Contents lists available at ScienceDirect

EBioMedicine

Research paper

Accuracy of four lateral flow immunoassays for anti SARS-CoV-2 antibodies: a head-to-head comparative study

Hayley E Jones[a,b,]*, Ranya Mulchandani[c], Sian Taylor-Phillips[d], A E Ades[a], Justin Shute[c],
Keith R Perry[c], Nastassya L Chandra[c], Tim Brooks[c], Andre Charlett[c], Matthew Hickman[a,b],
Isabel Oliver[b,c], Stephen Kaptoge[e], John Danesh[e], Emanuele Di Angelantonio[e], David Wyllie[c],
COMPARE study investigators, EDSAB-HOME investigators

[a] Population Health Sciences, Bristol Medical School, University of Bristol, UK
[b] NIHR Health Protection Research Unit (HPRU) in Behavioural Science and Evaluation, University of Bristol, UK
[c] Public Health England, UK
[d] Warwick Medical School, University of Warwick, UK
[e] Department of Public Health and Primary Care, University of Cambridge, UK

ARTICLE INFO

ABSTRACT

Background: SARS-CoV-2 antibody tests are used for population surveillance and might have a future role in individual risk assessment. Lateral flow immunoassays (LFIAs) can deliver results rapidly and at scale, but have widely varying accuracy.

Methods: In a laboratory setting, we performed head-to-head comparisons of four LFIAs: the Rapid Test Consortium's AbC-19[TM] Rapid Test, OrientGene COVID IgG/IgM Rapid Test Cassette, SureScreen COVID-19 Rapid Test Cassette, and Biomerica COVID-19 IgG/IgM Rapid Test. We analysed blood samples from 2,847 key workers and 1,995 pre-pandemic blood donors with all four devices.

Findings: We observed a clear trade-off between sensitivity and specificity: the IgG band of the SureScreen device and the AbC-19[TM] device had higher specificities but OrientGene and Biomerica higher sensitivities. Based on analysis of pre-pandemic samples, SureScreen IgG band had the highest specificity (98.9%, 95% confidence interval 98.3 to 99.3%), which translated to the highest positive predictive value across any pre-test probability: for example, 95.1% (95% uncertainty interval 92.6, 96.8%) at 20% pre-test probability. All four devices showed higher sensitivity at higher antibody concentrations ("spectrum effects"), but the extent of this varied by device.

Interpretation: The estimates of sensitivity and specificity can be used to adjust for test error rates when using these devices to estimate the prevalence of antibody. If tests were used to determine whether an individual has SARS-CoV-2 antibodies, in an example scenario in which 20% of individuals have antibodies we estimate around 5% of positive results on the most specific device would be false positives.

Funding: Public Health England.

Introduction

Tests for SARS-CoV-2 antibodies are used for population serosurveillance [1,2] and could in future be used for post-vaccination seroepidemiology. Given evidence that antibodies are associated with reduced risk of COVID-19 disease [3-7], antibody tests might also have a role in individual risk assessment [8], pending improved understanding of the mechanisms and longevity of immunity. Both uses require understanding of test sensitivity and specificity: these can be used to adjust seroprevalence estimates for test errors [9], while any test used for individual risk assessment would need to be shown to be sufficiently accurate, in particular, highly specific [10,11].

A number of laboratory-based immunoassays and lateral flow immunoassays (LFIAs) are now available, which detect IgG and/or IgM responses to the spike or nucleoprotein antigens [12-14]. Following infection with SARS-CoV-2, most individuals generate antibodies against both of these antigens [15]. Existing efficacious recombinant vaccines contain the spike antigen [16], therefore

* Corresponding author.
 E-mail address: hayley.jones@bristol.ac.uk (H.E. Jones).

Research in Context

Evidence before this study

We searched for evidence on the accuracy of the four devices compared in this study: OrientGene COVID IgG/IgM Rapid Test Cassette, SureScreen COVID-19 Rapid Test Cassette, Biomerica COVID-19 IgG/IgM Rapid Test and the UK Rapid Test Consortium's AbC-19[TM] Rapid Test. We searched Ovid MEDLINE (In-Process & Other Non-Indexed Citations and Daily), PubMed, MedRxiv/BioRxiv and Google Scholar from January 2020 to 16th January 2021. Search terms included device names AND ((SARS-CoV-2) OR (covid)). Of 303 records assessed, data were extracted from 24 studies: 18 reporting on the accuracy of the OrientGene device, 7 SureScreen, 2 AbC-19[TM] and 1 Biomerica. Only three studies compared the accuracy of two or more of the four devices. With the exception of our previous report on the accuracy of the AbC-19[TM] device, which the current manuscript builds upon, sample size ranged from 7 to 684. For details, see Supplementary Materials (Figure S1, Tables S1, S2).

The largest study compared OrientGene, SureScreen and Biomerica. SureScreen was estimated to have the highest specificity (99.8%, 95% CI 98.9 to 100%) and OrientGene the highest sensitivity (92.6%), but with uncertainty about the latter result due to small sample sizes. The other two comparative studies were small (n = 65, n = 67) and therefore provide very uncertain results.

We previously observed spectrum effects for the AbC-19[TM] device, such that sensitivity is upwardly biased if estimated only from PCR-confirmed cases. The vast majority of previous studies estimated sensitivity in this way.

Added value of this study

We performed a large scale (n = 4,842), head-to-head laboratory-based evaluation and comparison of four lateral flow devices, which were selected for evaluation by the UK Department of Health and Social Care's New Tests Advisory Group, on the basis of a survey of test and performance data available. We evaluated the accuracy of diagnosis based on both IgG and IgM bands, and the IgG band alone. We found a clear trade-off between sensitivity and specificity across devices, with the SureScreen and AbC-19[TM] devices being more specific and OrientGene and Biomerica more sensitive. Based on analysis of 1,995 pre-pandemic blood samples, we are 99% confident that SureScreen (IgG band reading) has the highest specificity of the four devices (98.9%, 95% CI 98.3, 99.3%).

By including individuals without PCR confirmation, and exploring the relationship between laboratory immunoassay antibody index and LFIA positivity, we were able to explore spectrum effects. We found evidence that all four devices have reduced sensitivity at lower antibody indices. However, the extent of this varies by device and appears to be less for other devices than for AbC-19.

Our estimates of sensitivity and specificity are likely to be higher than would be observed in real use of these devices, as they were based on majority readings of three trained laboratory personnel.

Implications of all the available evidence

When used in epidemiological studies of antibody prevalence, the estimates of sensitivity and specificity provided in this study can be used to adjust for test errors. Increased precision in error rates will translate to increased precision in seroprevalence estimates. If lateral flow devices were used for individual

risk assessment, devices with maximum specificity would be preferable. However, if, for example, 20% of the tested population had antibodies, we estimate that around 1 in 20 positive results on the most specific device would be incorrect.

vaccinated individuals generate only a response to this. LFIAs are small devices which produce results rapidly, without the need for a laboratory, and therefore have the potential to be employed at scale.

A Cochrane review found 38 studies evaluating LFIAs for SARS-CoV-2 antibodies already by late April 2020. However, results from most studies were judged to be at high risk of bias, and very few studies directly compared multiple devices [14]. Where direct comparisons have been performed, they have shown that accuracy of LFIAs varies widely across devices [13,17-20]. A key limitation of most studies is that sensitivity has been estimated only from individuals who previously had a positive PCR test. In a recent evaluation of one LFIA, the UK Rapid Test Consortium's "AbC-19[TM] Rapid Test" [21] (AbC-19 hereafter), we found evidence that this can over-estimate sensitivity [21]. We attributed this to PCR-confirmed cases tending to be more severe – particularly early in the pandemic, when access to testing was very limited. Since more severe disease is associated with increased antibody concentrations [22-24], which may be easier to detect, estimates of test sensitivity based on previously PCR-confirmed cases only are susceptible to "spectrum bias" [25,26].

