

Identification of undetected SARS-CoV-2 infections by clustering of Nucleocapsid antibody trajectories

Leslie R. Zwerwer^{*1,2,3}, Tim E. A. Peto^{#1,4,5,6}, Koen B. Pouwels^{#5,7}, Ann Sarah Walker^{#1,5,6}, and the COVID-19 Infection Survey team

contribution considered equal

1. Nuffield Department of Medicine, University of Oxford, Oxford, UK
2. Department of Health Sciences, University of Groningen, University Medical Center Groningen, The Netherlands
3. Center for Information Technology, University of Groningen, Groningen, The Netherlands.
4. Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK
5. The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford, Oxford, UK
6. The National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford, Oxford, UK
7. Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK

*Corresponding author: Leslie R. Zwerwer, Department of Health Sciences, University of Groningen, University Medical Center Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands. Email: l.r.zwerwer@rug.nl.

Date of submission: October 17, 2024

Keywords: COVID-19, SARS-CoV-2, nucleocapsid antibodies, PCR, undetected infections

Abstract

During the COVID-19 pandemic, numerous SARS-CoV-2 infections remained undetected. Serological testing could potentially aid their identification. We combined results from routine monthly nose and throat swabs, and self-reported positive swab tests, from a UK household survey, linked to national swab testing programme data from England and Wales, together with Nucleocapsid (N-) antibody trajectories clustered using a longitudinal variation of K-means to estimate the number of infections undetected by either approach (N=185,646). After combining N-antibody (hypothetical) infections with swab-positivity, we estimated that 7.4% of all true infections would have remained undetected, 25.8% by swab-positivity-only and 28.6% by trajectory-based N-antibody classifications only. Congruence with swab-positivity was much poorer using a fixed threshold to define N-antibody infections. Additionally, using multivariable logistic regression N-antibody seroconversion was more likely as age increased between 30 and 60 years, in non-white participants, those less (recently/frequently) vaccinated, for lower Ct values in the range above 30, in symptomatic and Delta (vs BA.1) infections. Comparing swab-positivity data sources showed that routine monthly swabs were not sufficient to detect infections by swab-positivity only and incorporating national testing programme/self-reported data substantially increased detection rates. Overall, whilst N-antibody serosurveillance can identify infections undetected by swab-positivity, optimal use requires trajectory-based analysis.

Introduction

To July 21, 2024, almost 776 million severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections have been reported worldwide¹. Nevertheless, many infections remained undetected and therefore the actual number is thought to be substantially higher^{2,3}. Serological testing can potentially provide information on undetected infections, thereby improving estimates of the number of previous infections^{4,5,6}.

Several studies have explored serological testing for SARS-CoV-2 infections by analysing either spike (S-) or nucleocapsid (N-) antibodies^{6,7,8,9}. Levels of both are in most people, at least temporarily, raised after SARS-CoV-2 infection. Because the most widely used SARS-CoV-2 vaccines target the spike protein, leading to increased S-antibody levels following vaccination, S-antibodies cannot easily be used to estimate how many people have been previously infected in populations with high vaccination rates, such as high-income countries^{10,11}. N-antibodies do not directly respond to the most commonly used mRNA and adenovirus SARS-CoV-2 vaccinations^{10,11}. Nevertheless, the sensitivity of N-antibodies to detect infections depends strongly on the population and thresholds used; previous studies have reported sensitivities ranging from 40-100%^{4,12,13,14,15,16}.

Various demographical characteristics have also been shown to affect N-antibody seroconversion following infection. For instance, several studies reported higher antibody titers, and hence higher seroconversion rates, in males^{17,18} and older individuals^{18,19,20}. Other factors influencing seroconversion include presence of symptoms/disease severity^{16,17,19,20}, hospitalisation^{17,18}, ethnicity²⁰ and body mass index^{18,19}. Moreover, while N-antibodies do not directly respond to most commonly used SARS-CoV-2 vaccinations, some studies have suggested N-antibody seroconversion might be reduced in vaccinated individuals^{9,13,16}. For instance, in a randomized controlled trial examining mRNA-1273 vaccine effectiveness, only

40% (95% confidence interval (CI): 27-54%; n=21) of 52 vaccination recipients showed N-antibody seroconversion after polymerase chain reaction (PCR) confirmed symptomatic infection with SARS-CoV-2 versus 93% (95%CI: 92-95%; n=605) of 648 placebo recipients¹³.

Studies aimed at generating learning from the pandemic rely on accurate estimates of infection, often inferred from PCR- and lateral flow test (LFT)-based surveillance. To assess the effectiveness of these systems, it is essential to quantify the number of infections they miss that could be identified from serology, and limitations of such serosurveillance (e.g. lower response rates among specific subgroups and with asymptomatic infections (which nevertheless can transmit onwards), impact of positivity thresholds). To our knowledge, there are no studies to date estimating the effectiveness of combining N-antibody seropositivity and PCR/LFT. Here, we therefore examine the ability of N-antibodies to identify prior (undetected by swab-positivity) SARS-CoV-2 infections in a general community-based cohort including vaccinated individuals, using clustering of longitudinal N-antibody trajectories. Additionally, we explore reasons for lack of seroconversion after PCR-confirmed SARS-CoV-2 infection, and the impact of defining infections based on different data sources.

Results

Population

Between February 28, 2021 and January 30, 2022, the period when N-antibodies were assayed within the COVID-19 Infection Survey (see **Methods**), 270,686 participants provided blood samples for serological testing (**Supplementary Fig. 1**), median 6 per participant. The median age at first N-antibody measurement was 55 years; 54.2% participants were female, 94.0% reported white ethnicity, 26.2% a long-term health condition and 5.0% reported working in healthcare (**Supplementary Table 1**). Respectively, 7.3%,

28.4%, 58.5% and 0.1% of participants had received 1, 2, 3 and 4 vaccinations by the end of the period in which they had N-antibodies measured (denoted their study period), with 5.7% participants remaining unvaccinated throughout. We defined swab-positive infections using positive and negative PCR results from routine monthly nose and throat swabs taken for the COVID-19 Infection Survey, positive swab PCR or LFT results from the national testing programmes in England and Wales or self-reported positive swab tests (see **Methods**). We aggregated swab-positive infections into four different classes: *No positive swab* before or during the participant's study period (81.2%), swab-positive infection *before* the participant's study period only (8.5%), swab-positive infection *during* the participant's study period only (9.9%) and swab-positive infection *before and during* the participant's study period (0.5%).

Clustering of N-antibody trajectories

In order to classify different types of N-antibody trajectories, we used a longitudinal variation of K-means in participants with ≥ 4 N-antibody measurements, in order to ensure the N-antibody trajectories had sufficient information to detect SARS-CoV-2 infections. This excluded 85,040 participants (**Supplementary Fig. 2**), who were slightly younger (**Supplementary Table 2**), as well as being more likely to report fewer vaccinations, as expected since those leaving the survey before January 2022 would have both fewer vaccinations and fewer measurements. Since all N-antibody measurements were censored at the lower and upper limits of quantification (respectively, 10 ng/mL and 200 ng/mL), clustering was not performed for 85,449 participants with no evidence of a previous infection (all N-antibody levels ≤ 10 ng/mL) and 326 participants with evidence of a previous infection (all N-antibody measurements ≥ 200 ng/mL) who were simply assigned to these two respective additional clusters. We therefore applied the longitudinal variation on K-means to identify 13 clusters in the remaining 99,871 participants (**Supplementary Fig. 2**) using absolute values (denoted identity clustering, 'id') and using \log_2 values (denoted 'log₂').

