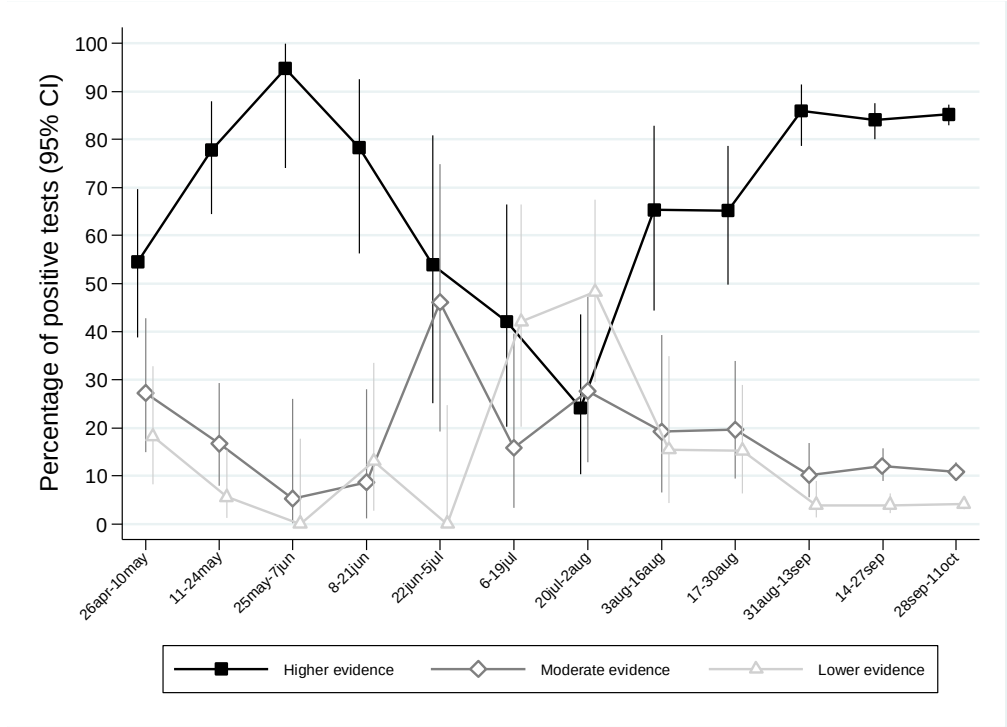


medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

C



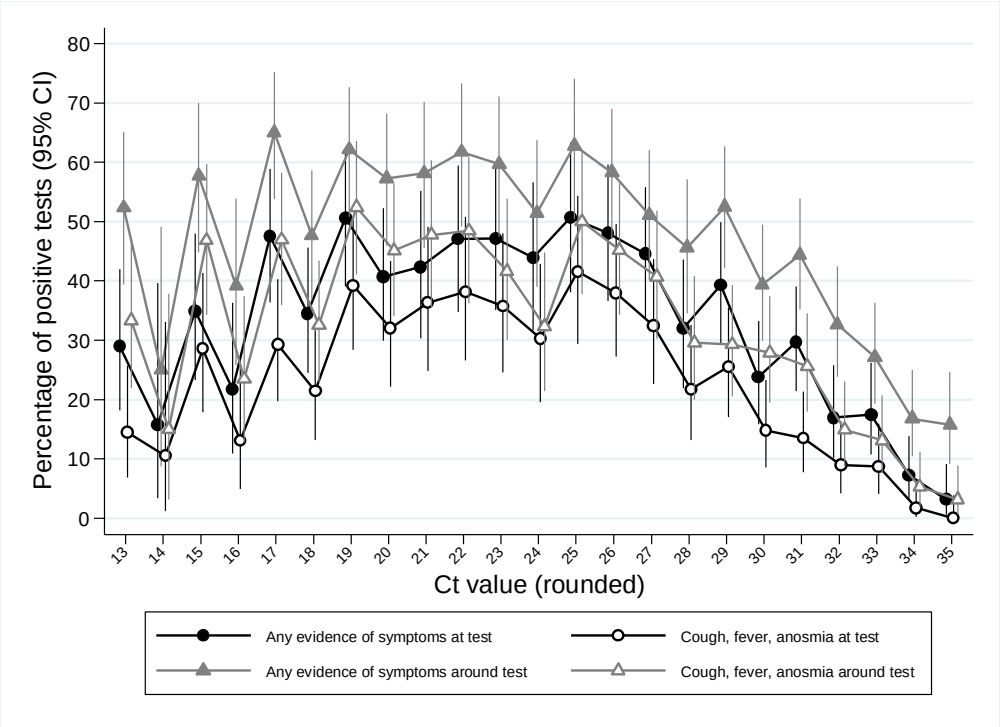
D



Note: in (A), points are classified by symptoms around the test. 16 individuals with missing data (in April (1), June (1), September (8), October (6)) are plotted as no evidence of symptoms. Line shows effect of calendar time from a five knot natural cubic spline, not adjusted for other factors (see Supplementary Methods).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2 Percentage reporting symptoms by Ct value