In this paper, we present a head-to-head comparison of the accuracy of AbC-19 and three other LFIAs, based on a large (n = 4,842) number of blood samples. The three additional devices were OrientGene "COVID IgG/IgM Rapid Test Cassette", SureScreen "COVID-19 Rapid Test Cassette", and Biomerica "COVID-19 IgG/IgM Rapid Test", hereafter referred to as OrientGene, SureScreen and Biomerica for brevity.

Methods

We analysed blood samples from 2,847 key workers participating in the EDSAB-HOME study and 1,995 pre-pandemic blood donors from the COMPARE study [27], in a laboratory setting. All samples were from distinct individuals. We evaluated each device using two approaches. First (Approach 1), we compared LFIA results with the known previous infection status of pre-pandemic blood donors ("known negatives") and the 268 EDSAB-HOME participants who reported previous PCR positivity ("known positives"). Second (Approach 2), we compared LFIA results with results on two sensitive laboratory immunoassays in EDSAB-HOME participants. Both approaches were pre-specified in our protocol (available at http://www.isrctn.com/ISRCTN56609224).

We have previously reported accuracy of the AbC-19 device based on the same sample set and overall approaches: these results are reproduced here for comparative purposes [21]. Following this previous work, and in particular due to the spectrum effects observed [21] we anticipated the estimates of sensitivity based on comparison with a laboratory immunoassay (Approach 2) but estimates of specificity based on pre-pandemic sera (Approach 1) to be the least susceptible to bias.

Lateral flow immunoassays

Devices (Table S3) were selected by the UK Department of Health and Social Care's New Tests Advisory Group, on the basis of a survey of test and performance data available. AbC-19, OrientGene and SureScreen devices contain SARS-CoV-2 Spike protein, or domains from it, while Biomerica contains Nucleoprotein. All four devices give qualitative positive or negative results. AbC-19 detects IgG only, while the

other three devices contain separate bands representing detection of IgG and IgM. We report results for these by two different scoring strategies: (i) "one band", in which we considered a result to be positive only if the IgG band was positive; and (ii) "two band", in which we considered results to be positive if either band was positive. In statistical analysis, these two readings were treated as separate "tests", such that our comparison was of seven tests in total. By definition, the "two band" reading of each device has sensitivity greater than or equal to, but specificity less than or equal to, the "one band" reading.

Study participants

The two sets of study participants have been described in full previously [21]. A flow diagram is provided (Fig. 1).

EDSAB-HOME (ISRCTN56609224) was a prospective study designed to assess the accuracy of LFIAs in key workers in England [24]. The research protocol is available at http://www.isrctn.com/ISRCTN56609224. Participants were convenience samples, recruited through their workplaces in three recruitment streams. Individuals in Streams A (fire and police officers: n = 1,147) and B (healthcare
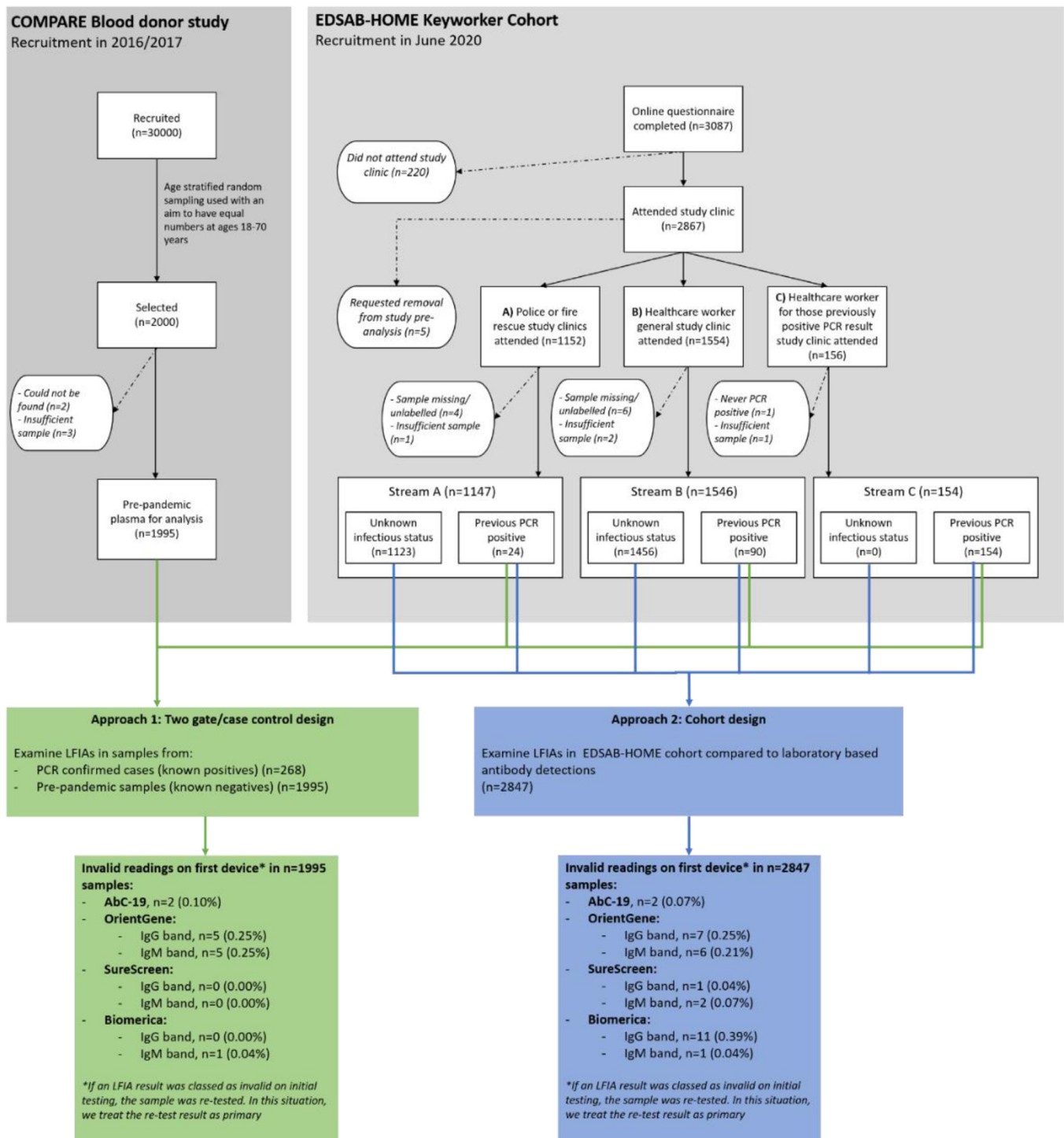

Fig. 1. Study flow diagram

workers: n = 1,546) were recruited without regard to previous SARS-CoV-2 infection status. Stream C (n = 154) consisted of additional healthcare workers who were recruited based on self-reported previous PCR positivity. Symptom history was not part of the eligibility criteria. During June 2020, all participants (n = 2,847) completed an online questionnaire and had a venous blood sample taken at a study clinic.

Sample size considerations, eligibility criteria, the recruitment process and demographic characteristics (Table S4) are described in the Supplementary Materials. Detailed information on participants from the study questionnaire (including symptom history, testing history and household exposure) is provided elsewhere [24]. In addition to the Stream C individuals, some Stream A/B individuals (n = 114) also self-reported previous PCR positivity. All self-reported PCR results were later validated by comparison with national laboratory records. We refer to the total (n = 268) individuals with a previous PCR positive result as "known positives" and to the remaining n = 2,579 EDSAB-HOME participants as "individuals with unknown previous infection status" at clinic visit. Twelve of 268 known positives reported having experienced no symptoms. Of the known positives reporting symptoms, the median (interquartile range) number of days between symptom onset and study clinic was 63 (52 to 75).