After careful examination of these 13 clusters from the two N-antibody transformations (**Supplementary Fig. 3**), we grouped them into four types: *flat*, *decreasing*, *increasing*, and those that first *decreased and then increased*. Biologically, the different categories broadly correspond to having no evidence of an infection before or during the study period, evidence of a previous infection before the study period only, evidence of a current infection during the study period only and evidence of a previous and current infection, respectively (**Supplementary Fig. 4 and 5**). An overall classification was obtained based on consensus: where the two transformations differed (N=9,644, 9.7%), often relating to smaller absolute increases which were magnified on the relative (log) scale, participants were classified using visualization of the trajectories (**Supplementary Fig. 6 and 7**). Interestingly, the N-antibody trajectories for 54 participants in cluster 13 using identity clustering and cluster 10 using log₂ transformed clustering implied two different infections during the participant's study period. Overall 20 (37.0%) of these participants had two or more swab-positive infections during their study period (compared to 350 (0.2%) among those with ≥ 4 N-antibody measurements).

Figure 1 shows the N-antibody trajectories for the final different trajectory-based classifications and swab-positive infection groups. More specifically, it shows that *flat* N-antibody trajectory-based classifications with *no positive swab before or during* the study period had relatively little variation. *Flat* N-antibody trajectory-based classifications with a swab-positive infection *before* the participant's study period had a marginal decrease in N-antibody levels overall. Moreover, N-antibody trajectories classified as *flat* with a swab-positive infection *during* or *before and during* the participant's study period had a marginal increase in N-antibody levels overall. In contrast, N-antibody trajectories classified as *decreasing* with *no positive swab before or during* the study period or a swab-positive infection *before* their study period showed a marked decrease in N-antibody levels.

Decreasing N-antibody trajectory-based classifications with a swab-positive infection *during* or *before and during* the participant's study period had decreasing then increasing N-antibody levels. Finally, regardless of swab-positivity group, all trajectories classified as *increasing* or *decreasing and increasing* had considerable increases in N-antibody levels.

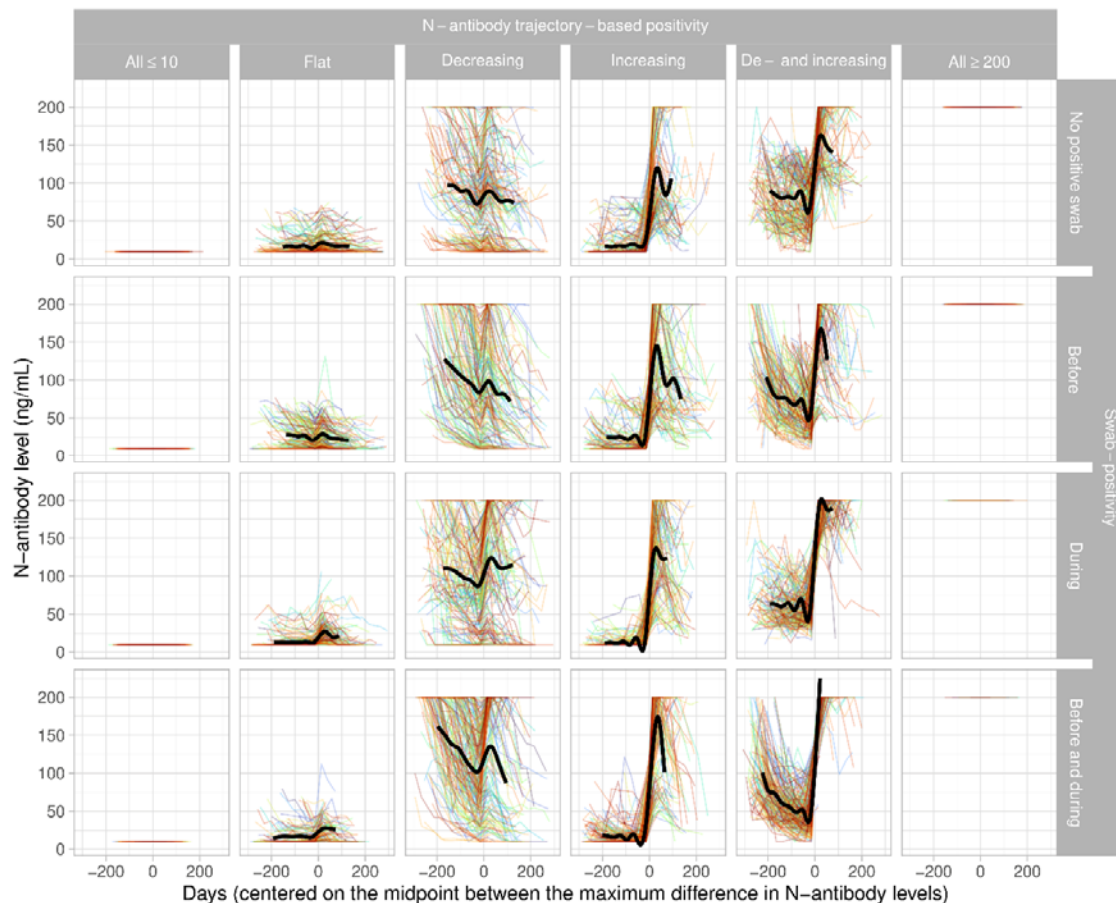


Fig 1. N-antibody trajectories for the different N-antibody and swab-positive infection groups (restricted to those with ≥ 4 N-antibody measurements, see **Supplementary Fig. 2**). For comparability, trajectories are centered on the midpoint between the maximum difference between any two consecutive measurements per participant. This approximates the hypothetical infection date for those with an N-antibody trajectory compatible with infection, but can create a small but arbitrary distortion in those without swab-positive infections and classified as *flat* or *decreasing*. Each frame contains a random sample of 200 N-antibody

trajectories (see **Fig. 2** for numbers and cell percentages). Black line depicts a generalised additive modelling smooth for the inner 90% of all observations in each cluster.

Figure 2 shows the number of participants in the different trajectory-based N-antibody classifications and swab-positive infection groups. Overall agreement between the N-antibody trajectory-based classification and swab-positive infections was 86.2% (95%CI: 86.0–86.3%) in all participants with ≥ 4 N-antibody measurements. 28.6% (28.1–29.2%) of the 25,404 swab-positive infections during the study period did not show any evidence of an infection from their N-antibody trajectories. Moreover, 25.8% (25.3–26.4%) of the 24,440 participants with *increasing/de- and increasing* N-antibody trajectories had no evidence of a swab-positive infection. For 18,128 (9.8%) participants, a swab-positive infection occurring during the participant’s study period was also detected by N-antibody trajectory-based analysis. For these participants, we estimated N-antibody (hypothetical) infection dates as 14 days before the midpoint between the two measurements with the maximum increase in N-antibody levels. Overall, most (61.5%) N-antibody (hypothetical) infection dates were within 15 days of the closest swab-positive date (**Supplementary Tables 3 and 4, Supplementary Fig. 8**), being ≥ 60 days in only 505 (2.8%) participants.



Fig 2. Number of participants classified in each N-antibody trajectory-based and swab-positive infection group. Of all 13,324 participants with a swab-positive infection *before/before and during* their study period, only 108 (0.8%) participants had two distinct swab-positive infections *before* their study period and of all 25,052 participants with a swab-positive infection *during/before and during* their study period, 350 (1.4%) participants had two or more swab-positive infections during their study period.

Note: showing raw counts, total percentages, row and column percentages.

