FAST TRACK

Accuracy of UK Rapid Test Consortium (UK-RTC) "AbC-19 Rapid Test" for detection of previous SARS-CoV-2 infection in key workers: test accuracy study

Ranya Mulchandani,[1*] Hayley E Jones,[2*] Sian Taylor-Phillips,[3] Justin Shute,[1] Keith Perry,[1] Shabnam Jamarani,[1] Tim Brooks,[1] Andre Charlett,[1] Matthew Hickman,[2] Isabel Oliver,[1] Stephen Kaptoge,[4] John Danesh,[4] Emanuele Di Angelantonio,[4] Anthony E Ades,[2] David H Wyllie,[1] on behalf of the EDSAB-HOME and COMPARE Investigators

[1]Public Health England, London, UK

[2]University of Bristol, Bristol, UK

[3]University of Warwick, Coventry, UK

[4]University of Cambridge, Cambridge, UK

*Contributed equally

Correspondence to: D Wyllie
David.wyllie@phe.gov.uk
(or @davidwyllie20 on Twitter:
ORCID 0000-0002-9155-6228)

ABSTRACT

OBJECTIVE
To assess the accuracy of the AbC-19 Rapid Test lateral flow immunoassay for the detection of previous severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection.

DESIGN
Test accuracy study.

SETTING
Laboratory based evaluation.

PARTICIPANTS
2847 key workers (healthcare staff, fire and rescue officers, and police officers) in England in June 2020 (268 with a previous polymerase chain reaction (PCR) positive result (median 63 days previously), 2579 with unknown previous infection status); and 1995 pre-pandemic blood donors.

MAIN OUTCOME MEASURES
AbC-19 sensitivity and specificity, estimated using known negative (pre-pandemic) and known positive (PCR confirmed) samples as reference standards and secondly using the Roche Elecsys anti-nucleoprotein assay, a highly sensitive laboratory immunoassay, as a reference standard in samples from key workers.

RESULTS
Test result bands were often weak, with positive/negative discordance by three trained laboratory staff for 3.9% of devices. Using consensus readings, for known positive and negative samples sensitivity was 92.5% (95% confidence interval 88.8% to 95.1%) and specificity was 97.9% (97.2% to 98.4%). Using an immunoassay reference standard, sensitivity was 94.2% (90.7% to 96.5%) among PCR confirmed cases but 84.7% (80.6% to 88.1%) among other people with antibodies. This is consistent with AbC-19 being more sensitive when antibody concentrations are higher, as people with PCR confirmation tended to have more severe disease whereas only 62% (218/354) of seropositive participants had had symptoms. If 1 million key workers were tested with AbC-19 and 10% had actually been previously infected, 84 700 true positive and 18 900 false positive results would be projected. The probability that a positive result was correct would be 81.7% (76.8% to 85.8%).

CONCLUSIONS
AbC-19 sensitivity was lower among unselected populations than among PCR confirmed cases of SARS-CoV-2, highlighting the scope for overestimation of assay performance in studies involving only PCR confirmed cases, owing to "spectrum bias." Assuming that 10% of the tested population have had SARS-CoV-2 infection, around one in five key workers testing positive with AbC-19 would be false positives.

STUDY REGISTRATION
ISRCTN 56609224.

WHAT IS ALREADY KNOWN ON THIS TOPIC

A previous study reported 97.7% sensitivity and 100% specificity of the AbC-19 Rapid Test to detect SARS-CoV-2 antibodies

However, these findings may be optimistic because the study excluded participants with unclear antibody status according to laboratory based immunoassays

Most studies evaluating accuracy of SARS-CoV-2 antibody tests have used a "two gate" or "diagnostic case-control" design, but the effect of spectrum bias on estimates of test sensitivity is unclear

WHAT THIS STUDY ADDS

Using pre-pandemic samples, the specificity of AbC-19 was 97.9% (97.2% to 98.4%), implying that, at 10% prevalence of previous infection, around one in five positive results would be incorrect

Evidence of spectrum bias existed, with ~10% higher sensitivity (absolute difference) in key workers with polymerase chain reaction confirmed infection, suggesting that two gate studies may overestimate sensitivity

Trained laboratory staff reported that test bands were often weak and disagreed on device positivity for 3.9% of tests, so accuracy may be lower if used by members of the public

Introduction

After infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), most, but not all, infected people generate antibodies against the viral spike (S) or nucleoprotein (N) antigen.[1 2] Several lateral flow immunoassays (LFIAs)—small, pregnancy test format devices that can deliver testing rapidly and at scale—have recently become available that detect antibodies against SARS-CoV-2 proteins. These devices have two potential main uses: population serosurveillance and assessment of individual risk of developing immunity to coronavirus disease 2019 (covid-19).[3-5] The US Food and Drug Administration recently (24 September 2020) licensed an LFIA for office use (that is, in supervised environments), on the basis that it may predict immunity (https://www.fda.gov/media/139789/download); if this is validated, then it could result in

widespread use of this class of devices. However, this application is also critically dependent on high test accuracy: if assay specificity is not sufficiently high, false positive results could result in non-immune people being assigned to high SARS-CoV-2 exposure environments.[6 7]

Evaluations of different LFIAs for SARS-CoV-2 have produced widely varying estimates of accuracy.[2 8] In some jurisdictions, large numbers of devices are available, with varying levels of performance characterisation.[9] In a recent Cochrane review,[2] most LFIA evaluations were considered to be at high risk of bias, owing to use of a "two gate" (also known as "diagnostic case-control") design.[10] These studies evaluate LFIAs on a set of pre-pandemic blood samples and on a separate set of convalescent samples taken from cases of SARS-CoV-2 confirmed by polymerase chain reaction (PCR). This study design has been associated with overestimation of test accuracy on average across multiple clinical settings,[10 11] referred to as "spectrum bias."[12] The sensitivity of SARS-CoV-2 antibody tests based on PCR confirmed cases could be overestimated if people with more severe disease are more likely to have been PCR tested than those with milder illness, and if, as is known to occur,[13-15] people with more severe illness produce higher antibody concentrations. This tendency would make confirmed cases easier to diagnose than cases in other people who, although previously infected, were not PCR tested. Despite the importance of obtaining real world estimates of test accuracy, however, potential spectrum bias has not been assessed to date in SARS-CoV-2 antibody testing. This is partly due to the need for larger sample sizes in the alternative approach of assessing test accuracy directly in a target cohort (a so-called "one gate" design) and partly due to the lack of a true gold standard test to assess previous infection in such a cohort.