COMPARE (ISRCTN90871183) was a 2016-2017 blood donor cohort study in England [27]. We performed stratified random sampling by age, sex and region to select 2,000 participants, of whom 1,995 had samples available for analysis. We refer to these samples as "known negatives".

Laboratory protocol

All tests were performed by experienced laboratory staff at PHE Colindale, London. All EDSAB-HOME samples were first tested with two laboratory immunoassays: Roche Elecsys®, which measures total (including IgG and IgM) antibodies against the Nucleoprotein, and EuroImmun Anti-SARS-CoV-2 ELISA assay, which measures IgG antibodies against the S protein S1 domain. Any immunoassay failing for technical reasons was repeated.

Lateral flow devices were stored in a temperature controlled room (thresholds 16-30°C, actuals from continuous monitoring system 19-20°C). Laboratory staff reading the LFIAs received on-site training from Abingdon Health (the manufacturers of the AbC-19 device), SureScreen and Biomerica. Laboratory staff discussed the evaluation, including use of the LFIA device, with OrientGene on a call. The manufacturers' instructions for use were followed, with plasma being pipetted into the devices, followed by the chase buffer supplied with the kits.

The first 350 COMPARE samples were interspersed randomly among EDSAB-HOME samples, with the remaining 1,650 COMPARE samples being analysed later. Each device was independently read by three members of staff. Readers were blind to demographic or clinical information on participants and to results on any previous assays. Readers scored LFIA test bands using the WHO scoring system for subjectively read assays: 0 ("negative"), 1 ("very weak but definitely reactive"), 2 ("medium to strong reactivity") or 7 ("invalid") [28]. As this scoring system does not clearly state how to categorise "weak" bands, our readers used a score of 1 for what they considered to be either "weak" or "very weak". The majority score was taken as the consensus reading. For assessment of test sensitivity and specificity, scores of 1 and 2 were grouped as "positive". If any band of a device was assigned a consensus score of 7 (invalid), the sample was re-tested and the re-test results taken as primary. We report numbers and proportions of invalid bands, and the total number and proportion of all devices with at least one invalid band.

We re-tested samples when LFIAs made apparent errors, on the following basis: for all four devices, we re-tested EDSAB-HOME samples if the result differed from the immunoassay composite reference standard of "positive on either Roche Elecsys® or EuroImmun, versus negative on both". For three of the four devices, we also re-tested all COMPARE samples that incorrectly tested positive. For Biomerica, due to a lack of devices and a high observed false positive rate, only false positives in the first batch of 350 samples were re-tested. Any re-test results are reported as secondary.

Two checks were made before pipetting to ensure that samples were in the correct position. Once results were available, an initial check was made to ensure that no obvious mistakes had been made by readers on the laboratory scoring sheet (i.e. that scoring was for the correct sample). Data were then manually entered into a spreadsheet for each test run and every result checked against the primary data (laboratory scoring sheet). Results were transferred from each test run spreadsheet to the main results sheet, and checked for correct alignment.

Estimation of sensitivity, specificity, positive and negative predictive values

Approach 1

We estimated LFIA accuracy through comparison of results with the known previous SARS-CoV-2 infection status of individuals. Specificity was estimated from all 1,995 "known negative" samples. The association of false positivity with age, sex and ethnicity was also explored. We estimated sensitivity from the 268 "known positive" EDSAB-HOME samples. Numbers of false negatives are also reported by time since symptom onset and separately for asymptomatic individuals.

Approach 2

We estimated LFIA accuracy through comparison with results on the Roche Elecsys® laboratory immunoassay in EDSAB-HOME samples. This was selected as the primary laboratory reference standard on the basis that it was the assay available to us that had the highest published accuracy for detection of recent SARS-CoV-2 infection at the time of sample collection, as per our protocol. We used the manufacturer recommended positivity threshold of 1.0. At this threshold, the assay has been estimated to have sensitivity of 97.2% (95% CI 95.4, 98.4%) and specificity of 99.8% (99.3, 100%) to previous infection [12]. As sensitivity analyses, we also report accuracy estimates based on comparison with EuroImmun and a composite reference standard of "positive on either laboratory assay versus negative on both". We treated EuroImmun results as positive if they were greater than or equal to the manufacturer "borderline" threshold of 0.8 [24].

In Approach 2, estimates of sensitivity were calculated separately for known positives and individuals with unknown previous infection status, to assess for potential spectrum bias [21]. Specificity was estimated from reference standard negative individuals among the "unknown previous infection status" population. As EDSAB-HOME Streams A and B comprise a "one gate" population [21,29], we also report Approach 2 results from all EDSAB-HOME streams A and B participants combined, regardless of previous PCR positivity.

Positive and negative predictive values: We estimated the positive and negative predictive value (PPV and NPV) for example scenarios of a 10%, 20% and 30% pre-test probability. To calculate these, we used estimates of specificity based on pre-pandemic sera (Approach 1) and sensitivity based on comparison with Roche Elecsys® in individuals with unknown previous infection status (Approach 2). As noted above, we anticipated that these estimates of the respective parameters would be the least susceptible to bias.

Statistical analysis

Statistical analysis was performed in R4.0.3 and Stata 15. Sensitivity and specificity were estimated by observed proportions based on each reference standard, with 95% CIs computed using Wilson's

method. Logistic regressions with age, sex and ethnicity as covariates were used to explore potential associations with false positivity. To further explore potential associations with age, we also fitted fractional polynomials and plotted the best fitting functional form for each test.

In comparing the sensitivity and specificity of the seven "tests", we used generalised estimating equations (GEE) to account for conditional correlations among results [30]. For example, in Approach 1 we fitted separate GEE regressions, with test as a covariate, to the "known positives" and to the "known negatives". We used independence working covariance matrices [30].

We obtained 95% uncertainty intervals (UIs) around PPVs, NPVs, differences in sensitivity and differences in specificity using Monte Carlo simulation. This is a commonly used approach for propagating uncertainty in functions of parameters, used frequently for example in decision modeling [31]. We sampled one million iterations from a multivariate normal distribution (using R function "mvnorm") for each set of GEE regression coefficients, using the parameter estimates and robust variance-covariance matrix. We calculated each function of parameters of interest (e.g. the PPV at 20% prevalence) at each iteration. We report the median value across iterations as the parameter estimate and 2.5th and 97.5th percentiles as 95% uncertainty intervals. We also present ranks (from 1 to 7) for each set of sensitivity, specificity and PPV estimates. These were similarly computed at each iteration of the Monte Carlo simulations, and summarised by medians, 2.5th and 95.5th percentiles across simulations [32]. We further report the proportion of simulations for which each test was ranked first, i.e. the probability the test is the "best" with regard to each measure.

Assessment of spectrum effects in test sensitivity

Within the Approach 2 analysis of Roche Elecsys® positives, we report the absolute difference between sensitivity estimated from PCR-confirmed cases and sensitivity estimated from individuals with unknown previous infection status, with 95% UI.

Among individuals who were positive on Roche Elecsys®, we also examined the relationship between the amount of antibody present and the likelihood of lateral flow test positivity. We categorised anti-Nucleoprotein (Roche Elecsys®) and, separately, anti-S1 (EuroImmun) antibody indices into bins containing similar number of samples and calculated the observed sensitivity (i.e. proportion of positive results) with 95% CI for each LFIA in each bin.

To aid visual assessment of the relationship between antibody index and sensitivity, we also plotted exploratory dose-response curves. The shape of fitted curve was selected using the Akaike information criteria, using the drc package in R [33].

Ethics statement

The EDSAB-HOME study was approved by the NHS Research Ethics Committee (Health Research Authority, IRAS 284980, date 2nd June 2020) and the PHE Research Ethics and Governance Group (REGG, NR0198, date 21st May 2020). All participants gave written informed consent. Ethical approval for use of samples from the COMPARE study is covered by approval from NHS Research Ethics Committee Cambridge East (ref 15/EE/0335, date 18th December 2015).