Estimated number of true infections

Using both N-antibody trajectory-based classifications and swab-positive infections (including multiple swab-positive infections per participant), we identified 31,716 infections during the study period. 24,440 (77.1%; 95% CI 76.6–77.5%) of these detected infections were identified using N-antibody trajectory-based analysis, 25,404 (80.1%; 79.7–80.5%)

were detected by swab-positivity and 18,128 (57.2%; 56.6–57.7%) were detected by both swab-positivity and N-antibody trajectory-based analysis. Assuming that both types reflected true infections and there were no false-positives, using a method dependent capture-recapture model we estimated that in total there would have been 34,249 (34,115–34,383) infections during the study period among all participants with ≥ 4 N-antibody measurements. Of those infections 7.4% (7.0–7.8%) remained undetected with either method, 25.8% (25.5–26.1%) by swab-positivity and 28.6% (28.4–28.9%) by N-antibody trajectory-based classification.

In subgroup analyses estimating the percentage of true infections undetected by either methods results were slightly different (see **Supplementary Table 5**). Overall, when stratifying by vaccination status 4.8–10.9% of the true infections were undetected, with respectively 6.6% (95% CI 5.1–8.2%) and 10.9% (9.9–11.9%) of all infections remaining unidentified in unvaccinated participants and participants with 3 or 4 vaccinations. Moreover, respectively 59.7% (50.6–68.4%), 5.8% (5.4–6.2%) and 7.4% (6.8–8.1%) of all true infections were undetected by either method during the Alpha, Delta and BA.1 epoch.

Sensitivity analyses reclassifying the 505 participants with ≥ 60 days between the N-antibody (hypothetical) infection date and closest swab-positive infection date gave comparable results. Where the swab-positive infection date was ≥ 60 days before the N-antibody (hypothetical) infection date, we classified the infection as detected by swab-positivity only, and as N-antibody only when the swab-positive infection date was ≥ 60 days after the N-antibody (hypothetical) infection date. Under these assumptions, of all detected infections, 24,139 (76.1%; 95% CI 75.6–76.6%) were detected using N-antibody trajectory-based analysis, 25,200 (79.5%; 79.0–79.9%) using swab-positivity and 17,623 (55.6%; 55.0–56.1%) by both methods. Under the assumption that neither methods identifies any false positives, of a total 34,517 (34,374–34,663) estimated true infections during the study period,

8.1% (7.7 – 8.5%) would have been undetected by both swab-positivity and trajectory-based N-antibody positivity.

Next, we performed a sensitivity analysis using the manufacturer’s proposed N-antibody seropositivity threshold of 30 ng/mL²¹. Using both N-antibody threshold-based classifications and swab-positive infections we detected 39,511 (hypothetical) infections, of which 32,702 (82.8%; 95% CI 82.4–83.1%) were identified using this fixed 30 ng/mL threshold. Further, a much smaller percentage (25,404, 64.3%; 63.8–64.8%) of all detected (hypothetical) infections were swab-positive. 18,595 (47.1%; 46.6–47.6) infections were detected by both methods. Hence, under the assumption of no false positives as above, of a total of 44,676 (44,460–44,874) estimated true infections during the study period, 11.6% (11.1–12.0%) would have been missed by both swab-positivity and infections defined by the fixed N-antibody threshold (compared to 7.4% (7.0–7.8%) using swab-positivity and N-antibody trajectory-based classification).

Associations with lack of N-antibody response

Subsequently we compared participant characteristics between swab-positive infections with *increasing* or *decreasing and increasing* N-antibodies trajectories (i.e. responders) and *flat* or *decreasing* N-antibody trajectories (i.e. non-responders) (**Supplementary Fig. 9, Supplementary Table 6**). In a multivariable model, we found significantly lower odds of non-response (i.e. higher odds of seroconversion) as age increased between 30 and 60 years and in non-white participants (**Supplementary Table 7**). We also found that vaccination influenced N-antibody non-response, with significantly lower odds of non-response in unvaccinated participants, and those that were less recently vaccinated or had fewer vaccinations. Furthermore, higher cycle threshold (Ct) values in the range above 30 were associated with significantly greater odds of non-response. Additionally, participants

with symptoms were significantly less likely to be non-responders. Finally, compared to infections during the Delta epoch, infections during the BA.1 epoch were significantly more likely to be N-antibody non-responders.

Using different data sources to define infections and trajectory-based vs. fixed threshold-based positivity

Next, we compared positivity based on N-antibody trajectories and the fixed 30 ng/mL threshold using different data sources to define swab-positive infections. Across the different data sources, the percentage of participants with N-antibody (hypothetical) infections that were identified using swab-positivity ranged between 29.6-78.8% for the trajectory-based classification and between 22.4-60.5% for the fixed 30 ng/mL threshold (**Fig. 3a**). Using only results from the routinely scheduled swabs (i.e. from the COVID-19 Infection Survey alone), identification of infection was poor with between 70.4-77.6% of N-antibody responders remaining unidentified. The percentage identified through swab-positivity showed the highest increase moving from defining swab-positive infections using the survey only to the survey plus the national testing programme from England and Wales (covering 89% of participants).

Notably, there was a considerable difference in the percentage of swab-positive infections identified among N-antibody (hypothetical) infections comparing the trajectory-based classification and the threshold-based classification, with the threshold-based classification identifying consistently more participants as having N-antibody (hypothetical) infections during their study period (also **Supplementary Table 8**). **Supplementary Fig. 10** visualises the trajectories of N-antibody negative and N-antibody positive participants using the two classifications, stratified by swab-positivity (all during the participant's study period). It shows that most participants with N-antibody negative trajectory-based and N-antibody

positive threshold-based classification had no positive swab (87.4%) and N-antibody trajectories mostly had a marginal increase, to just above the 30 ng/ml threshold. Trajectories from participants with N-antibody positive trajectory-based and negative threshold-based classifications had large increases in N-antibody levels but 59.5% still had no positive swabs during their study period.

Finally, the percentage of all N-antibody negative participants with no positive swab decreased from 99.0% to 94.7% with increasing richness of data source for the trajectory-based N-antibody classification and from 99.0% to 94.8% for the threshold-based classification (**Fig. 3b**). Interestingly, including participants thinking they had had COVID-19 as a positivity criterion (without any swab-positive) made only marginal differences in the percentage of participants without (swab)-positive infections among N-antibody negatives.