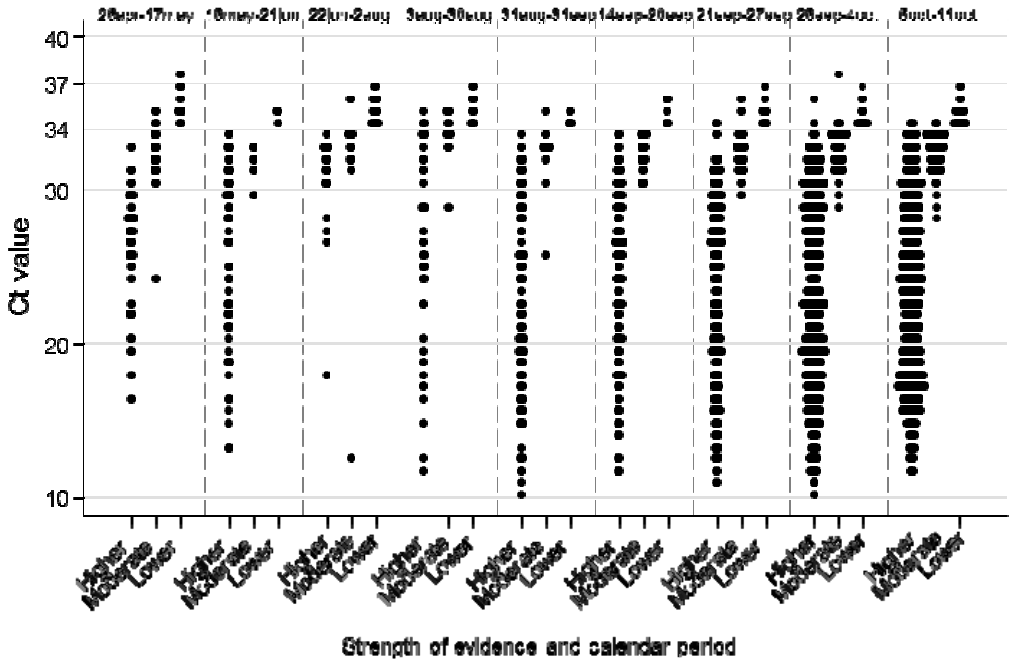


Note: Ct values under 13 and over 35 grouped with 13 and 35 respectively.

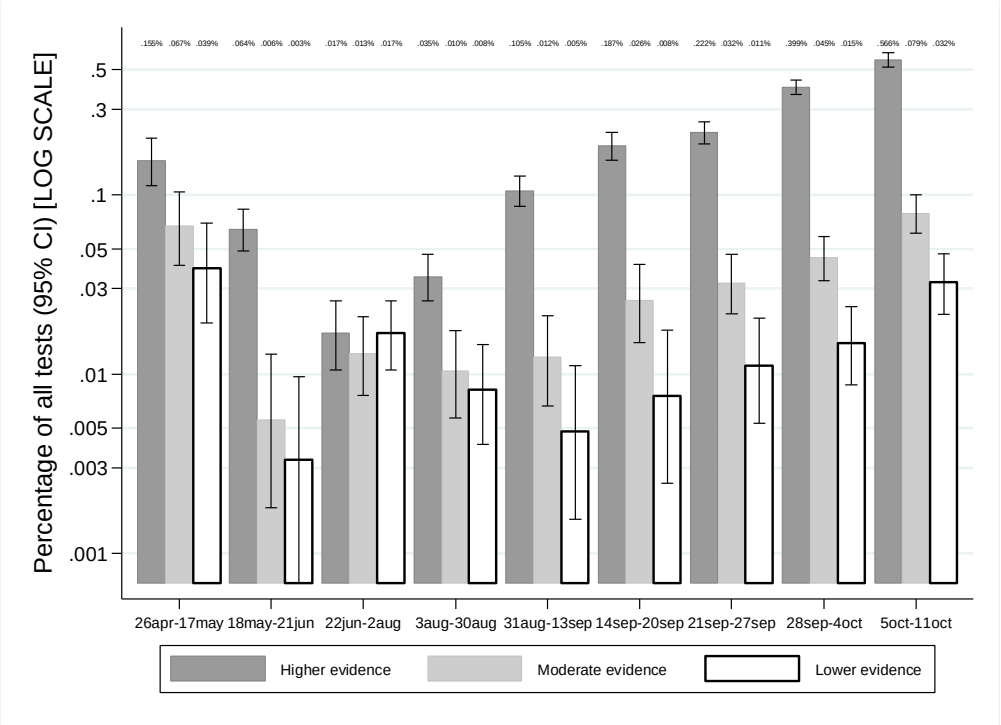
medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3 CT values (A) and percentage positive of all tests (B) by level of evidence and time