Role of the funding source

The study was commissioned by the UK Government's Department of Health and Social Care (DHSC), and was funded and implemented by Public Health England, supported by the NIHR Clinical Research Network Portfolio. The DHSC had no role in the study design, data collection, analysis, interpretation of results, writing of the manuscript, or the decision to publish.

Results

Main results

Figs. 2 and 3 show results from Approach 1 and from the Approach 2 analysis of individuals with unknown previous infection status at clinic visit, which are mutually exclusive and exhaustive subsets of the 4,842 samples. Also shown are results from the Approach 2 sensitivity analyses with alternative reference standards. Estimated differences between the sensitivity and specificity of tests, with 95% UIs, are shown in Tables S5 and S6.

Both approaches show a clear trade-off between sensitivity and specificity, with SureScreen 1 band and AbC-19 having higher specificities but lowest sensitivities, while OrientGene and Biomerica have higher sensitivities but lower specificities.

From Approach 1, SureScreen 1 band is estimated to have higher specificity but lower sensitivity than AbC-19, whereas the two tests appeared comparable (although with all point estimates marginally favouring SureScreen) from Approach 2. Resulting from this, we estimate the one band reading of the SureScreen device to have the highest PPV.

Approach 1: comparison with known previous infection status

Approach 1 estimates are shown in Tables 1 (specificity) and 2 (sensitivity).

SureScreen 1 band reading was estimated to have 98.9% specificity (95% CI 98.3, 99.3), with high certainty (99%) of this being the highest. This is 1.0% (95% UI 0.2, 1.8%) higher than the specificity of AbC-19 (Table S6), which was ranked 2nd (95% UI 2nd, 4th). There was no strong evidence of any association between false positivity and age for any device (Table S7) although there was some indication that Biomerica 1 band specificity might decline in older adults (Figure S2). With the exception of an apparent association of false positivity of the AbC-19 device with sex, which we have reported previously [21], there was no indication of specificity varying by sex or ethnicity (Table S8).

SureScreen 1 band was, however, estimated to have the lowest sensitivity when this was estimated from PCR-confirmed cases only (Table 2: 88.8%, 95% CI 84.5, 92.0%), 3.7% (95% UI 0.5, 7.1%) lower than AbC-19 (Table S5).

Approach 2: comparison with laboratory immunoassay results in EDSAB-HOME samples

Among the 268 "known positives", nine were negative on Roche Elecsys®. Removing these from the denominator slightly increased point estimates of sensitivity (Table 3), but had no notable impact on rankings. Among the 2,579 individuals with unknown previous infection status, 354 were positive on Roche Elecsys®. Point estimates of sensitivity were lower for all seven tests in this population than among known positives (see below). In this population, there was evidence that both the OrientGene and Biomerica devices have higher sensitivity than SureScreen or AbC-19 (Table S5). There was no evidence of a difference between the sensitivity of SureScreen and AbC-19 (absolute difference in favour of SureScreen = 0.8%, 95% UI -2.2, 3.9%). Increases in sensitivity in the 2 band versus 1 band reading of OrientGene and Biomerica devices were minimal.

Based on the 2,225 individuals with unknown previous infection status who were negative on Roche Elecsys®, specificity estimates were very similar to those from Approach 1 for SureScreen and Biomerica, but around 1% higher for AbC-19 and OrientGene (Table 3). The ranking of devices was consistent across the two approaches, but the observed difference in specificity between SureScreen and AbC-19 was much reduced in Approach 2 (difference = 0.1%, 95% UI -0.4 to 0.6%, Table S6).

Figs. 2, 3 and Tables S9, S10 show results from sensitivity analyses on the 2,579 samples from individuals with unknown previous
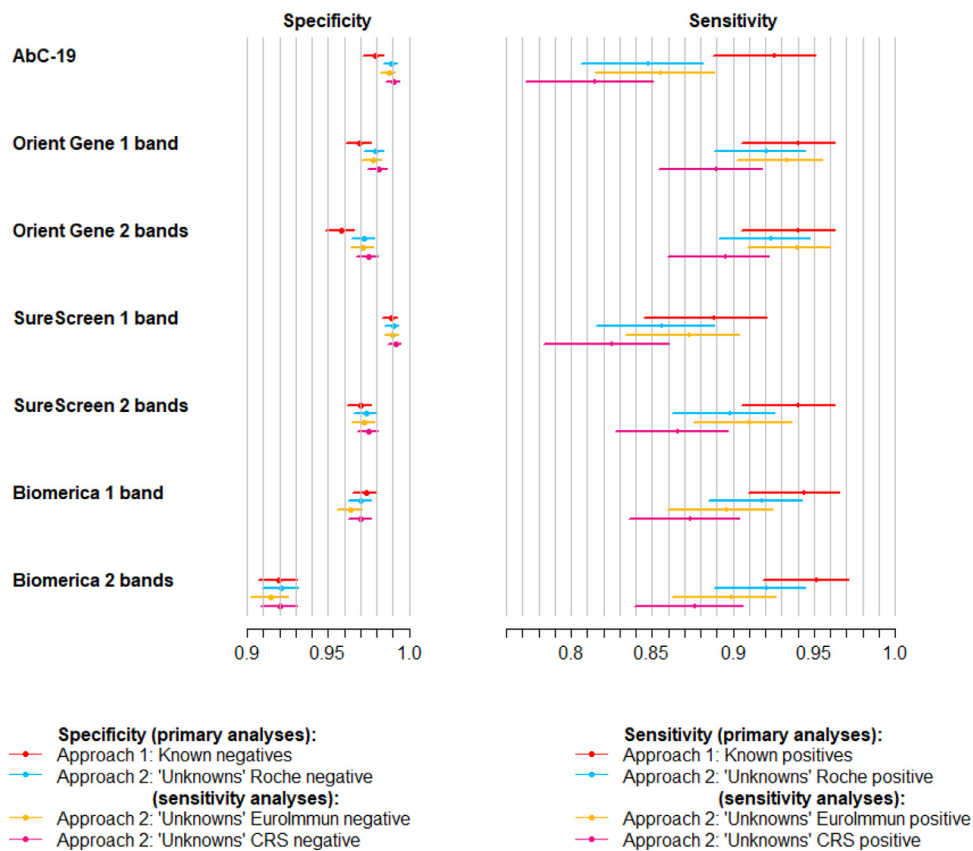
Fig. 2. Sensitivity and specificity of lateral flow devices, with 95% confidence intervals. Four sets of estimates are shown: (i) Approach 1, i.e. specificity from analysis of known negatives and sensitivity from known positives (sample size: n = 1,995 for specificity, n = 268 for sensitivity); (ii) Approach 2 analysis of individuals with unknown previous infection status ("unknowns"), calculated against Roche Elecsys® reference standard (sample size: n = 2,225 for specificity, n = 354 for sensitivity); (iii) Approach 2 sensitivity analysis: analysis of unknowns compared with alternative EuroImmun reference standard (n = 2,233 for specificity, n = 346 for sensitivity); (iv) Approach 2 sensitivity analysis: analysis of unknowns compared with alternative composite reference standard (CRS) of positive on either Roche Elecsys® or EuroImmun versus negative on both (n = 2,207 for specificity, n = 372 for sensitivity).

infection status. When EuroImmun was taken as the reference standard, estimates of specificity were robust, while sensitivity appeared slightly higher for AbC-19, OrientGene and SureScreen, but slightly lower for Biomerica, although with overlapping CIs. All devices were estimated to have slightly lower sensitivity when evaluated against the composite reference standard. OrientGene was ranked highest for sensitivity across all three immunoassay reference standards, but with Biomerica appearing as a close contender when evaluated against Roche Elecsys®.

Table S11 shows sensitivity and specificity estimated from all EDSAB-HOME Streams A and B ("one gate" study), based on comparison with each of the three immunoassay reference standards. Rankings of devices were quite robust to inclusion of PCR-confirmed cases.