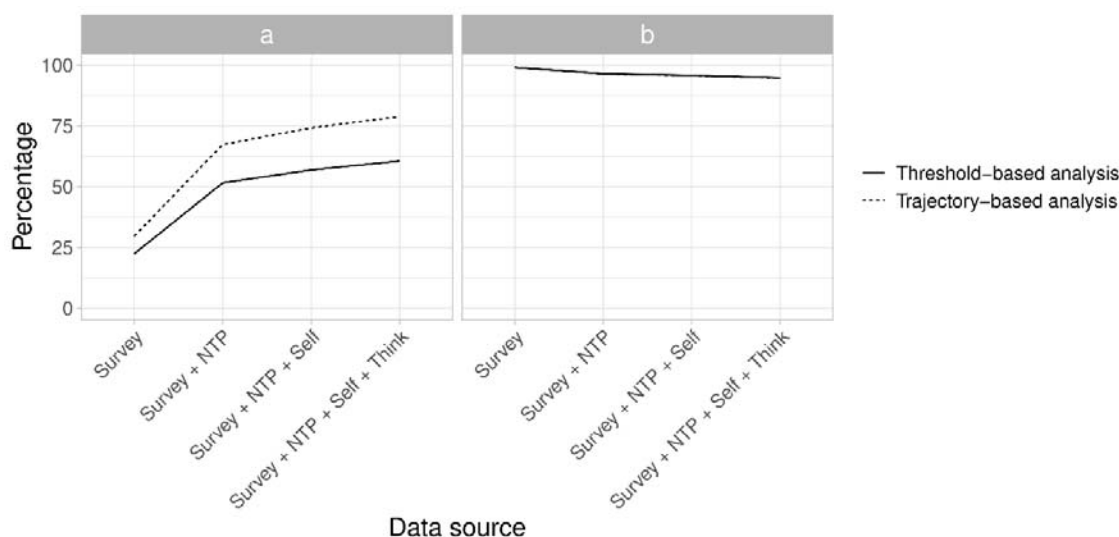


Fig. 3. Comparison of the N-antibody trajectory-based classification and fixed 30 ng/mL classification across different data sources used to define swab-positive infections. (a) Percentage of participants with N-antibody (hypothetical) infections that were identified using swab-positivity (b) Percentage of participants without N-antibody (hypothetical) infections with no positive swab. Survey: only using positive and negative swab PCR test

results from the COVID-19 Infection Survey to define swab-positive infections; NTP: using positive PCR and LFT swab test results from national testing programmes in England and Wales; Self: using self-reported positive swab test results; Think: reports on thinking one had had COVID-19.

Discussion

During the recent COVID-19 pandemic, hundreds of millions of individuals tested positive for a SARS-CoV-2 infection. However, due to a considerable number of asymptomatic individuals, the true number of infections remains unknown². In this study we used data from a large broadly representative UK household survey and examined the efficacy of detecting prior (undetected) SARS-CoV-2 infections in the general population by clustering of N-antibody trajectories. We found that under the assumption that swab-positives and N-antibody positives both reflect true infections, 7.4% of all true SARS-CoV-2 infections would have remained unidentified from both swab results and N-antibody trajectories (compared to 25.8% by swab-only and 28.6% by trajectory-based N-antibody classifications only).

As far as we are aware, no other study has examined the efficacy of combining swab-positivity and N-antibody serological testing to identify undetected SARS-CoV-2 infections. However, several studies have examined the ascertainment rate for swab-positivity alone in the UK. For instance, Colman et al. (2023) estimated that, after SARS-Cov-2 testing become widely available in the UK, 60-70% of all infections remained undetected by national healthcare and community testing programmes by calibrating reported cases to the swab-positivity rate from the COVID-19 infection survey, while accounting for the incubation period distribution, and the time-dependent test sensitivity of PCR and lateral flow tests²². Nightingale et al. (2022) estimated the under-ascertainment rate at 75.0% using swab-

positivity from the COVID-19 infection survey as a ground truth to estimate the performance of the national testing programmes²³. Here, we took a different approach and estimated the ‘true’ number of infections by applying a log-linear capture-recapture model to infections detected by positive swab tests - from the COVID-19 infection survey or national testing programmes, or self-reported positive swab test – or infections detected through the clustering analysis performed on N-antibody results. Focusing purely on the swabs included in this comparison, 25.8% of all infections were missed. Further, estimates of undetected infections vary by age group, variant, and region, which may be related to differences in symptoms/disease severity, public sentiment and availability of testing²².

Notably, even using these increasingly rich data sources on swab-positivity, still 25.8% of all true infections remained undetected by swab-positivity during the study period. This may be for a variety of reasons. Firstly, the timing of the swab is of critical importance and can result in false-negative results^{24, 25}. Monthly swabs in the survey were a trade-off between expected duration of PCR positivity (mean 21 days from infection in human challenge studies²⁶) and costs, aiming to identify >80% of infections. Secondly, viral load can be unequally distributed throughout the body³, with higher viral RNA concentrations in stool and sputum²⁷, and related to severity of symptoms²⁸; low viral load could cause false-negative swab results²⁵. Other reasons for false-negative swabs include viral genetic variation and challenges with self-sampling²⁵.

Consistent with a much smaller study among hospitalised individuals, we found that a little over a quarter of all swab-positive infections did not seroconvert in terms of N-antibodies⁴. However, several other studies also report lower percentages of non-responders among swab-positives^{12, 14, 16, 18, 19, 29, 30}. Nevertheless, these studies did not use N-antibody trajectories to define seroconversion, which we showed made important gains in identification, and predominately focused on specific subgroups, such as healthcare workers

and/or had small sample size. Consistent with most other studies, we found no association between seroconversion rates and gender^{9, 16, 19, 20}. We found N-antibody seroconversion rates increased as age increased between 30 and 60 years, consistent with higher antibody titers (and thus higher seroconversion rates) in older individuals in some studies^{18, 19, 20}, although one study found higher seroconversion rates among younger age groups compared to individuals ≥ 65 years (adjusted for vaccination)¹⁶. Again consistent with the literature, we found seroconversion was more likely among individuals who reported non-white ethnicity²⁰, were less (recently or frequently) vaccinated^{9, 13, 16}, infections with lower Ct values in the range above 30²⁰ (a proxy for viral load³¹) and symptomatic infections^{19, 20}. However, in contrast to one previous study, we also found that participants with an infection during the BA.1 epoch were significantly less likely to seroconvert compared to participants with an infection during the Delta epoch¹⁶. This could potentially relate to the low number of asymptomatic infections in this previous study¹⁶, since the proportion of asymptomatic infections is significantly higher for Omicron compared to Delta infections³² and we, and others, have shown that individuals with asymptomatic infections are less likely to seroconvert^{19, 20}.

Where participants were identified as having been infected using both approaches, estimated N-antibody (hypothetical) infection dates were mostly within 15 days of the closest swab-positive date. Nonetheless, the percentage of N-antibody (hypothetical) infections identified using swab-positivity was highly dependent on the data source. We incrementally tested adding the different data sources into swab-positivity definitions, reflecting their likely level of ascertainment. Using the survey swab-positivity alone, only approximately a quarter of all N-antibody trajectory-based infections were identified. The use of data from national testing programmes vastly increased infection identification rates, although on their own, they provide a poor level of ascertainment (as above^{22, 23}) and incorporation of unbiased swab

positivity testing data from the COVID-19 infection survey has been demonstrated to be essential to reconstruct the epidemic³³. Using ‘thinking one had COVID-19’ as a positivity criterion only modestly increased the number of N-antibody infections identified, whilst having a marginal impact on the percentage of false-negatives, which is remarkable considering that an earlier study showed that in the UK only 51.5% of all individuals recognises common COVID-19 symptoms³⁴. Compared to threshold-based N-antibody positivity classifications (based on the manufacturer’s threshold), trajectory-based classification was consistently more aligned with swab-positivity. The threshold-based classification identified considerably more (hypothetical) infections whose trajectories were relatively flat but elevated and never tested positive by swab. These relatively flat but elevated antibody trajectories could potentially reflect cross-reactivity^{10, 35}.