This study helps to reduce uncertainties in the evaluation of a new SARS-CoV-2 rapid antibody test, the "AbC-19 Rapid Test" (AbC-19 hereafter). Produced by the UK Rapid Test Consortium (UK-RTC), the test contains trimeric SARS-CoV-2 S protein attached to a membrane, which, when recognised by IgG antibodies, results in formation of a band visible to the user. Professional use CE marking has been obtained by self-certification. The PANDEMIC study from Ulster University, which is not yet peer reviewed, reports sensitivity of 97.7% (n=304; 95% confidence interval 95.7% to 99.3%) and specificity of 100% (n=350; 95% confidence interval reported as 100% to 100%) to detect presence of antibodies.[16] This suggests that the AbC-19 test gives no false positive results. However, these may be overestimates of test accuracy, as blood samples from a number of people were excluded from this test evaluation on the basis of results of other tests (pre-selection). We evaluated AbC-19 in the detection of previous infection with SARS-CoV-2 by using venous blood samples collected from UK key workers in the EDSAB-HOME (Evaluating Detection of SARS-CoV-2 AntiBodies at HOME) study[15] and pre-pandemic samples from blood donors.

Methods

Study participants

In total, across the two parts of the study, we analysed blood samples from 4842 people: 2847 people recruited through the EDSAB-HOME study,[15] and an additional 1995 pre-pandemic samples from the COMPARE blood donor study. Each sample was from a separate individual.

EDSAB-HOME study: recruitment

EDSAB-HOME is a study of key workers in the UK (ISRCTN 56609224).[15] People were eligible to enrol in the study if they were currently working at their place of work, aged 18 years or over, and able to read English and had an in-use personal email address and mobile phone number. Anyone who was currently experiencing covid-19 compatible symptoms, had experienced any such symptoms in the previous seven days, met the government's criteria for "exceptionally vulnerable," was unable to read normal sized print, or was taking part in any covid-19 vaccine trials was not eligible for enrolment. Symptom history was not part of the eligibility criteria. All data and samples were collected in June 2020 and form a convenience sample. The purpose of the study was to assess the accuracy of LFIAs suitable for deployment at scale in key workers in the UK. All participants completed an online questionnaire, which included information on any previous tests they had had for SARS-CoV-2, their beliefs about whether they had had covid-19, and any signs and symptoms that they or their household members had experienced since 1 January 2020. On the basis of this information, we categorised participants by using a version of the World Health Organization's criteria for confirmed, suspected, and probable cases, adapted to the UK context and taking into account the confirmation of community transmission on 5 March 2020 (supplementary material, table S1).[15] Previous PCR status was self-reported and validated by comparison with national laboratory records after recruitment. Recruitment into the study was through three streams: fire and police officers (stream A), healthcare workers who were recruited without any restrictions on previous PCR status (stream B), and healthcare workers recruited only if they had a history of SARS-CoV-2 PCR positivity (stream C) (fig 1). On attending a study clinic, each participant had a venous blood sample taken, which was anticoagulated using EDTA. In participants with previous PCR positivity, onset of illness occurred a median of 63 (interquartile range 52-75) days before the clinic visit. Plasma was separated by centrifugation and stored at −80°C until use. No samples were frozen and thawed more than five times. Demographic details are in the supplementary materials (table S2), as is additional information on the relation between PCR testing and symptoms.

COMPARE study

We used stratification by age, sex, and region to randomly select 2000 participants as "known negative" samples from among a pool of almost 29 029
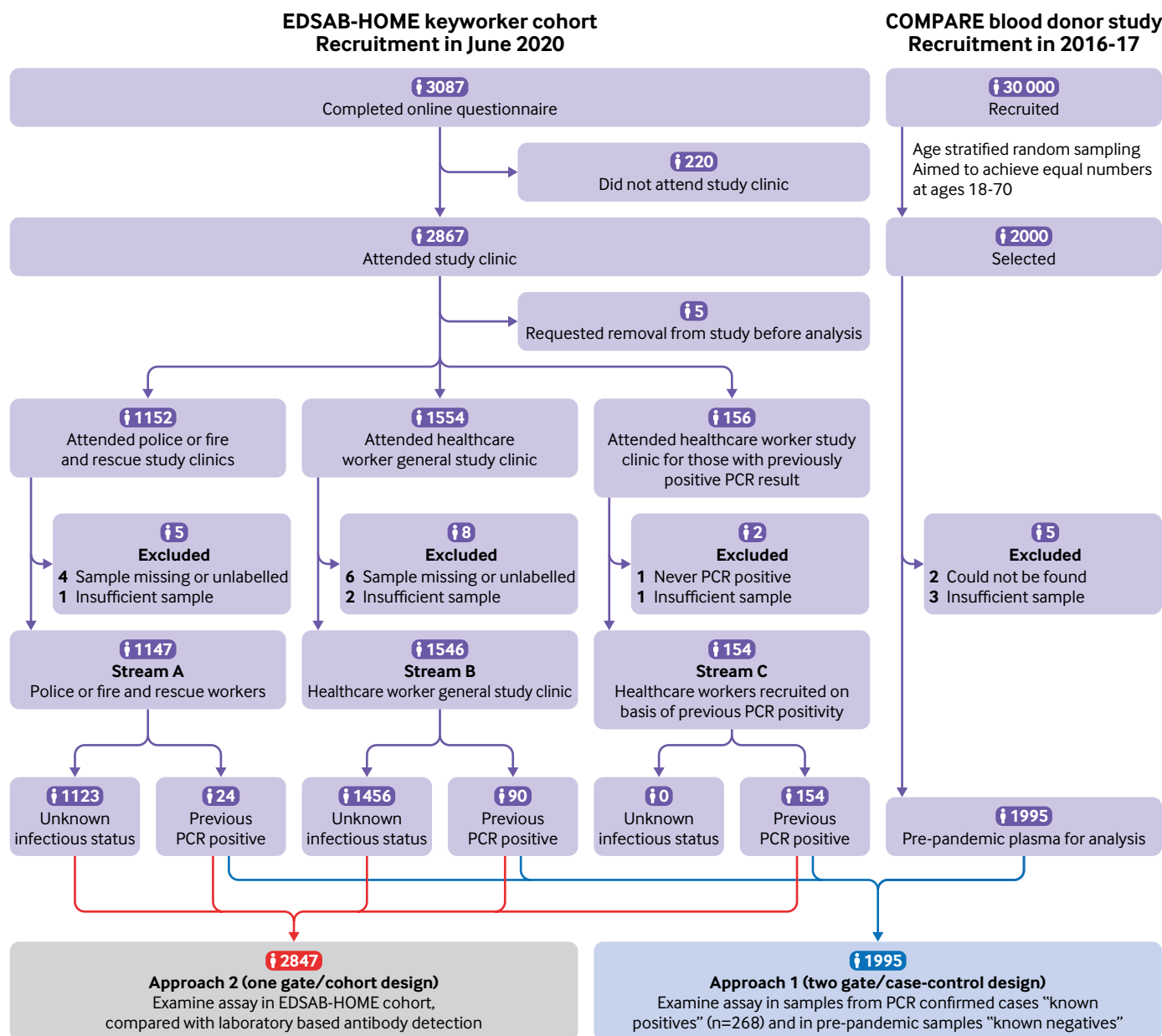
Fig 1 | Flow diagram of recruitment and analysis. PCR=polymerase chain reaction