A



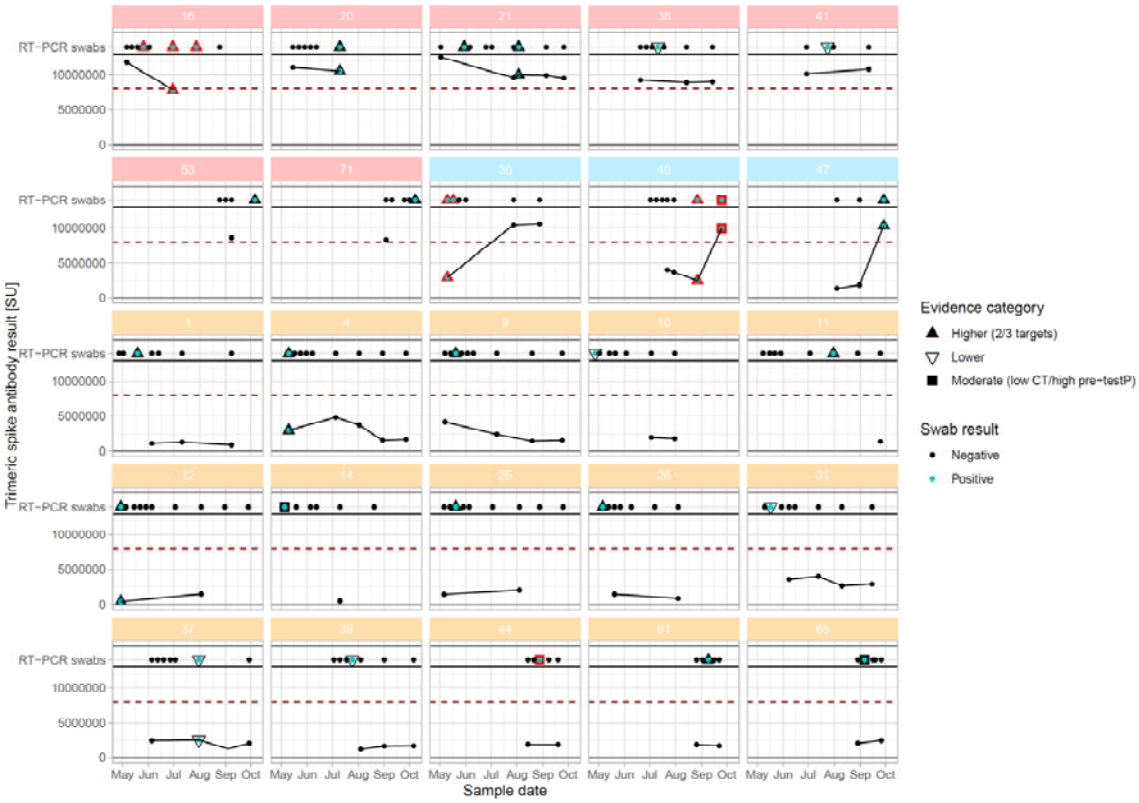
B



Note: calendar time categorised to reflect changes shown in Figure 1A and give roughly similar numbers of positives per category where these are longer than one week.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 4 Longitudinal swabbing and antibody testing trajectories for participants with positive RT-PCR despite antecedent anti-spike IgG antibody presence, who seroconverted post-swab-positive, and who did not seroconvert within three months