The main study strength is our use of a longitudinal variation of K-means to identify infections from N-antibody trajectories, rather than using an arbitrary fixed threshold. By comparing how antibody levels respond over time, this allowed us to still classify participants with “blunted” responses as having been infected. However, our study has several limitations. Firstly, N-antibody measurements were obtained using one assay only, which was ultimately not commercialised. Secondly, we only applied one clustering method and due to computational limitations were not able to optimize the number of clusters. However, in contrast to most other studies that aim to cluster a high-dimensional space, we clustered time-series, which allowed for visualisation and thorough inspection of clusters without projection methods that depend on hyperparameters and interpretation such as Uniform Manifold Approximation and Projection³⁶. Moreover, swab-positivity allowed careful triangulation of each cluster, overall leading to a biologically plausible classification for most participants (**Fig.1&2**). Next, by necessity all measurements below or above the lower and upper limits of quantification were censored, potentially leading to incorrect N-antibody trajectory-based

classifications. For instance, fully censored participants could have had a considerable increase or decrease in N-antibody measurements, which was no longer visible due to the censoring. Furthermore, the number of participants with a SARS-CoV-2 infection before their study period is most likely an underestimation of the true number of infections, given lack of widespread testing in the first wave in March-May 2020, and recruitment of most survey participants from July-October 2020. Also, detection of previous infections using N-antibodies depends on the durability of seropositivity, with S-antibody response (before vaccination) in general more persistent than N-antibody response²⁹. Estimating the percentage of infections that remained undetected by swab-positivity and N-antibodies depended on PCR tests not being subject to false-positives and N-antibody trajectories not being subject to cross-reactivity. Previous analyses using the COVID-19 Infection Survey have shown that specificity of the PCR testing protocol was really high, alleviating concerns about potential false-positives resulting from PCR testing³⁷. Specificity has also been suggested to be very high for the N-antibody tests³⁸. Moreover, we used a method dependent capture-recapture model, in which the probabilities of detecting true infections varied by method, but not per individual and infection episode, which over simplifies reality as seen in the sensitivity analysis (**Supplementary Table 5**). Data from national testing programmes in Northern Ireland and Scotland were not available (11% of survey population); to mitigate this we also included self-reported positive swab results which had very high agreement with national testing data in England and Wales (>95%). Finally, we had no information on symptom severity, which could also be related to N-antibody seropositivity¹⁹.

In conclusion, we used N-antibody trajectories from a large broadly representative UK household survey to examine the total number of undetected SARS-CoV-2 infections. Whilst N-antibodies serosurveillance can be used to improve estimates of the number of

previous infections, for optimal use, trajectory-based analysis is required over threshold-based analysis.

Methods

Data collection

Data came from the UK's Office for National Statistics (ONS) COVID-19 Infection Survey (ISRCTN21086382, protocol on <https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets>), a large longitudinal survey inviting all individuals aged 2 years or older living within randomly selected private households across the UK to participate. Following verbal consent, study workers visited each household, and recruited all consenting residents aged 2 years or older who provided written informed consent (from parents/carers for those under 16 years; those aged 10-15 years also provided written assent). Participants could also provide optional consent for subsequent weekly visits in the first month and then monthly, up to the latest of March 2023, when they became no longer resident at the selected address or no longer wished to participate (98% consented to post-enrolment visits). Ethical approval was obtained from the South Central Berkshire B Research Ethics Committee (20/SC/0195).

Data was collected on participants socio-demographic characteristics; at each assessment, data was collected on behaviours and vaccination status, and participants provided a nose and throat swab for PCR testing (self-taken; parents/carers took swabs for those under 12 years) (details in **Supplementary File 1**). Initially, those aged ≥ 16 years from a random 10-20% households were asked for optional consent to give monthly venous blood samples for serological testing; this was expanded to a larger randomly selected subgroup of households from April 2021 using capillary blood sampling to examine vaccine responses

(prioritising those with longer survey participation). Moreover, any participant ≥ 16 years testing PCR-positive through December 2021 was invited to provide blood samples on their subsequent monthly follow up visits.

Serological testing and definition of infections

Levels of SARS-CoV-2 S-antibody (throughout) and N-antibody (between February 28, 2021 and January 30, 2022 to monitor initial responses to the vaccination programme) were tested on venous or capillary blood samples using an enzyme-linked immunosorbent assay (ELISA) detecting anti-trimeric spike and nucleocapsid IgG developed by the University of Oxford. Before 26 February 2021, the S-antibody assay used fluorescence detection, with a positivity threshold of 8 million units validated on banks of known SARS-CoV-2-positive and -negative samples³⁹. After this, the S-antibody used a commercialized CE-marked version of the assay, the Thermo Fisher OmniPATH 384 Combi SARS-CoV-2 IgG ELISA (Thermo Fisher Scientific), with the same antigen and colorimetric detection, reporting normalized results in ng/mL of mAb45 monoclonal antibody equivalents (details in ⁷) and using $42 \square \text{ ng/mL}$ as the threshold for an IgG-positive or -negative result (corresponding to the 8 million units with fluorescence detection). SARS-CoV-2 N-antibody levels were tested using a research-use only assay (details in²¹). Lower and upper limits of quantification were 10 and 200 ng/mL respectively.

The study period was defined as the period in which participants had N-antibody measurements available. All survey data after the participant's study period was excluded from this analysis. We defined 'infection episodes' using results from swab test results as in⁴⁰. In brief we used all positive and negative PCR test results from the survey, linked information about positive only PCR and LFT from the national testing programmes in England and Wales (not available for Scotland and Northern Ireland), self-reported positive

swab tests from all participants (as national testing data was not available in Scotland/Northern Ireland; very high (>95%) agreement for participants in England and Wales). To reflect the fact that some individuals can test positive on PCR for extended periods of time when testing is independent of symptoms/case contacts as in the survey (in contrast to national testing programmes), whereas others have reinfections (confirmed by sequencing) after only short periods of time, we incorporated information from genetic sequencing, S-gene presence/absence, and Ct values, together with negative PCR test results from the survey only⁴⁰.

Classifying N-antibody trajectories

We clustered similar N-antibody trajectories in participants with ≥ 4 measurements together using a longitudinal variation of K-means with a dynamic time-warping loss function to account for varying periods of availability of N-antibody measurements, and gaps in each participant's trajectory due to missed visits or failed assays (details in **Supplementary file 1**)^{41, 42, 43}. Characteristics of those with < 4 vs ≥ 4 N-antibody measurements were compared using standardised differences. Participants with all N-antibody measurements either ≤ 10 or ≥ 200 were not formally clustered but assigned to two additional clusters. Due to the large sample size, optimisation of the number of clusters was not computationally feasible. Therefore, we chose to fit the largest number of clusters which was still computationally feasible to converge within 2 days ($n=13$, taking 40 hours on 10 cores).

To reflect the fact that both absolute and relative changes in N-antibody levels might indicate infection, we clustered N-antibody trajectories firstly using absolute values (denoted identity clustering, 'id') and secondly, using \log_2 values (denoted 'log₂'). Five different initialisations were used for each, with a maximum of 50 iterations, returning the clustering

solution with the lowest sum of squared dynamic time-warping distances between each trajectory and the corresponding cluster centroid (i.e. minimal inertia)⁴³.

We then visualised the N-antibody trajectories in each cluster together with a generalised additive model smooth (function 'geom_smooth(method='gam')' from ggplot2⁴⁴), and arbitrarily classified them based on expected trajectories following infection (**Fig.1**). We then took the consensus of the id and log₂ N-antibody classifications, with manual reconciliation where these disagreed (see Results), and compared the combined final classification with swab-positive infections as defined above (**Supplementary Fig. 2**).