Re-test results are shown in Table S12.

Positive and negative predictive values

Based on the sets of estimates that we consider least susceptible to bias (see Methods), we are 99% confident that SureScreen 1 band reading has the highest PPV. This ranking does not depend on pretest probability (Table 4, Figure S4). At a pre-test probability of 20%, we estimate SureScreen 1 band reading to have a PPV of 95.1% (95% UI 92.6, 96.8%), such that we would expect approximately one in twenty positive results to be incorrect.

OrientGene and Biomerica have the highest ranking NPVs. There is very little difference between the NPVs for the one or two band readings of these devices (Table 4).

Spectrum effects

For all seven tests, point estimates of sensitivity were lower among individuals with unknown previous infection status who were positive on Roche Elecsys® than among PCR-confirmed cases, with strong statistical evidence of a difference for all tests except OrientGene (Table 3). The greatest observed difference was for AbC-19.

Fig. 4 shows that all devices were more sensitive at higher antibody concentrations. This effect was most marked in the devices with lower sensitivity, particularly AbC-19. All LFIAs had high sensitivity at the highest anti-S IgG concentrations, but at lower concentrations many lateral flow tests were falsely negative (Figure S3, Tables S13, S14).

Usability issues

Very few bands or devices produced invalid readings (Table S15). Laboratory assessors reported that SureScreen bands were intense, well defined and easy to read, and that OrientGene bands were also easy to read. For Biomerica, some gradients and streaking in band formation were observed, which led to devices taking slightly longer to read. As we have reported previously, AbC-19 bands were often weak visually [21].

All devices showed some variability in reading across three assessors. Concordance was highest for the SureScreen IgG band: there were no discrepancies in the reading of this for 98.7% (98.3, 98.9%) of devices (Table S16). Positive OrientGene, Biomerica and SureScreen IgG bands all tended to be stronger than AbC-19 bands: for example, across the 613 EDSAB-HOME samples that were positive on Roche
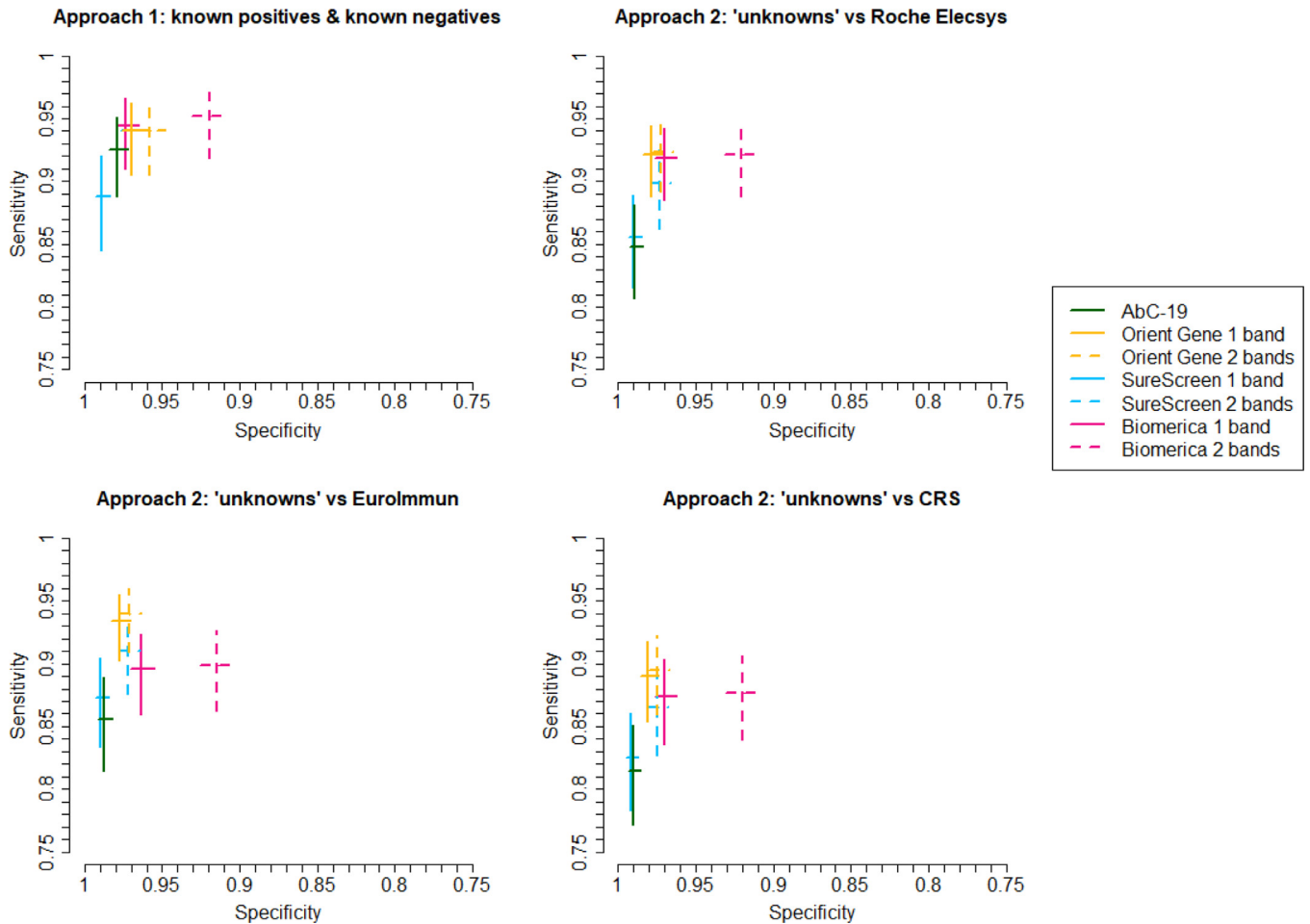
Fig. 3. Sensitivity and specificity of lateral flow devices, with 95% confidence intervals, plotted in Receiver Operator Characteristic space. Four sets of estimates are shown: (i) Approach 1, i.e. specificity from analysis of known negatives and sensitivity from known positives (sample size: n = 1,995 for specificity, n = 268 for sensitivity); (ii) Approach 2 analysis of individuals with unknown previous infection status ("unknowns"), calculated against Roche Elecsys® reference standard (sample size: n = 2,225 for specificity, n = 354 for sensitivity); (iii) Approach 2 sensitivity analysis: analysis of unknowns compared with alternative EuroImmun reference standard (n = 2,233 for specificity, n = 346 for sensitivity); (iv) Approach 2 sensitivity analysis: analysis of unknowns compared with alternative composite reference standard (CRS) of positive on either Roche Elecsys® or EuroImmun versus negative on both (n = 2,207 for specificity, n = 372 for sensitivity). NB SureScreen 2 band overlays Orient Gene 1 band in the first panel.

Elecsys®, 76%, 69% and 79% showed "medium to strong reactivity" respectively, compared with 44% of AbC-19 devices (Table S17). Concordance was lower for reading of IgM than IgG bands. IgM bands, when read as positive, were also often weak.

Discussion

We found evidence that SureScreen (when reading the IgG band only) and AbC-19 have higher specificities than OrientGene and Biomerica, but the latter have higher sensitivities. We can confidently conclude that SureScreen 1 band reading has ~99% specificity, since this estimate was robust across two large discrete samples sets. In contrast, estimates of the specificity of AbC-19 and OrientGene varied slightly across Approaches 1 and 2. As Approach 2 denominators are subject to some misclassification error, we consider the estimates of specificity based on pre-pandemic samples to be most reliable.

The sensitivities of OrientGene and Biomerica appeared comparable based on a reference standard of Roche Elecsys® (anti-N) immunoassay, whereas OrientGene appeared to have higher sensitivity when an alternative (anti-S) reference standard was used. This difference is not surprising since Biomerica also measures anti-N response whereas OrientGene (and the other two devices studied) measures anti-S response. For all four devices, there was some evidence of lower sensitivity to detect lower concentrations of antibody. This spectrum effect appeared strongest for the AbC-19 test and weakest

for OrientGene. Due to spectrum effects, we consider Approach 2 estimates of sensitivity to be the most realistic.