Note: Each facet represents a single individual; nasopharyngeal swab RT-PCR test results are shown at the top of each facet and anti-spike IgG antibody results (Oxford immunoassay) in the main body of the facet. The presence of symptoms at the time of a swab is denoted with a red background at the positive swab timepoint; asymptomatic cases are denoted with a black background at the positive swab timepoint. The red dashed line represents the antibody assay threshold for positivity (8 million SU). The facet header colours reflect three distinct categories of individuals: (i) individuals who were swab-positive despite antecedent anti-spike IgG antibody presence (rose-pink); (ii) individuals who seroconverted in response to being swab-positive (light blue); and (iii) individuals who did not have a positive anti-spike IgG antibody within three months of a positive swab (light orange). The remaining cases, where these trajectories were unclear, are represented in **Supplementary Figure 4**.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1 Genes detected in positive swabs

	All positives (N=1892)		First positive per participant (N=1516)	
Number of genes detected	n (%)	Median CT* (IQR) [range]	n (%)	Median CT* (IQR) [range]
1	345 (18%)	33.6 (32.3-34.6) [12.7-37.6]	307 (20%)	33.7 (32.5-34.7) [12.7-37.3]
2	185 (10%)	31.5 (29.8-32.8) [10.3-36.3]	138 (9%)	31.5 (29.6-33.0) [10.3-36.3]
3	1362 (72%)	22.8 (18.2-27.4) [10.5-34.2]	1071 (71%)	21.8 (17.7-27.0) [10.5-33.8]
Genes detected				
N only	243 (13%)	33.7 (32.5-34.7) [29.0-37.6]	213 (14%)	33.8 (32.6-34.7) [29.0-37.1]
ORF1ab only	83 (4%)	32.7 (31.9-33.8) [24.0-35.7]	75 (5%)	33.0 (31.9-33.9) [24.0-35.7]
S only**	19 (1%)	35.0 (34.3-36.1) [12.7-37.3]	19 (1%)	35.0 (34.3-36.1) [12.7-37.3]
N+ORF1ab	158 (8%)	31.3 (29.8-32.6) [10.3-36.3]	113 (7%)	31.2 (29.6-32.8) [10.3-36.3]
S+ORF1ab	9 (0.5%)	28.9 (26.1-31.0) [16.2-34.7]	8 (0.5%)	28.8 (24.5-32.1) [16.2-34.7]
N+S	18 (1%)	32.8 (32.3-33.1) [28.2-35.2]	17 (1%)	32.8 (32.3-33.1) [28.2-35.2]
N+S+ORF1ab	1362 (72%)	22.8 (18.2-27.4) [10.5-34.2]	1071 (71%)	21.8 (17.7-27.0) [10.5-33.8]

* taking the mean CT per positive swab across positive gene targets (Spearman rho=0.99 for each pair of genes, p<0.0001)

** through mid-May only: after this samples positive for the S gene only were not called positive overall.

Note: comparing first vs subsequent positives per participant, exact p<0.0001 for both number of genes detected and specific genes detected.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.25.20219048; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2 Evidence supporting positive test results and impact on other factors

	Strength of evidence for true infection			
	Higher	**Moderate**	**Lower**	**p (exact)**
Number (col %) (N=1892)	1547 (82%)	233 (12%)	112 (6%)	
Factors determining classification				
Number of genes detected (row %)	3: 1362 (88%) 2: 185 (12%)	1: 233 (100%)	1: 112 (100%)	
CT, median	24.0	32.8	34.9	
CT, n (row %) <34*	1543 (99.7%)	212 (91%)	0 (0%)	
Symptoms around test, n (row %)	790/1532 (52%)	76/232 (33%)	0/112 (0%)	<0.0001 (exc lower)
Occupational risk**, n (row %)	67/1356 (5%)	10/201 (5%)	0 (0%)	0.98 (exc lower)
Other factors				
Cough, fever, anosmia around test, n (row %)	560 (36%)	30 (13%)	0 (0%)	<0.0001
First positive test n (row %) (vs subsequent positive test)	1209 (78%)	198 (85%)	109 (97%)	<0.0001
First test in study, n (row %) (vs follow-up ie prior negative in study)	500 (32%)	85 (36%)	40 (36%)	0.37
Whole genome sequence obtained†	95/137 (69%)	5/32 (16%)	0/15 (0%)	<0.0001
Any other household member ever positive‡	691/1376 (50%)	46/196 (23%)	5/96 (5%)	<0.0001

* approximate 97.5th percentile of CT in higher evidence positives through 2 August when classification first applied.

** reported working in a patient-facing healthcare role/care/residential home.

† whole genome sequenced successfully out of attempted (other positives not found or not yet attempted)

‡ denominator households with 2 or more study participants.

Note: classification arbitrarily determined on 2 August based on the number of genes detected, CT values and pre-test probability (see Methods).