Estimating infection dates

For participants with an N-antibody trajectory compatible with infection, we estimated the (hypothetical) infection date (the first date a participant would have tested positive on a nose and throat swab) assuming that the infection occurred 14 days before the midpoint between the two measurements with the maximum increase in N-antibody levels, given it takes on average ten days for N-antibodies to rise after developing symptoms⁸, the incubation period is approximately 6.5 days⁴⁵ and on average it takes 2.5 days from infection to swab-positivity⁴⁶. We then compared this (hypothetical) infection date estimated from N-antibody measurements with actual swab-positive infection dates (as defined above) for all participants with infections identified using both methods. Where participants had multiple swab-positive infections, we compared the closest swab-positive infection date to the N-antibody (hypothetical) infection date.

Estimating the total number of infections

To estimate the number of true infections in those participants with ≥ 4 N-antibody measurements, we used a capture-recapture model⁴⁷. This technique fits a loglinear model to the number of infections identified by swabs, N-antibody trajectories and their intersection to

estimate the number of infections missed by either methods. To reflect the fact that the number of true infections was equal for both methods, we used a closed population model. We accounted for heterogeneity in the infection detection probabilities of swabs and N-antibody trajectories by fitting a method dependent capture-recapture model, which allows the probabilities of detection to vary for swabs and N-antibody trajectories. To prevent overfitting, we chose not to model heterogeneity between infection episodes, meaning that all infection episodes had the same probability of being detected within each method. For participants with multiple swab-positive infections, we considered the closest swab-positive infection to the N-antibody (hypothetical) infection date detected by both methods and all other swab-positive infections detected by swab-positivity only. Moreover, we assumed that both swab-positives and N-antibody (hypothetical) infections reflected true infections (i.e. no false-positives).

We performed three sensitivity analyses. Firstly, we performed a subgroup analysis in which we calculated the number of true infections for different vaccination statuses and epochs. Both were determined at time of the infection, which was at the swab-positive date when available and otherwise at the N-antibody (hypothetical) infection date. Dependent on the time of infection the SARS-CoV-2 epoch was defined as Alpha when it was between December 7, 2020–May 16, 2021, Delta between May 17, 2021–December 12, 2021 and BA.1 between December 13, 2021–February 20, 2022, which was the first Monday where S-positivity for the corresponding variant was above 50% in the full survey population. Secondly, we reclassified all participants with ≥ 60 days between the estimated infection dates from the two methods. Where the swab-positive date was >60 days before the N-antibody (hypothetical) infection date, we classified the infection as swab-positivity only, and as N-antibody only when the swab-positive infection date was ≥ 60 days after the N-antibody

(hypothetical) infection date. Lastly, we classified N-antibody trajectories using the manufacturer's proposed N-antibody seropositivity threshold of 30 ng/mL²¹.

Associations with participant characteristics

We investigated lack of N-antibody seroconversion amongst participants with swab-positive infections and ≥ 4 N-antibody measurements (in whom seroconversion could be assessed as above) using logistic regression including all demographics and information related to the infection as covariates, that is age, sex, ethnicity, healthcare worker, long-term health condition, vaccination status at time of the swab-positive infection, Ct values, symptoms and the SARS-CoV-2 epoch (complete case analysis; details in **Supplementary Fig. 9**; results for other covariates similar excluding Ct values (most missing data) from the model). Participants with a N-antibody (hypothetical) infection and ≥ 60 days between the two infection dates were excluded from this analysis, as were a small number of participants with an earlier infection identified only by S-antibody seropositivity, as this could possibly be a marker of (previous) unregistered vaccination. We additionally excluded a very small number of infections before May 17, 2021 (emergence of Delta) (N=104/17,419 (0.6%)).

For participants with *increasing/de- and increasing* N-antibody trajectories who had multiple swab-positive infections, we considered the closest swab-positive infection to the N-antibody (hypothetical) infection date a responder and all other swab-positive infections non-responders. Vaccination was considered at the swab-positive infection date. Since there was limited variability in the time since vaccination for participants with 1, 3 and 4 vaccinations at the swab-positive infection (i.e. <250 participants were vaccinated >3 months ago), we aggregated time since vaccination and number of vaccinations into 7 different vaccination categories: not vaccinated, 1 vaccination, 2 vaccinations ≤ 3 months ago, 2 vaccinations 3–6 months ago, 2 vaccinations >6 months ago and 3 or 4 vaccinations, ignoring vaccinations ≤ 14

days before the swab-positive infection date. Dependent on the swab-positive infection date the SARS-CoV-2 epoch was again defined as Alpha when it was between December 7, 2020–May 16, 2021, Delta between May 17, 2021–December 12, 2021 and BA.1 between December 13, 2021–February 20, 2022. We initially fitted models with smooths for continuous covariates (age, Ct values); for interpretability, final models used piecewise linear effects with knots chosen based on visualisations of these smooths.

Other definitions of infections

Finally, we compared N-antibody (hypothetical) infections with infections defined using different data sources, specifically (i) only positive (and negative) swab PCR test results from the survey, (ii) positive and negative PCR results from the survey and positive swab PCR or LFT results from national testing programmes in England or Wales, (iii) positive and negative PCR results from the survey, positive swab PCR or LFT results from national testing programmes in England or Wales, and self-reported swab-positives and (iv) positive and negative PCR results from the survey, positive swab PCR or LFT results from national testing programmes in England or Wales, self-reported swab-positives and self-reports that participants thought they had had COVID-19. For each, we estimated the percentage of swab-positive infections among those with N-antibody (hypothetical) infections and the percentage of participants without swab-positive infections among those without N-antibody (hypothetical) infections. Sensitivity analysis used classifications based on the manufacturer's threshold.

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. Individuals can apply to be an accredited researcher using the

short form on https://researchaccreditation.service.ons.gov.uk/ons/ONS_registration.ofml. Accreditation requires completion of a short free course on accessing the SRS. To request access to data in the SRS, researchers must submit a research project application for accreditation in the Research Accreditation Service (RAS). Research project applications are considered by the project team and the Research Accreditation Panel (RAP) established by the UK Statistics Authority at regular meetings. Project application example guidance and an exemplar of a research project application are available. A complete record of accredited researchers and their projects is published on the UK Statistics Authority website to ensure transparency of access to research data. For further information about accreditation, contact Research.Support@ons.gov.uk or visit the SRS website.

Code availability

A copy of the analysis code is available at: https://github.com/UMCG-Global-Health/COVID-19_N-antibodies (<https://doi.org/10.5281/zenodo.13934702>).

6. References

1. World Health Organization. WHO Coronavirus (COVID-19) Dashboard.) (2024).
2. Bohning D, Rocchetti I, Maruotti A, Holling H. Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. *Int J Infect Dis* **97**, 197-201 (2020).
3. Ota S, Sugawa S, Suematsu E, Shinoda M, Izumizaki M, Shinkai M. Possibility of underestimation of COVID-19 prevalence by PCR and serological tests. *J Microbiol Immunol Infect* **55**, 1076-1083 (2022).