participants recruited into the COMPARE study, a pre-pandemic (2016-17) blood donor cohort in England (ISRCTN 90871183). Of these participants, 1995 had samples available for analysis (fig 1). The selection strategy aimed to maximise national geographical representativeness of the sample, with approximately equal numbers of people across participating age groups (18-70 years). Demographic details are in the supplementary materials (table S3).

Approach 1: PCR positive cases and pre-pandemic samples as reference standard

We used two approaches to estimate the accuracy of AbC-19, both pre-specified (fig 1). In approach 1, we estimated the sensitivity and specificity of AbC-19 by using two sets of samples with known previous

infection status: pre-pandemic samples ("known negatives") from the COMPARE study (n=1995) and samples collected from participants in EDSAB-HOME who self-reported having had a previous PCR positive test for SARS-CoV-2 ("known positives") (n=268).

Approach 2: Laboratory immunoassays as reference standard

When the EDSAB-HOME protocol was written, large scale development and evaluation of SARS-CoV-2 immunoassays was an emerging field. We pre-specified the use of a laboratory reference standard indicative of previous SARS-CoV-2 infection as a single assay with the highest published accuracy available to us after sample collection; as anticipated in the protocol, this was the Roche Elecsys assay, which

measures antibodies against the N protein.[17] We also analysed all samples with the EUROIMMUN anti-S IgG immunoassay, which detects antibodies against the S protein S1 domain, one domain present in the AbC-19 test antigen. We repeated immunoassays that failed for technical reasons, in accordance with usual clinical practice, and excluded samples with insufficient volume for testing from analyses. Although the primary reference standard used here is the Roche Elecsys immunoassay (threshold 1.0), having previously noted the strong positive correlation between Roche Elecsys and EUROIMMUN assay results (ρ=0.93[15]), in sensitivity analyses we also used a EUROIMMUN immunoassay (threshold 0.8, as described[15]) and a composite reference standard defined as positive on either Roche Elecsys or EUROIMMUN versus negative on both. We made these comparisons separately for the known positive (n=268) and all other EDSAB-HOME samples (n=2579), to assess for potential spectrum bias associated with estimating test sensitivity from PCR confirmed cases relative to other people with evidence of previous infection. We refer to the latter sample set (n=2579) as samples from people with unknown previous infection status. We analysed samples from 2847 EDSAB-HOME study participants (fig 1), a group comprised of police, fire, and healthcare staff who are of working age and worked throughout lockdown.[15]

Laboratory protocol
Test evaluation (product FG-FD51919, lot A2007003) was performed in an accredited WHO Pre-Qualification Evaluating Laboratory based in PHE Colindale, London, by experienced laboratory staff following training by the manufacturer. Trained staff dropped 2.5 µL of plasma into the lateral flow device followed by 100 µL of chase buffer; devices give a qualitative (positive or negative) result for presence of IgG antibodies and include a control band. A result was considered positive if two bands were present and negative if only the control band was present. Devices in which a control band did not develop were considered to have given invalid results (n=5) and were repeated. Three readers independently read each device using the WHO scoring system for subjectively read assays (0=negative; 1=very weak, but definitely reactive; 2=medium to strong reactivity).[18] For assessment of test sensitivity and specificity, and in measuring discordance in AbC-19 positivity across the multiple readers, scores of 1 and 2 were grouped together as "positive." If the three independent readers disagreed on the positivity of a sample, the majority reading was taken as the "overall" or consensus test result in our primary analysis, as per the WHO protocol.[18]

None of the staff who ran the laboratory immunoassays had access to any clinical information on the samples, and laboratory immunoassays were all completed before any testing of the index test. For all EDSAB-HOME samples and the first 350 pre-pandemic samples analysed, readers were blind to the reference standard (including randomising order of known

positives and negatives), as well as to any information on study participants or results on other assays. The additional 1650 pre-pandemic samples were analysed after the other samples, for which complete blinding was difficult. We retested known negative samples if they tested positive on the AbC-19 device. For all other samples, we retested if the result was different from the composite reference standard defined above. Our primary results are those based on the initial set of readings for each sample. Results following retesting are reported as secondary, as only discordant samples were retested, which does not reflect how the test would be used in the real world. More details are in the supplementary materials.

Statistical analysis
We estimated 95% confidence intervals around proportions by using Wilson's method in R 3.5.1. In approach 1, we estimated AbC-19 test sensitivity and specificity as the proportion of known positive samples that tested positive and the proportion of known negative samples that tested negative, respectively. Estimates of sensitivity are also presented by days since onset of symptoms and separately for people who reported no symptoms. Test specificity estimates are further reported by age group.

In approach 2, AbC-19 test sensitivity and specificity are reported for each of the immunoassay reference standards described above. AbC-19 test results are also stratified by adapted WHO categorisation. To facilitate evaluation of potential spectrum bias associated with estimating test sensitivity from PCR confirmed cases, we stratified approach 2 estimates by whether participants were known to have had PCR confirmed SARS-CoV-2 infection (known positive sample set: n=268) or whether their previous infection status was unknown before antibody testing (n=2579). However, to reflect the "one gate" nature of EDSAB-HOME streams A and B, we also report as a secondary analysis results from all stream A and B participants, regardless of previous PCR positivity (n=2693).