Notably, none of the four devices met the UK Medicines and Health products Regulatory Agency's requirement of sensitivity >98% for the use case of individual level risk assessment [11], even in our least conservative analytical scenario, which we expect to over-estimate sensitivity. On the other hand, the basis for this criterion is unclear, as we would expect high specificity to be the key consideration for this potential use case.

Major strengths of this work include its size and performance of all LFIAs on an identical sample set. This design is optimal for comparing test accuracy [34]. Inclusion of laboratory immunoassay positive cases without PCR confirmation is an additional key strength over most previous studies in this field (see Research in Context panel and Table S2): this allowed assessment and quantification of spectrum effects. Antibody test sensitivity may have been over-estimated by studies that have quantified this from previously PCR-confirmed cases alone, particularly if blood samples were taken at a point when access to PCR testing was very limited.

A limitation of our study is that tests were conducted in a laboratory setting, with the majority reading across three expert readers being taken as the result. For devices with discrepancies between readers, the accuracy of a single reader can be expected to lower [21]. Accuracy may be lower still if devices were read by individuals at home with less or no training, and may differ if device reading

Table 1

Specificity of lateral flow devices: Approach 1 (known negatives). Estimates based on analysis of 1,995 pre-pandemic samples. CI = confidence interval, UI = uncertainty interval based on percentiles from Monte Carlo simulation, TNs = true negatives, FPs = false positives, "Probability best" = the proportion of Monte Carlo simulations in which the test had the highest specificity. Note: these AbC-19™ results have been published previously (21) and are reproduced here for comparative purposes.

	Lateral flow immunoassay						
	AbC-19™	Orient Gene 1 band	Orient Gene 2 bands	SureScreen 1 band	SureScreen 2 bands	Biomerica 1 band	Biomerica 2 bands
TNs	1,953	1,934	1,911	1,973	1,935	1,942	1,835
FPs	42	61	84	22	60	53	160
Specificity	97.9%	96.9%	95.8%	98.9%	97.0%	97.3%	92.0%
(95% CI)	(97.2, 98.4)	(96.1, 97.6)	(94.8, 96.6)	(98.3, 99.3)	(96.1, 97.7)	(96.5, 98.0)	(90.7, 93.1)
Rank	2	4	6	1	4	3	7
(95% UI)	(2, 4)	(3, 5)	(6, 6)	(1, 1)	(3, 5)	(2, 5)	(7, 7)
Probability best	0.01	0.00	0.00	0.99	0.00	0.00	0.00

Table 2

Sensitivity of lateral flow devices: Approach 1 (known positives). Estimates based on analysis of 268 individuals self-reporting previous PCR confirmed infection. CI = confidence interval, UI = uncertainty interval based on percentiles from Monte Carlo simulation, TPs = true positives, FNs = false negatives, "Probability best" = the proportion of Monte Carlo simulations in which the test had the highest sensitivity. Note: these AbC-19™ results have been published previously (21) and are reproduced here for comparative purposes.

	Lateral flow immunoassay						
	AbC-19™	Orient Gene 1 band	Orient Gene 2 bands	SureScreen 1 band	SureScreen 2 bands	Biomerica 1 band	Biomerica 2 bands
TPs	248	252	252	238	252	253	255
FNs	20	16	16	30	16	15	13
Sensitivity (95% CI)	92.5% (88.8, 95.1)	94.0% (90.5, 96.3)	94.0% (90.5, 96.3)	88.8% (84.5, 92.0)	94.0% (90.5, 96.3)	94.4% (91.0, 96.6)	95.1% (91.9, 97.1)
Rank	6	4	4	7	3	3	1
(95% UI)	(2, 6)	(1, 6)	(1, 6)	(7, 7)	(1, 6)	(1, 6)	(1, 5)
Probability best	0.01	0.08	0.08	0.00	0.14	0.04	0.64
False negatives by days since between symptom onset and blood sample:							
Asymptomatic (n=12)	5	3	3	5	4	4	4
8-21 days (n=5)	0	0	0	0	0	0	0
22-35 days (n=20)	1	1	1	2	1	1	1
36-70 days (n=142)	6	4	4	12	5	2	1
≥71 days (n=89)	8	8	8	11	6	8	7

Table 3

Sensitivity and specificity of lateral flow devices: Approach 2. Comparison with Roche Elecsys® immunoassay in EDSAB-HOME samples, stratified by previous PCR positivity. CI = confidence interval, UI = uncertainty interval based on percentiles from Monte Carlo simulation, "Probability best" = the proportion of Monte Carlo simulations in which the test had the highest sensitivity or specificity. Note: the AbC-19™ results have been published previously (21) and are reproduced here for comparative purposes.

	AbC-19™	Orient Gene 1 band	Orient Gene 2 bands	SureScreen 1 band	SureScreen 2 bands	Biomerica 1 band	Biomerica 2 bands
Analysis of 268 PCR-confirmed cases							
Reference standard of Roche Elecsys®: 259 positive							
False negatives	15	12	12	23	10	8	6
Sensitivity (95% CI)	94.2% (90.7, 96.5)	95.4% (92.1, 97.3)	95.4% (92.1, 97.3)	91.1% (87.0, 94.0)	96.1% (93.0, 97.9)	96.9% (94.0, 98.4)	97.7% (95.0, 98.9)
Ranked sensitivity (95% UI)	6 (3, 7)	4 (2, 6)	4 (2, 6)	7 (6, 7)	3 (1, 6)	2 (1, 6)	1 (1, 5)
Probability best sensitivity	0.00	0.01	0.01	0.00	0.14	0.05	0.79
Analysis of 2,579 individuals with unknown previous infection status							
Reference standard of Roche Elecsys®: 354 positive and 2,225 negative							
False negatives	54	28	27	51	36	29	28
Sensitivity (95% CI)	84.7% (80.6, 88.1)	92.1% (88.8, 94.5)	92.4% (89.1, 94.7)	85.6% (81.6, 88.9)	89.8% (86.2, 92.6)	91.8% (88.5, 94.2)	92.1% (88.8, 94.5)
Ranked sensitivity (95% UI)	7 (6, 7)	2 (1, 4)	1 (1, 4)	6 (6, 7)	5 (3, 5)	4 (1, 5)	3 (1, 4)
Probability best sensitivity	0.00	0.07	0.50	0.00	0.00	0.06	0.36
False positives	24	47	62	22	59	66	176
Specificity (95% CI)	98.9% (98.4, 99.3)	97.9% (97.2, 98.4)	97.2% (96.4, 97.8)	99.0% (98.5, 99.3)	97.3% (96.6, 97.9)	97.0% (96.2, 97.7)	92.1% (90.9, 93.1)
Ranked specificity (95% UI)	2 (1, 2)	3 (3, 4)	5 (4, 6)	1 (1, 2)	4 (3, 6)	6 (3, 6)	7 (7, 7)
Probability best specificity	0.37	0.00	0.00	0.63	0.00	0.00	0.00
Absolute difference in sensitivity: known positives vs individuals with unknown previous infection status (95% UI)	9.5% (7.7, 11.3)	3.3% (-0.2, 6.6)	3.0% (-0.5, 6.3)	5.5% (2.2, 8.7)	6.3% (3.0, 9.7)	5.1% (2.3, 8.0)	5.6% (3.0, 8.3)

Table 4

Positive predictive value (PPV) and negative predictive value (NPV), with 95% uncertainty intervals (UIs), for example scenarios of 10%, 20% and 30% pre-test probability. "Probability best" = the proportion of Monte Carlo simulations in which the test had the highest PPV or NPV. Specificity was estimated from 1,995 pre-pandemic samples (Table 1) and sensitivity from 354 Roche Elecsys® positives with unknown previous infection status at clinic visit (Table 3).