4. Liu W, *et al.* Evaluation of Nucleocapsid and Spike Protein-Based Enzyme-Linked Immunosorbent Assays for Detecting Antibodies against SARS-CoV-2. *J Clin Microbiol* **58**, (2020).
5. Suhandynata RT, Hoffman MA, Kelner MJ, McLawhon RW, Reed SL, Fitzgerald RL. Longitudinal Monitoring of SARS-CoV-2 IgM and IgG Seropositivity to Detect COVID-19. *J Appl Lab Med* **5**, 908-920 (2020).
6. Tsuchida T, *et al.* Back to normal; serological testing for COVID-19 diagnosis unveils missed infections. *J Med Virol* **93**, 4549-4552 (2021).
7. Wei J, *et al.* Anti-spike antibody response to natural SARS-CoV-2 infection in the general population. *Nat Commun* **12**, 6250 (2021).
8. Van Elslande J, *et al.* Antibody response against SARS-CoV-2 spike protein and nucleoprotein evaluated by four automated immunoassays and three ELISAs. *Clin Microbiol Infec* **26**, (2020).
9. Navaratnam AMD, *et al.* Nucleocapsid and spike antibody responses following virologically confirmed SARS-CoV-2 infection: an observational analysis in the Virus Watch community cohort. *Int J Infect Dis* **123**, 104-111 (2022).
10. Assis R, *et al.* Distinct SARS-CoV-2 antibody reactivity patterns elicited by natural infection and mRNA vaccination. *NPJ Vaccines* **6**, 132 (2021).
11. van den Hoogen LL, *et al.* Seropositivity to Nucleoprotein to detect mild and asymptomatic SARS-CoV-2 infections: A complementary tool to detect breakthrough infections after COVID-19 vaccination? *Vaccine* **40**, 2251-2257 (2022).

12. Burbelo PD, *et al.* Sensitivity in Detection of Antibodies to Nucleocapsid and Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus 2 in Patients With Coronavirus Disease 2019. *J Infect Dis* **222**, 206-213 (2020).
13. Follmann D, *et al.* Antinucleocapsid Antibodies After SARS-CoV-2 Infection in the Blinded Phase of the Randomized, Placebo-Controlled mRNA-1273 COVID-19 Vaccine Efficacy Clinical Trial. *Ann Intern Med* **175**, 1258-1265 (2022).
14. Van Elslande J, Gruwier L, Godderis L, Vermeersch P. Estimated Half-Life of SARS-CoV-2 Anti-Spike Antibodies More Than Double the Half-Life of Anti-nucleocapsid Antibodies in Healthcare Workers. *Clin Infect Dis* **73**, 2366-2368 (2021).
15. Alfego D, Sullivan A, Poirier B, Williams J, Adcock D, Letovsky S. A population-based analysis of the longevity of SARS-CoV-2 antibody seropositivity in the United States. *EClinicalMedicine* **36**, 100902 (2021).
16. Grebe E, *et al.* Detection of Nucleocapsid Antibodies Associated with Primary SARS-CoV-2 Infection in Unvaccinated and Vaccinated Blood Donors. *Emerg Infect Dis* **30**, 1621-1630 (2024).
17. Piccoli L, *et al.* Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **183**, 1024-1042 e1021 (2020).
18. Grzelak L, *et al.* Sex Differences in the Evolution of Neutralizing Antibodies to Severe Acute Respiratory Syndrome Coronavirus 2. *J Infect Dis* **224**, 983-988 (2021).
19. Gudbjartsson DF, *et al.* Humoral Immune Response to SARS-CoV-2 in Iceland. *N Engl J Med* **383**, 1724-1734 (2020).

20. Lumley SF, *et al.* The Duration, Dynamics, and Determinants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibody Responses in Individual Healthcare Workers. *Clin Infect Dis* **73**, e699-e709 (2021).
21. Donaldson M, McBride J. Performance Evaluation Report for the N antibody assay used for research purposes in the protocol. (2021). Available at: <https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/n-antibody-assay-performance>
22. Colman E, Puspitarani GA, Enright J, Kao RR. Ascertainment rate of SARS-CoV-2 infections from healthcare and community testing in the UK. *J Theor Biol* **558**, 111333 (2023).
23. Nightingale ES, *et al.* The local burden of disease during the first wave of the COVID-19 epidemic in England: estimation using different data sources from changing surveillance practices. *BMC Public Health* **22**, 716 (2022).
24. 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* **173**, 262-267 (2020).
25. Bahreini F, Najafi R, Amini R, Khazaei S, Bashirian S. Reducing False Negative PCR Test for COVID-19. *Int J MCH AIDS* **9**, 408-410 (2020).
26. Killingley B, *et al.* Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nat Med* **28**, 1031-1041 (2022).
27. Wolfel R, *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465-469 (2020).

28. Fajnzylber J, *et al.* SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun* **11**, 5493 (2020).
29. Terpos E, *et al.* SARS-CoV-2 antibody kinetics eight months from COVID-19 onset: Persistence of spike antibodies but loss of neutralizing antibodies in 24% of convalescent plasma donors. *Eur J Intern Med* **89**, 87-96 (2021).
30. Van Elslande J, *et al.* Longitudinal follow-up of IgG anti-nucleocapsid antibodies in SARS-CoV-2 infected patients up to eight months after infection. *J Clin Virol* **136**, 104765 (2021).
31. Walker AS, *et al.* Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. *Elife* **10**, (2021).
32. Yu W, Guo Y, Zhang S, Kong Y, Shen Z, Zhang J. Proportion of asymptomatic infection and nonsevere disease caused by SARS-CoV-2 Omicron variant: A systematic review and analysis. *J Med Virol* **94**, 5790-5801 (2022).
33. Birrell PJ, *et al.* Real-time modelling of the SARS-CoV-2 pandemic in England 2020-2023: a challenging data integration. *arXiv preprint arXiv:240804178*, (2024).
34. Smith LE, Potts HWW, Amlot R, Fear NT, Michie S, Rubin GJ. Adherence to the test, trace, and isolate system in the UK: results from 37 nationally representative surveys. *BMJ* **372**, n608 (2021).
35. Yamaoka Y, *et al.* Whole Nucleocapsid Protein of Severe Acute Respiratory Syndrome Coronavirus 2 May Cause False-Positive Results in Serological Assays. *Clin Infect Dis* **72**, 1291-1292 (2021).

36. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**, 861 (2018).
37. Pouwels KB, *et al.* Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey. *Lancet Public Health* **6**, e30-e38 (2021).
38. Fox T, *et al.* Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev* **11**, CD013652 (2022).
39. 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* **20**, 1390-1400 (2020).
40. Wei J, *et al.* Risk of SARS-CoV-2 reinfection during multiple Omicron variant waves in the UK general population. *Nat Commun* **15**, 1008 (2024).
41. Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recogn* **44**, 678-693 (2011).
42. Chabchoub Y, Fricker C. Classification of the Velib Stations Using Kmeans, Dynamic Time Wrapping and Dbw Averaging Method. *2014 International Workshop on Computational Intelligence for Multimedia Understanding (Iwcim)*, (2014).
43. Tavenard R, *et al.* Tslearn, A Machine Learning Toolkit for Time Series Data. *J Mach Learn Res* **21**, (2020).
44. Wickham H. ggplot2 : Elegant Graphics for Data Analysis. In: *Use R!*,). 2nd edn. Springer International Publishing : Imprint: Springer, (2016).

45. Wu Y, Kang L, Guo Z, Liu J, Liu M, Liang W. Incubation period of COVID-19 caused by unique SARS-CoV-2 strains: a systematic review and meta-analysis (vol 5, e2228008, 2022). *Jama Netw Open* **5**, (2022).
46. Hellewell J, *et al.* Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *Bmc Med* **19**, (2021).
47. Baillargeon S, Rivest L-P. Rcapture: Loglinear Models for Capture-Recapture in R. *Journal of Statistical Software* **19**, 1 - 31 (2007).

Acknowledgements

We are grateful for the support of all COVID-19 Infection Survey participants.

The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, Department of Health, or UKHSA. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data.

This work uses research datasets (<https://doi.org/10.57906/r47r-1735>) which may not exactly reproduce National Statistics aggregates.