Additional post hoc analyses assessed the association between age, sex, and ethnicity and specificity, using multivariable logistic regression; anti-S1 antibody indexes (EUROIMMUN) and AbC-19 test sensitivity; and the level of disagreement between three independent reviewers. These analyses are described in the supplementary materials.

Sample size calculations
Study size was based on sample size calculations done as part of the EDSAB-HOME research protocol (see supplementary materials).

Positive and negative predictive values
We calculated positive and negative predictive values for prevalence of previous infection of 5%, 10%, 15%, 20%, and 25% as functions of estimated sensitivity and specificity. We calculated 95% confidence intervals by using a simulations based approach in R, accounting for the sampling variation (uncertainty) in both

estimates. We used the estimates of specificity based on pre-pandemic sera (known negatives). We present two sets of estimates of positive predictive value and negative predictive value based on alternative estimates of test sensitivity: sensitivity based on analysis of PCR confirmed cases and sensitivity based on comparison with the primary (Roche) immunoassay reference standard in people with unknown previous infection status.

Accounting for multiple readers

For each sample set (known positives, known negatives, and unknown previous infection status), we report numbers of disagreements between the three independent reviewers. In the presence of discordant results across readers, we anticipated that estimating test sensitivity and specificity based on the majority reading would lead to a slight upward bias in these estimates. To explore the potential extent of this, within approach 1 we did additional analyses in which each reading was treated as a separate "test." We fitted multinomial distributions to tables of cross classified readings, assuming a log-linear model for underlying probabilities. Full details are available in the supplementary materials.

Patient and public involvement

As part of the EDSAB-HOME study design, patient and public involvement was provided by a public panel run through NIHR Birmingham Biomedical Research Centre and NIHR Applied Research Collaboration West Midlands. The panel commented on study design, acceptability, and strategies for effective recruitment. The questionnaire, communications with the participants, and all public facing documentation were informed and refined by this engagement. We did not seek public engagement in the design of the laboratory aspects of the study or in the interpretation of the results.

Results

Approach 1: sensitivity and specificity in "known positives" and "known negatives"

In 1995 pre-pandemic known negative samples from the COMPARE study, test specificity was 97.9% (95% confidence interval 97.2 to 98.4%) (table 1). Specificity was slightly lower in women (96.9%, 95.6% to 97.8%) than in men (98.9%, 98.0% to 99.4%) (adjusted odds ratio favouring false positivity in women 2.9, 95% confidence interval 1.5 to 5.9). After retesting of discordant samples, a female excess was still evident (adjusted odds ratio 4.7, 1.4 to 17) (table S4).

Among 268 (9.4%) EDSAB-HOME participants who had had a positive PCR test for SARS-CoV-2, test sensitivity was 92.5% (88.8% to 95.1%) (table 2). In this group, the first positive PCR test had occurred a median of 59 days previously. Days since first positive test was 21 or more in 251/268 cases. Sensitivity was estimated to be substantially lower among the 12 participants who reported no symptoms (sensitivity 58.3%, 32.0% to 80.7%) than among the 256 who reported symptoms (sensitivity 94.1%, 90.6% to 96.4%), and may decline slightly in PCR positive people with longer times between symptoms and blood sampling (table 2; figure S1).

False negative AbC-19 tests occurred predominantly in samples with lower EUROIMMUN anti-S indices (fig 2, A; figure S1). Three participants (all asymptomatic) had clear negative results on both immunoassays in addition to the AbC-19 device.

Approach 2: analysis of "unknown previous infection status" EDSAB-HOME study samples

In EDSAB-HOME participants with unknown previous infection status (all those who had not had a positive PCR test; fig 2, B), antibody indices from anti-N and anti-S assays were highly correlated (Spearman's ρ=0.93; P<0.001). We observed clear delineation between two main clusters of participants (with versus without antibodies), whereas the true previous infection status of a small minority of samples is less clear.

Among participants with unknown previous infection status testing positive on the Roche reference standard, 62% (218/354) reported symptoms compatible with covid-19 since 1 January 2020. The remaining 136 included 55 (16%) people who reported no compatible symptoms and 81 (23%) who "did not know" whether they had had any compatible symptoms.

Table 3 shows estimates of sensitivity and specificity of the AbC-19 test based on comparison with the three immunoassay reference standards, across all EDSAB-HOME study participants. Among the unknown previous infection status samples, sensitivity was 84.7% (80.6% to 88.1%) and specificity was 98.9% (98.4% to 99.3%) on the basis of the Roche reference standard (table 3). This was lower than sensitivity among the PCR confirmed cases, which was estimated as 94.2% (90.7% to 96.5%) on the basis of the same reference standard. In sensitivity analyses, we observed a similar pattern on the basis of the EUROIMMUN reference standard and a composite of the two assays (table 3). This is expected given the strong positive correlations between the immunoassay results studied (fig 2).

Table 1	Approach 1: specificity of AbC-19 Rapid Test estimated from 1995 pre-pandemic samples ("known negatives")			
Result	True negatives	False positives	Total known negatives	Specificity, % (95% CI)
Primary result (result of first test*)	1953	42	1995	97.9 (97.2 to 98.4)
After re-testing of 42 false positives only	1979	16	1995	99.2 (98.7 to 99.5)

*2/1995 samples yielded invalid results (control band failed) on first test and were repeated; re-test value (both negative) was used for these as part of primary result.

Table 2 | Approach 1: sensitivity of AbC-19 Rapid Test based on samples from 268 people with previous positive polymerase chain reaction test ("WHO confirmed" category: "known positives")

Time since symptom onset	True positives	False negatives	Total known positives	Sensitivity, % (95% CI)
All	248	20	268	92.5 (88.8 to 95.1)
Asymptomatic	7	5	12	58.3 (32.0 to 80.7)
All symptomatic	241	15	256	94.1 (90.6 to 96.4)
8-21 days	5	0	5	100 (56.6 to 100.0)
22-35 days	19	1	20	95.0 (76.4 to 99.7)
36-70 days	136	6	142	95.8 (91.1 to 98.0)
≥71 days	81	8	89	91.0 (83.3 to 95.4)
All after retests*	251	17	268	93.7 (90.1 to 96.0)

*Only subset of samples retested: see main text for details.