	AbC-19™	Orient Gene 1 band	Orient Gene 2 bands	SureScreen 1 band	SureScreen 2 bands	Biomerica 1 band	Biomerica 2 bands
Positive predictive value (PPV)							
PPV at 10% pre-test probability (95% UI)	81.7% (77.1, 85.6)	77.0% (72.6, 80.8)	70.9% (66.7, 74.8)	89.6% (84.7, 93.0)	76.8% (72.1, 81.0)	79.3% (74.2, 83.6)	56.1% (52.1, 59.9)
PPV at 20% pre-test probability (95% UI)	91.0% (88.3, 93.0)	88.3% (85.6, 90.5)	84.6% (81.8, 87.0)	95.1% (92.6, 96.8)	88.2% (85.3, 90.5)	89.6% (86.6, 92.0)	74.2% (71.0, 77.1)
PPV at 30% pre-test probability (95% UI)	94.5% (92.8, 95.8)	92.8% (91.1, 94.2)	90.4% (88.5, 92.0)	97.1% (95.5, 98.1)	92.8% (90.9, 94.3)	93.7% (91.7, 95.2)	83.1% (80.7, 85.2)
Ranked PPV (UI)	2 (2, 4)	4 (2, 5)	6 (6, 6)	1 (1, 1)	4 (2, 5)	3 (2, 5)	7 (7, 7)
Probability best PPV	0.01	0.00	0.00	0.99	0.00	0.00	0.00
Negative predictive value (NPV)							
NPV at 10% pre-test probability (95% UI)	98.3% (97.9, 98.7)	99.1% (98.7, 99.4)	99.1% (98.8, 99.4)	98.4% (98.0, 98.8)	98.8% (98.4, 99.2)	99.1% (98.7, 99.4)	99.1% (98.7, 99.3)
NPV at 20% pre-test probability (95% UI)	96.3% (95.3, 97.0)	98.0% (97.2, 98.6)	98.0% (97.2, 98.6)	96.5% (95.5, 97.3)	97.4% (96.6, 98.1)	97.9% (97.1, 98.6)	97.9% (97.0, 98.5)
NPV at 30% pre-test probability (95% UI)	93.7% (92.2, 95.0)	96.6% (95.3, 97.6)	96.7% (95.4, 97.7)	94.1% (92.6, 95.4)	95.7% (94.3, 96.8)	96.5% (95.1, 97.5)	96.4% (95.0, 97.5)
Ranked NPV (UI)	7 (6, 7)	2 (1, 4)	2 (1, 4)	6 (6, 7)	5 (3, 5)	3 (1, 5)	3 (1, 5)
Probability best NPV	0.00	0.13	0.48	0.00	0.00	0.29	0.10
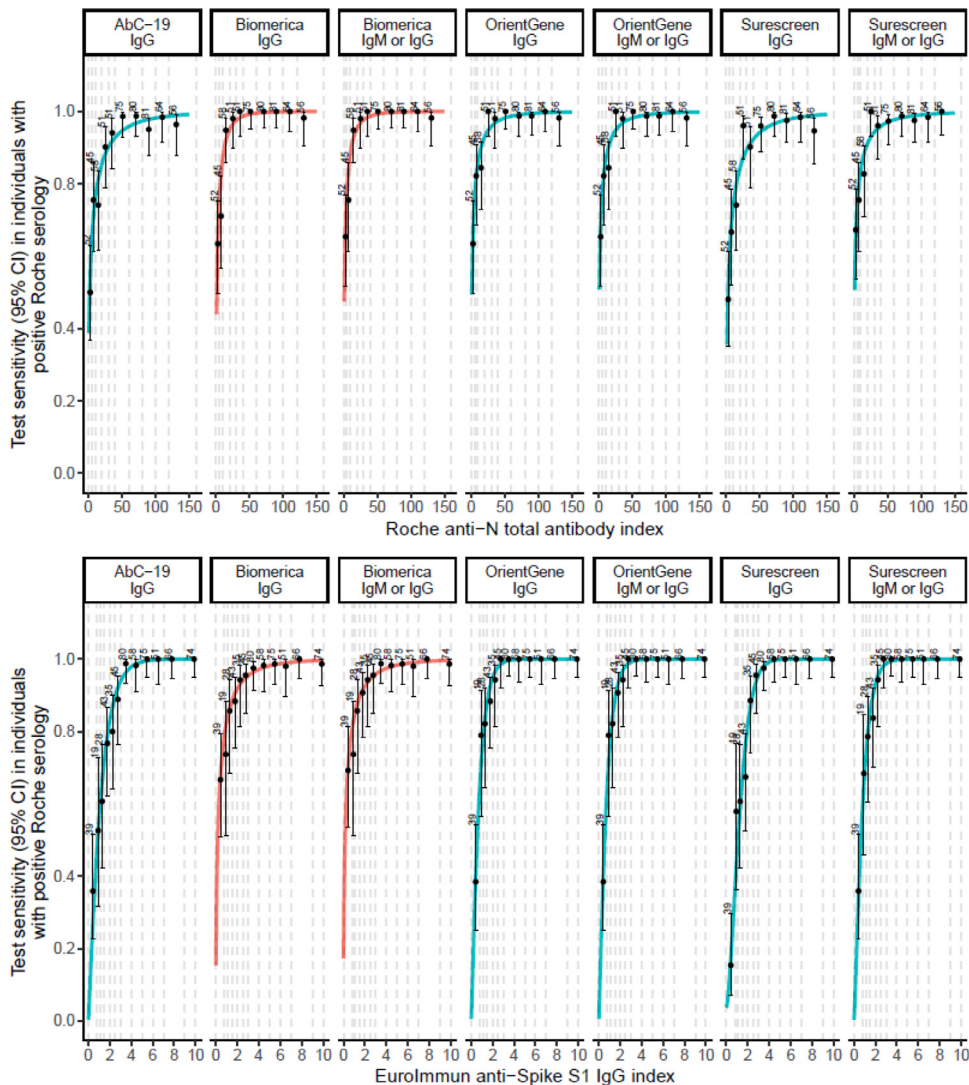
Fig. 4. Sensitivity of lateral flow devices, with 95% confidence intervals, by antibody index (categorised into groups of approximately equal size), among n = 613 EDSAB-HOME participants who were positive on Roche Elecsys®. Top panel: sensitivity by anti-Nucleoprotein antibody (Roche Elecsys®); Bottom Panel: sensitivity by anti-Spike IgG (EuroImmun). Lines show exploratory dose-response curves based on a Weibull function. Red lines: lateral flow test contains nucleoprotein; Blue lines: lateral flow test contains spike proteins.

technologies were used. Variation by reader type seems particularly likely for the devices that our laboratory assessors found more difficult to read, due to weak bands or gradients and streaking in band formation. SureScreen IgG band, followed by OrientGene IgG band, had the highest concordance across readers, who also reported these bands to be easy to read.

An ongoing difficulty in this field is the ambiguity as to whether the true parameters of interest are sensitivity and specificity to previous infection, to presence of particular antibodies, or to "immunity". Our estimates are best interpreted as sensitivity and specificity to "recent" SARS-CoV-2 infection (Approach 1) or presence of an antibody response (Approach 2). These can be expected to correlate very highly since most individuals seroconvert [15] and both the anti-S and anti-N antibody response are highly specific to SARS-CoV-2 [12]. Although we believe that, due to spectrum effects, our estimates of sensitivity based on a reference standard of Roche Elecsys® are more reliable than those based on previous PCR confirmation, we note that this assay may itself make some errors, and that evaluation against this assay may tend to favour LFIAs measuring anti-N responses. We explored this with sensitivity analyses using two alternative reference standards.