The COVID-19 Infection Survey team

David W. Eyre^{5,6,8}, Nicole Stroesser^{1,4,5,6}, Philippa C. Matthews^{1,9,10}, Jia Wei^{1,6,8}, Ian Diamond¹¹, Ruth Studley¹¹, Nick Taylor¹¹, Emma Rourke¹¹, Tina Thomas¹¹, Dawid Pienaar¹¹, Joy Preece¹¹, Sarah Crofts¹¹, Lina Lloyd¹¹, Michelle Bowen¹¹, Daniel Ayoubkhani¹¹, Russell

Black¹¹, Antonio Felton¹¹, Megan Crees¹¹, Joel Jones¹¹, Esther Sutherland¹¹, Derrick W. Crook¹, Emma Pritchard¹, Karina-Doris Vihta¹, Alison Howarth¹, Brian D. Marsden¹, Kevin K. Chau¹, Lucas Martins Ferreira¹, Wanwisa Dejnirattisai¹, Juthathip Mongkolsapaya¹, Sarah Hoosdally¹, Richard Cornall¹, David I Stuart¹, Gavin Screaton¹, Katrina Lythgoe⁸, David Bonsall⁸, Tanya Golubchik⁸, Helen Fryer⁸, John N Newton¹², John I Bell¹³, Stuart Cox¹³, Kevin Paddon¹³, Tim James¹³, Thomas House¹⁴, Julie Robotham¹⁵, Paul Birrell¹⁵, Helena Jordan¹⁶, Tim Sheppard¹⁶, Graham Athey¹⁶, Dan Moody¹⁶, Leigh Curry¹⁶, Pamela Brereton¹⁶, Ian Jarvis¹⁷, Anna Godsmark¹⁷, George Morris¹⁷, Bobby Mallick¹⁷, Phil Eeles¹⁷, Jodie Hay¹⁸, Harper VanSteenhouse¹⁸, Jessica Lee¹⁹, Sean White²⁰, Tim Evans²¹, Lisa Bloemberg²⁰, Katie Allison²¹, Anouska Pandya²¹, Sophie Davis²¹, David I Conway²², Margaret MacLeod²², Chris Cunningham²²

8 Big Data Institute, University of Oxford, Oxford UK

9 The Francis Crick Institute, 1 Midland Road, London, UK

10 Division of Infection and Immunity, University College London, London, UK

11 Office for National Statistics, Newport, UK

12 Office of the Regius Professor of Medicine, University of Oxford, Oxford, UK

13 Oxford University Hospitals NHS Foundation Trust, Oxford, UK

14 University of Manchester, Manchester, UK

15 UK Health Security Agency, London, UK

16 IQVIA, London, UK

17 National Biocentre, Milton Keynes, UK.

18 Glasgow Lighthouse Laboratory, London, UK

19 Department of Health and Social Care, London, UK

20 Welsh Government, Cardiff, UK

21 Scottish Government, Edinburgh, UK

22 Public Health Scotland, Edinburgh, UK

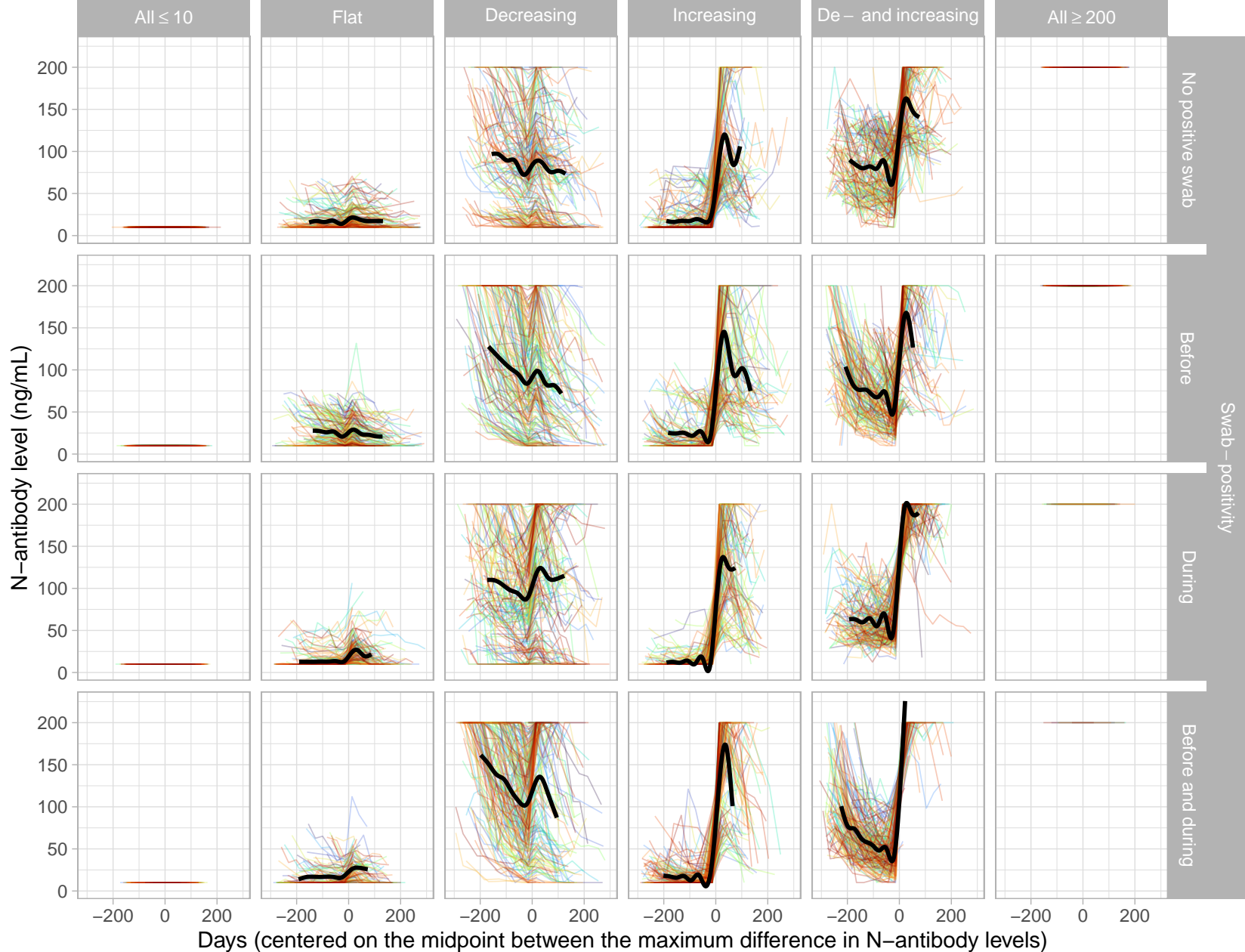
Author Contributions

The COVID-19 Infection Survey was designed, planned and conducted by ASW, KBP and the COVID-19 Infection Survey Team. This specific analysis was designed by ASW, KBP, LRZ and TEAP. LRZ conducted all statistical analysis. LRZ, ASW and KBP drafted the manuscript and all authors contributed to interpretation of the data and results and revised the manuscript. TEAP, KBP, and ASW contributed equally. All authors approved the final version of the manuscript.

Competing interests

This study was funded by the UK Health Security Agency and the Department of Health and Social Care with in-kind support from the Welsh Government, the Department of Health on behalf of the Northern Ireland Government and the Scottish Government. ASW and 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 the UK Health Security Agency (UK HSA) (NIHR200915). ASW is also supported by the NIHR Oxford Biomedical Research Centre. KBP is also supported by the Huo Family Foundation. There are no other conflicts of interest.

N – antibody trajectory – based positivity



Swab-positivity