Retesting discordant samples did not substantially alter these estimates (table S4). Test accuracy estimates did not vary across recruitment stream A or B (tables S5 and S6). Based on comparison of test results in all EDSAB-HOME stream A and B samples with the Roche reference standard, regardless of previous PCR positivity (one gate estimate), AbC-19 test sensitivity was estimated as 86.6% (83.2% to 89.4%) (table S7).

Fig 2 | Relation between AbC-19 test results and immunoassay indices. A: AbC-19 Rapid Test results by EUROIMMUN and Roche Elecsys assay titre, showing whether participant was asymptomatic, among 268 "known positives" (approach 1). B: AbC-19 results by EUROIMMUN and Roche Elecsys titre among all 2579 "unknown previous infection status" EDSAB-HOME stream A and B participants (ie, all study participants except cases confirmed by polymerase chain reaction) (approach 2). Lines reflect Roche threshold (1.0) and EUROIMMUN rule-out (0.8) and rule-in (1.1) thresholds recommended by manufacturers. 0.8 threshold was used for EUROIMMUN in this study

Between reader variation

With the AbC-19 device, many bands were weak visually; for example, among the 462 samples from streams A and B that tested positive with the Roche reference standard, consensus scoring found that 62 (13%) had no visible band, 199 (43%) had a band scored 1, and 201 (44%) had a band scored 2. Expert readers differed in rating the device as positive or negative in 189/4842 (3.9%) devices, requiring a consensus decision (table S8). A log-linear regression model fitted to the known positives and known negatives estimated the sensitivity and specificity of an individual reader reading the device as 91.4% (95% credible interval 88.2% to 93.9%) and 97.6% (97.0% to 98.1%), respectively (supplementary material, table S8). These point estimates are lower than the estimates derived when using the consensus approach used in WHO evaluations with sensitivity and specificity of 92.5% (88.8% to 95.1%) and 97.9% (97.2% to 98.4%) (table 1 and table 2).

Different test sensitivity in PCR positive versus immunoassay positive samples

Among participants from streams A and B with immunoassay positive results, we investigated why AbC-19 test sensitivity in people who had not had a previous positive PCR test was less than in those who had had confirmation by PCR. As the AbC-19 test detects anti-S antibody, we compared results with anti-S antibody indices obtained using the EUROIMMUN immunoassay. PCR positive participants had higher anti-S antibody titres (median 4.14 v 5.63 units; P=0.001; fig 3, A). As the AbC-19 test band strength observed by the three observers increases, so too does the anti-S antibody index detected by EUROIMMUN immunoassays, showing that band strength is proportional to antibody concentrations (fig 3, B). Importantly, at low antibody concentrations, corresponding to EUROIMMUN indices of less than 3, the AbC-19 test is insensitive (fig 4). However, such low concentrations are common: 84/263 (32%) participants who had previously been PCR positive had anti-S antibody indexes less than 3, compared with 146/372 (39%) who had not but were seropositive (fig 3, A). In summary, compared with a less selected population, fewer PCR positive people had the low antibody concentrations at which the AbC-19 test is insensitive.

Positive and negative predictive values

Table 4 shows estimates of positive and negative predictive value. Given the different estimates of sensitivity noted above from approaches 1 and 2, we observed some differences across the two sets of estimates of positive and negative predictive value presented. We consider the approach 2 estimates to be most applicable to a potential target testing population of adult frontline workers seeking antibody testing, as we would expect people with PCR confirmation to be less likely to seek antibody testing.

Assuming a scenario of testing 1 million key workers with unknown previous infection status, of whom 10% had actually been previously infected, we estimate that approximately 84 700 true positive and

Table 3 | Approach 2: AbC-19 results on all EDSAB-HOME samples, compared with immunoassay reference standards

Reference	AbC-19 positive	AbC-19 negative	Total	Proportion, % (95% CI)
"Known positive" samples (n=268)				
Roche Elecsys:				
Positive	244	15	259	Sensitivity of AbC-19: 94.2 (90.7 to 96.5)
Negative	4	5	9	Agreement: 55.6 (26.7 to 81.1)
EUROIMMUN:				
Positive	240	10	250	Sensitivity of AbC-19: 96.0 (92.8 to 97.8)
Negative	8	10	18	Agreement: 55.5 (33.7 to 75.4)
Composite reference standard:				
Positive	247	16	263	Sensitivity of AbC-19: 93.9 (90.3 to 96.2)
Negative	1	4	5	Agreement: 80.0 (37.6 to 99.0)
Total	248	20	268	Proportion positive on AbC-19: 92.5 (88.8 to 95.1)
All "unknown previous infection status" samples (n=2579)				
Adapted WHO classification:				
Suspected	168	228	396	42.4 (37.7 to 47.3)
Early probable	12	133	145	8.3 (4.8 to 13.9)
Uncertain	22	123	145	15.2 (10.2 to 21.9)
No	122	1771	1893	6.4 (5.4 to 7.6)
Roche Elecsys:				
Positive	300	54	354	Sensitivity of AbC-19: 84.7 (80.6 to 88.1)
Negative	24	2201	2225	Specificity of AbC-19: 98.9 (98.4 to 99.3)
EUROIMMUN:				
Positive	296	50	346	Sensitivity of AbC-19: 85.5 (81.5 to 88.9)
Negative	28	2205	2233	Specificity of AbC-19: 98.7 (98.2 to 99.1)
Composite reference standard:				
Positive	303	69	372	Sensitivity of AbC-19: 81.5 (77.2 to 85.1)
Negative	21	2186	2207	Specificity of AbC-19: 99.0 (98.5 to 99.4)
Total	324	2255	2579	Proportion positive on AbC-19: 12.6 (11.3 to 13.9)