Although there is strong evidence that presence of antibody response correlates with reduced risk [3,4], our estimates should not be directly interpreted as sensitivity and specificity to detect "immunity" or to detect "any" previous infection (given declining antibody response over time). Further, our study describes test accuracy following natural infection, not after vaccination. Estimates of sensitivity would require further validation in vaccinated populations if the tests were to be used for post-vaccination monitoring. Notably, antigen choice precludes both Biomerica and Roche Elecsys® from this use case. An additional limitation of our analyses is that we did not quantify the accuracy of tests used in sequence, e.g. check positive results on Test A with a confirmatory Test B [35]. Finally, we estimated device accuracy in key workers in the UK and we note that accuracy may not be generalisable to other populations.

If these devices are used for seroprevalence estimation, our estimates of LFIA accuracy can be used to adjust for test errors [9]. The "one gate" estimates of sensitivity would likely be the most appropriate for this. For the alternative potential use case of individual risk assessment (pending improved understanding of immunity), it would be desirable to use the most specific test or that with the highest PPV, which we estimate to be SureScreen 1 band reading, followed by AbC-19. At a 20% seroprevalence, we estimate that around 1 in 20 SureScreen IgG positive readings would be a false positive. Confirmatory testing, possibly with a second LFIA, would be an option, although requires evaluation.

Contributors

DW, RM, HEJ, STP, AEA, TB, AC, MH and IO planned the study. KRP and JS planned the laboratory based investigation. SK, JD, EDA, and DW planned the specificity investigations. DW, RM, EDSAB-HOME site investigators and COMPARE investigators collected/provided samples. RB, EL, and TB collated samples and performed assays. KRP and JS conducted experiments. HEJ, DW and SK did the statistical analyses. NC performed the rapid review of previous evidence. HEJ and DW wrote the paper, which all authors critically reviewed. Data have been verified by HEJ and DW.

Declaration of Competing Interest

JS and KP report financial activities on behalf of WHO in 2018 and 2019 in evaluation of several other rapid test kits. MH declares unrelated and unrestricted speaker fees and travel expenses in last 3 years from MSD and Gillead. JD has received grants from Merck, Novartis, Pfizer and AstraZeneca and personal fees and non-financial support from Pfizer Population Research Advisory Panel. Outside of this work, RB and EL perform meningococcal contract research on behalf of PHE for GSK, Pfizer and Sanofi Pasteur. All other authors declare no conflicts of interest.

Data sharing

SARS-CoV-2 antibody test result data on the 2,847 EDSAB-HOME study participants have been deposited on Mendeley Data (DOI: 10.17632/gnsf982vrb.1), with a data dictionary. The data set contains individual level results on the four LFIAs and two laboratory immunoassays. Study recruitment stream, previous PCR positivity and (among individuals self-reporting previous PCR positivity) whether symptomatic disease was experienced and days between symptom onset and venous blood sample being taken are also provided. COMPARE study data are provided in aggregate form in the supplementary materials (Table S18). This table provides cross tabulations of results on the four LFIA devices.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103414.

References

[1] Anand S, Montez-Rath M, Han J, Bozeman J, Kerschmann R, Beyer P, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. Lancet 2020.

[2] Poustchi H, Darvishian M, Mohammadi Z, Shayanrad A, Delavari A, Bahadorimonfared A, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. Lancet Infect Dis 2020.

[3] Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, et al. Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. N Engl J Med 2020.

[4] Hanrath AT, Payne BAI, Duncan CJA, Group NHT, van der Loeff IS, Baker KF, et al. Prior SARS-CoV-2 infection is associated with protection against symptomatic reinfection. J Infect 2020.

[5] Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet 2020;396(10249):467–78.

[6] Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet 2021;397(10269):99–111.

[7] Hall V, Foulkes S, Charlett A, Atti A, Monk EJ, Simmons R, et al. Do antibody positive healthcare workers have lower SARS-CoV-2 infection rates than antibody negative healthcare workers? Large multi-centre prospective cohort study (the SIREN study), England: June to November 2020.

[8] Voo TC, Clapham H, Tam CC. Ethical Implementation of Immunity Passports During the COVID-19 Pandemic. J Infect Dis 2020;222(5):715–8.

[9] Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. Am J Epidemiol 1978;107(1):71–6.

[10] Watson J, Richter A, Deeks J. Testing for SARS-CoV-2 antibodies. 2020;370:m3325.

[11] MHRA. Target Product Profile: antibody tests to help determine if people have immunity to SARS-CoV-2, Version 2. London: MHRA; 2020.

[12] Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. Lancet Infect Dis 2020;20(12):1390–400.

[13] Flower B, Brown JC, Simmons B, Moshe M, Frise R, Penn R, et al. Clinical and laboratory evaluation of SARS-CoV-2 lateral flow assays for use in a national COVID-19 seroprevalence survey. Thorax 2020.

[14] Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane Database System Rev 2020;6(6):Cd013652.

[15] Long Q-X, Liu B-Z, Deng H-J, Wu G-C, Deng K, Chen Y-K, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020;26(6):845–8.

[16] World Health Organization. COVID-19 - Landscape of novel coronavirus candidate vaccine development worldwide 2020 [Available from: https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines (Last accessed 09/04/2021).

[17] Van Elslande J, Houben E, Depypere M, Brackenier A, Desmet S, Andre E, 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(8):1082–7.

[18] Adams ER, Ainsworth M, Anand R, Andersson MI, Auckland K, Baillie JK, et al. Antibody testing for COVID-19: a report from the national COVID scientific advisory panel. 2020;5(139):139.

[19] Conklin SE, Martin K, Manabe YC, Schmidt HA, Miller J, Keruly M, et al. Evaluation of serological SARS-CoV-2 lateral flow assays for rapid point of care testing. J Clin Microbiol 2020.

[20] Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, et al. Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. Nat Biotechnol 2020;38(10):1174–83.

[21] Mulchandani R, Jones HE, Taylor-Phillips S, Shute J, Perry K, Jamarani S, et al. 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. BMJ 2020;371:m4262.

[22] Seow J, Graham C, Merrick B, Acors S, Pickering S, Steel KJA, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. Nat Microbiol 2020;5(12):1598–607.

[23] Ripperger TJ, Uhrlaub JL, Watanabe M, Wong R, Castaneda Y, Pizzato HA, et al. Detection, prevalence, and duration of humoral responses to SARS-CoV-2 under conditions of limited population exposure. medRxiv.

[24] Mulchandani R, Taylor-Philips S, Jones HE, Ades A, Borrow R, Linley E, et al. Association between self-reported signs and symptoms and SARS-CoV-2 antibody detection in UK key workers. J Infection 2021.

[25] Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299(17):926–30.

[26] Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282(11):1061–6.

[27] Bell S, Sweeting M, Ramond A, Chung R, Kaptoge S, Walker M, et al. Comparison of four methods to measure haemoglobin concentrations in whole blood donors (COMPARE): A diagnostic accuracy study. Transfus Med 2020.

[28] World Health Organization. In vitro diagnostics and laboratory technology. WHO performance evaluation protocols; 2021. [Available from: https://www.who.int/diagnostics_laboratory/evaluations/alter/protocols/en/ Last accessed 09/04/2021.

[29] Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem 2005;51(8):1335–41.

[30] Pepe MS. The statistical evaluation of medical tests for classification and prediction, xvi. Oxford; New York: Oxford University Press; 2003. p. 302.

[31] Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Med Decis Making 1985;5(2):157–77.

[32] Welton NJ. Evidence synthesis for decision making in healthcare, xii. Chichester, West Sussex: John Wiley & Sons; 2012. p. 282.

[33] Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. PLoS One 2015;10(12):e0146021.

[34] Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006;332(7549):1089–92.

[35] Ripperger TJ, Uhrlaub JL, Watanabe M, Wong R, Castaneda Y, Pizzato HA, et al. Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low-Prevalence Communities and Reveal Durable Humoral Immunity. Immunity 2020;53(5):925-33 e4.