Results are presented separately for "known positive" (WHO "confirmed" category: n=268) and "unknown previous infection status" (all other EDSAB-HOME participants: n=2579) samples.
Composite reference standard was "positive on at least one of Roche Elecsys or EUROIMMUN versus negative on both."
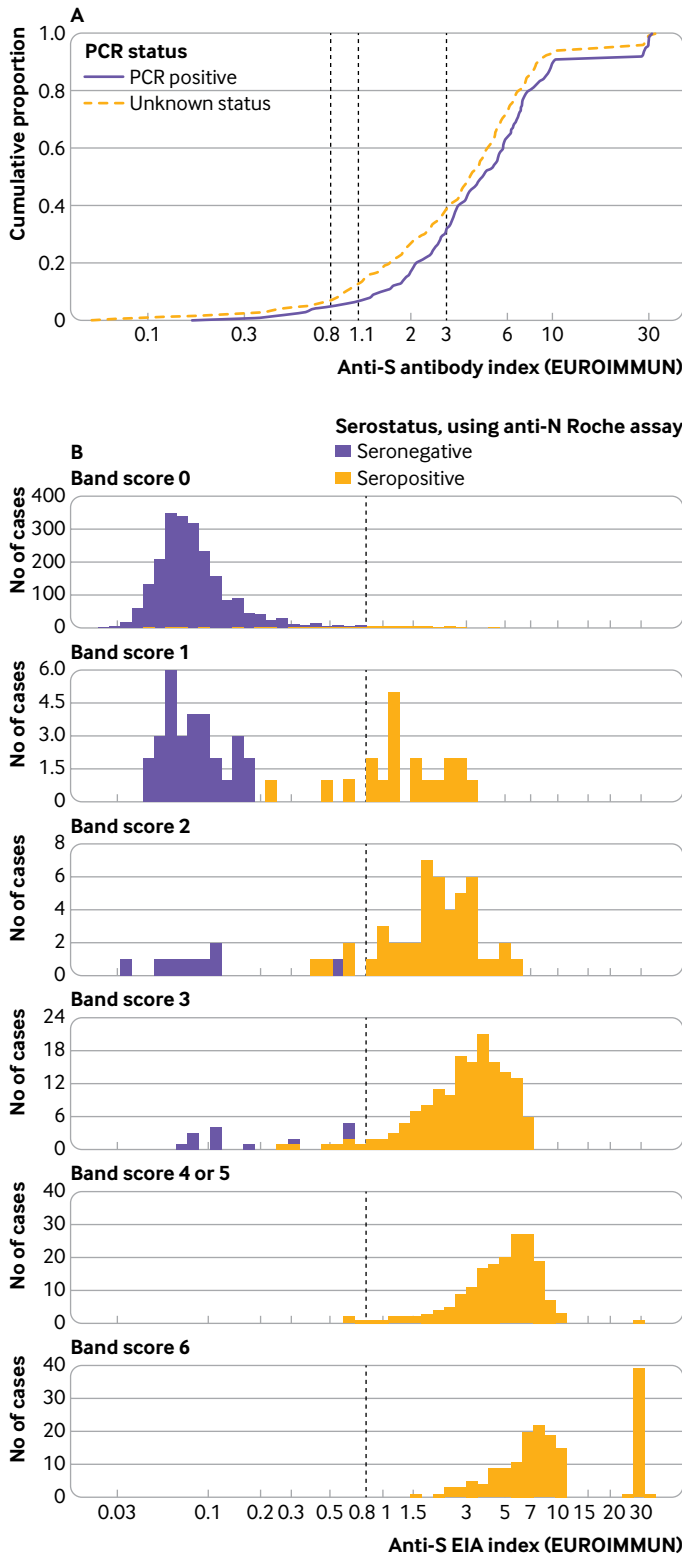
Fig 3 | Relation between AbC-19 test results and anti-S immunoassay indices. A: cumulative anti-S (EUROIMMUN) antibody index, stratified by polymerase chain reaction (PCR) status. B: AbC-19 test band strength and its relation to anti-S (EUROIMMUN) antibody index. Test bands strength refers to sum of test band intensities (scored as 0, 1, or 2) by three observers. EIA=enzyme immunoassay

18 900 false positive results would occur, meaning that approximately one in five positive test results would

be false positive (positive predictive value of 81.7%, 76.8% to 85.8%; table 4).

Discussion

In a series of analyses involving two complementary and distinct cohorts, we made several observations relevant to the performance of emerging, scalable antibody tests for SARS-CoV-2. Firstly, on the basis of an analysis of pre-pandemic blood samples from almost 2000 blood donors in England, we estimated the specificity of the AbC-19 test to be 97.9% (95% confidence interval 97.2% to 98.4%). That is, 2.1% of people who did not have a previous SARS-CoV-2 infection incorrectly tested positive. Next, we analysed samples from 2847 key workers in the UK. Among those who self-reported previous PCR positivity, we estimated the sensitivity of the AbC-19 test to be 92.5% (88.8% to 95.1%). In this group, participants with PCR confirmed covid-19 had symptom onset a median of 63 (interquartile range 52-75) days before the sample was taken, which is sufficient for antibody development.[2 17] This group is not, however, a representative sample of all people who have been infected with SARS-CoV-2: we would expect asymptomatic and lower severity cases to be under-represented, particularly given limited test availability in the early stage of the pandemic when many people became ill. We found evidence that this estimate of test sensitivity is upwardly biased, owing to lack of representativeness, by comparing AbC-19 test results with results on a sensitive immunoassay[17] separately among those with previous PCR confirmation versus those with unknown previous infection status before antibody testing. Among the participants with PCR confirmed infection, comparison with this alternative reference standard produced a similar estimate of sensitivity (94.2%, 90.7% to 96.5%).

By contrast, among EDSAB-HOME participants who did not know or were unsure of their previous infection status (that is, all participants who had not had a previous positive PCR test, a likely user group for the test), we estimated the sensitivity of the AbC-19 test to be lower (84.7%, 80.6% to 88.1%) using the same immunoassay reference standard. This "spectrum bias"[12] effect occurred because the PCR positive participants had, on average, higher antibody concentrations than other seropositive people. As people who have had SARS-CoV-2 previously confirmed by PCR would probably be less likely to seek antibody testing than other people, we consider this lower estimate (84.7%) to be a more appropriate estimate of test sensitivity for a potential use case in which people were to choose to take the test to find out their own previous infection status. For an alternative potential use of population serosurveillance,[3 4] the estimate of test sensitivity of 86.6% (83.2% to 89.4%), which derives from an analysis of all EDSAB-HOME stream A and B participants, irrespective of PCR positivity, may be more appropriate. Finally, we observed that AbC-19 test bands were often weak, a finding also reported in another laboratory based evaluation of AbC-19.[16] In

Fig 4 | Sensitivity of AbC-19 test at varying levels of anti-S antibody index. LFIA=lateral flow immunoassay

standard" test to evaluate a new test against is lacking. In this study, we aimed to estimate sensitivity and specificity to detect previous SARS-CoV-2 infection. Our approach to estimating test specificity based on pre-pandemic samples (collected in 2016-17) is therefore robust, as we can be confident that these samples were taken from people never infected with SARS-CoV-2. Unbiased estimation of test sensitivity for this target is, unfortunately, more difficult. We took presence of antibodies according to a sensitive immunoassay to be a proxy for previous infection.[1][2][15] After infection, most, but not all, infected people seroconvert.[12] However, antibody concentrations decline over time.[13][15][19] As our approach will necessarily miss any individuals without antibodies, we may have overestimated the sensitivity of the AbC-19 test to detect previous infection, particularly over a longer time period. Furthermore, results on the immunoassays we used as reference standards are themselves subject to some small degree of misclassification as to whether antibodies are present.

Secondly, we evaluated AbC-19 in a laboratory setting, using venous blood samples, rather than having participants do the test themselves using finger prick samples. Tests were read manually, as specified in the manufacturer's instructions for use; we used a majority of three reads as the test result, which we have shown may overestimate test sensitivity by about 1% compared with a single trained laboratory reader. The sensitivity and specificity of the test if used by untrained members of the general public at home may be lower still, a factor that would need further investigation if these tests were to be used by the general population.[3] Thirdly, our cohort of key workers included few people aged over 65 years, suggesting the need for further evaluation of the assay at older ages when risk of covid-19 is substantially higher.

Finally, of the 354 participants with a positive Roche immunoassay result who had unknown previous infection status before antibody testing (the denominator for our estimate of 84.7% test sensitivity), 62% reported symptoms compatible with covid-19 since 1 January 2020. Although study eligibility criteria did not depend on symptoms, people who believed that they had had covid-19 because of previous symptoms may have been more likely to volunteer. This could mean that our estimate of test sensitivity is upwardly biased.

Comparison with other studies
Most previous evaluations of LFIAs have used a "two gate" design, in which tests are performed on two groups of samples: PCR positives and pre-pandemic sera.[2][4][20-22] This study design remains an important tool in evaluating accuracy, owing to the lack of a gold standard test, as discussed above. However, we have shown that using only PCR confirmed cases can overestimate test sensitivity. Therefore, results from these previous studies may be optimistic and should be interpreted with caution. Notably, the PCR confirmed cases in our study were key workers, few of whom were

our study, discordant interpretation by highly trained laboratory technicians accounted for almost 4% of tests.

Strengths and weaknesses of study
Key strengths of this study include precise estimation of test specificity owing to study of almost 2000 pre-pandemic samples, and evaluation in a cohort of almost 2900 key workers consisting of people both with known and with unknown previous infection status, enabling study of potential spectrum bias. Our study also has potential limitations. Firstly, in common with all evaluations of SARS-CoV-2 antibody tests, a "gold

Table 4 | Positive predictive value and negative predictive value for alternative values of prevalence of previous infection. Values are percentages (95% credible intervals)

	Prevalence of previous infection				
	5%	10%	15%	20%	25%
Positive predictive value					
Based on sensitivity estimated from PCR confirmed cases*	69.8 (63.1 to 75.7)	83.0 (78.3 to 86.8)	88.6 (85.2 to 91.3)	91.6 (89.1 to 93.7)	93.6 (91.6 to 95.2)
Based on sensitivity estimated against Roche Elecsys reference standard in people with unknown previous infection status†	67.9 (61.1 to 74.1)	81.7 (76.8 to 85.8)	87.6 (84.0 to 90.6)	91.0 (88.2 to 93.1)	93.1 (90.9 to 94.8)
Negative predictive value					
Based on sensitivity estimated from PCR confirmed cases*	99.6 (99.4 to 99.7)	99.2 (98.7 to 99.4)	98.7 (98.0 to 99.1)	98.1 (97.2 to 98.8)	97.5 (96.3 to 98.4)
Based on sensitivity estimated against Roche Elecsys reference standard in people with unknown previous infection status†	99.2 (99.0 to 99.4)	98.3 (97.8 to 98.7)	97.3 (96.6 to 97.9)	96.3 (95.3 to 97.1)	95.1 (93.8 to 96.1)

Two sets of estimates are presented, using alternative estimates of test sensitivity. For both versions, test specificity of 97.9% (97.2% to 98.4%) is assumed (based on analysis of pre-pandemic samples: table 1).
*Sensitivity 92.5% (88.8% to 95.1%), based on analysis of polymerase chain reaction (PCR) confirmed cases (table 2).
†Sensitivity 84.7% (80.6% to 88.1%), based on comparison with immunoassay reference standard in people with unknown true previous infection status (table 3).

admitted to hospital and a small minority of whom were asymptomatic. The extent of spectrum bias may be greater if sensitivity is estimated using only PCR positive hospital patients.

To our knowledge, only one previous evaluation of the AbC-19 test has been described in a medRxiv pre-print article that is yet to be peer reviewed.[16] That previous study reported estimates of sensitivity and specificity of 97.7% (95.7% to 99.3%) and 100% (100% to 100%), respectively (notably the latter confidence interval, around an observed proportion of 350/350, is reported more appropriately as 98.95% to 100% on the manufacturer's website[23]). Our estimates of both sensitivity and specificity are, therefore, lower than those from the earlier study. In a low prevalence setting, the difference in specificity would be especially important: for example, if 1 million people were tested, of whom 10% had been previously infected with SARS-CoV-2, the previous study predicts no false positives using the AbC-19 test, whereas our study predicts 18 900 false positives. The difference in results was probably driven by sample selection factors: Robertson et al estimated specificity by using only samples that were negative on three immunoassays.[16] This means, for example, that any pre-pandemic samples that falsely tested positive using any of three immunoassays would be excluded. Similarly, sensitivity was estimated from only samples that were positive on EUROIMMUN and at least one other immunoassay. Exclusion of samples with unclear "true" antibody status can be expected to lead to overestimation of both test sensitivity and specificity.

Meaning of study: possible explanations and implications for clinicians and policy makers

For a population with 10% prevalence of previous infection, we estimated a positive predictive value of 81.7% for the AbC-19 test. This means that about one in five positive AbC-19 tests would be a false positive. If the AbC-19 test were to be used for mass population screening in a relatively low prevalence setting, we would anticipate a large number of false positive results (for example, 18 900 for every 1 million tests carried out). Consequently, AbC-19 test positives

would need to be checked with a confirmatory assay, a practice followed in HIV and syphilis serodiagnosis, if such an assay for SARS-CoV-2 could be shown to be sufficiently specific for this purpose. The false positive rate may be higher still if accuracy of the test is lower when deployed in the field than observed in our laboratory study. We also emphasise that our study does not answer the question of whether previous infection, or presence of antibodies, implies immunity, nor the duration of antibody or immune response. The alternative use of LFIAs for population surveillance requires statistical adjustments for imperfect sensitivity and specificity.[3][4][24] The precise adjustment, however, would depend on whether the objective was to estimate the proportion previously infected or simply the proportion with antibody. Considering the decline in antibody over time, the adjustment needed to estimate the proportion previously infected would be difficult to calculate. Finally, if people were to use and read the test themselves, rather than the test being read by trained laboratory technicians, or if a digital reader were developed, the accuracy of the test as read using these alternative approaches would need to be evaluated.

Conclusions

In summary, we found evidence that the sensitivity of the AbC-19 test is lower among populations with unknown previous infection status than among PCR confirmed cases of SARS-CoV-2. This highlights the scope for overestimation of SARS-CoV-2 antibody test sensitivity in other studies, in which sensitivity has been estimated from only PCR confirmed cases. Assuming a prevalence of 10% for previous SARS-CoV-2 infection, and that the performance observed in this laboratory study can be reproduced in field use, our findings suggest that around one in five key workers testing positive with the AbC-19 assay would be false positives.

1 Long Q-X, Liu B-Z, Deng H-J, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020;26:845-8. doi:10.1038/s41591-020-0897-1

2 Deeks JJ, Dinnes J, Takwoingi Y, et al, Cochrane COVID-19 Diagnostic Test Accuracy Group. Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane Database Syst Rev 2020;6:CD013652.

3 Atchison C, Pristerà P, Cooper E, et al. Usability and acceptability of home-based self-testing for SARS-CoV-2 antibodies for population surveillance. Clin Infect Dis 2020;ciaa1178.

4 Flower B, Brown JC, Simmons B, et al. Clinical and laboratory evaluation of SARS-CoV-2 lateral flow assays for use in a national COVID-19 seroprevalence survey. Thorax 2020;thoraxjnl-2020-215732. doi:10.1136/thoraxjnl-2020-215732

5 Voo TC, Clapham H, Tam CC. Ethical Implementation of Immunity Passports During the COVID-19 Pandemic. J Infect Dis 2020;222:715-8. doi:10.1093/infdis/jiaa352

6 Watson J, Richter A, Deeks J. Testing for SARS-CoV-2 antibodies. BMJ 2020;370:m3325. doi:10.1136/bmj.m3325

7 Medicine & Healthcare products Regulatory Agency. Target Product Profile: antibody tests to help determine if people have immunity to SARS-CoV-2: Version 2. 2020. https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-antibody-tests-to-help-determine-if-people-have-recent-infection-to-sars-cov-2-version-2.

8 Van Elslande J, Houben E, Depypere M, et al. Diagnostic performance of seven rapid IgG/IgM antibody tests and the Euroimmun IgA/IgG ELISA in COVID-19 patients. Clin Microbiol Infect 2020;26:1082-7. doi:10.1016/j.cmi.2020.05.023

9 US Food and Drug Administration. FAQs on Testing for SARS-CoV-2. 2020. https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/faqs-testing-sars-cov-2.

10 Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem 2005;51:1335-41. doi:10.1373/clinchem.2005.048595

11 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061-6. doi:10.1001/jama.282.11.1061

12 Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926-30. doi:10.1056/NEJM197810262991705

13 Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. Nat Microbiol 2020. doi:10.1038/s41564-020-00813-8

14 Ripperger TJ, Uhrlaub JL, Watanabe M, et al. Detection, prevalence, and duration of humoral responses to SARS-CoV-2 under conditions

of limited population exposure.2020. https://www.medrxiv.org/content/10.1101/2020.08.14.20174490v1 .

15 Mulchandani R, Taylor-Phillips S, Jones H, et al. Self assessment overestimates historical COVID-19 disease relative to sensitive serological assays: cross sectional study in UK key workers.2020. https://www.medrxiv.org/content/10.1101/2020.08.19.20178186v1. doi:10.1101/2020.08.19.20178186

16 Robertson LJ, Moore JS, Blighe K, et al. SARS-CoV-2 antibody testing in a UK population: detectable IgG for up to 20 weeks post infection.2020. https://www.medrxiv.org/content/10.1101/2020.09.29.20201509v1.

17 Public Health England, Nuffield Department of Medicine, University of Oxford, Oxford University Hospitals NHS Foundation Trust. *Evaluation of sensitivity and specificity of four commercially available SARS-CoV-2 antibody immunoassays.* Public Health England, 2020.

18 World Health Organization. WHO protocol for performance laboratory evaluation of HCV serology assays. 2017. https://www.who.int/diagnostics_laboratory/evaluations/alter/171219_protocol_pqdx_040_v6_hcv.pdf?ua=1.

19 Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid Decay of Anti-SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *N Engl J Med* 2020;383:1085-7. doi:10.1056/NEJMc2025179

20 Whitman JD, Hiatt J, Mowery CT, et al. Test performance evaluation of SARS-CoV-2 serological assays.2020. https://www.medrxiv.org/content/10.1101/2020.04.25.20074856v2. doi:10.1101/2020.04.25.20074856

21 Adams E, 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. doi:10.12688/wellcomeopenres.15927.1

22 Lassaunière R, Frische A, Harboe ZB, et al. Evaluation of nine commercial SARS-CoV-2 immunoassays.2020. https://www.medrxiv.org/content/10.1101/2020.04.09.20056325v1. doi:10.1101/2020.04.09.20056325

23 Abingdon Health. Preliminary report of an independent evaluation of AbC-19. 2020. https://www.abingdonhealth.com/preliminary-report-independent-evaluation-abc19/.

24 Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71-6. doi:10.1093/oxfordjournals.aje.a112510

Web appendix: Supplementary materials