UK Health Security Agency

Evaluation of lateral flow device performance within the National Testing Programme

Date of reporting period: 8 November 2020 to 21 March 2022

Contents

Evaluation team

Project sponsor: Susan Hopkins, UK Health Security Agency (UKHSA)

Principal investigators: Tom Fowler, UKHSA, and Tim Peto, University of Oxford

Co-investigators: Joanna Cole-Hamilton, Rida Saquib, Nick Germanacos, Edward Blandford,

Sarah Tunkel, Mark Stockbridge, Somya Agrawal and Jeanette Hall, all UKHSA

Biostatisticians: Matthias Futschik, Iftekhar Khan, both UKHSA, and David Eyre, University of Oxford

Data scientists: Olutoye Olatunbosun, Jegan Davis-Jones, Rishi Nursimloo and Zarkash Islam, all UKHSA

Virology advisors: Andrew Dodgson and Paul Klapper, both UKHSA and Manchester University NHS Foundation Trust, and Malur Sudhanva, UKHSA and King's College Hospital NHS Foundation Trust

Participant recruitment: Bhupinder Singh and Mahmood Batasi, both UKHSA

Project managers: Fauzia Ashraf, Bibian Mukasa and Nick Sharp, all UKHSA

Abstract

The performance of the various Lateral Flow Devices (LFDs)[1] deployed in different test settings and services within the National Testing Programme (NTP) was evaluated by comparing their performance metrics to the initial baseline performance determined by pre-deployment service evaluations conducted between November 2020 and January 2021.

Methods

Paired LFD and qRT-PCR tests were actively and prospectively sought, for the explicit purpose of ongoing evaluation of LFD performance, from individuals testing in different settings and services. The paired tests were taken at the same time, and the 2 results were matched to each other. Statistical outcomes such as sensitivity were compared with the results from the baseline studies.

Data was organised into 6 analysis sets defined by the type of LFD kit used in a particular test setting (self-testing (SELF), assisted testing (ASSIST)) and service (symptomatic (SYMP) or asymptomatic (ASYMP)). Sensitivity, specificity, probability of a true positive, positive predicted value (PPV), negative predicted value (NPV) and void rates were reported. Statistical methods included logistic regression modelling adjusted for covariates (such as viral concentration, vaccination status, symptom status, variant, and so on), along with 95% confidence intervals. An allowable departure from baseline sensitivity was set at -10% (a non-inferiority margin).

Results

A total of nearly 75,000 paired LFD-PCR samples were provided for this analysis between 8 November 2020 and 21 March 2022. For all 6 analysis sets, there was no statistical evidence that post-deployment sensitivity was inferior to the baseline. Specificity was found to be higher than or similar to baseline in post-deployment. Viral concentration and symptomatic disease were found to be statistically significant predictors of a true positive. All other outcomes showed similar or improved performance compared to the baseline.

Conclusions

This report concludes that the LFD kits used in the National Testing Programme were performing at or above expected levels during deployment. The evaluation was not designed to evaluate why performance during deployment may be different to baseline as a primary

[1] See full list of LFDs in Section 1.1 below.

outcome, and we were unable to draw conclusions as to why performance improved. This may warrant further investigation. The evaluation concludes that LFDs have provided sufficient diagnostic performance for use as part of a public health intervention for the SARS-CoV-2 pandemic.

Commonly used abbreviations

Abbreviation	Description
CI	Confidence Interval
Ct	Cycle Threshold
EUA	Exceptional Use Authorisation
LFD	Lateral Flow Device
LTS	Local Testing Site
MHRA	Medicines and Healthcare Products Regulatory Agency
NPV	Negative Predictive Value
NTP	National Testing Programme
OE	Ongoing Evaluation
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PPV	Positive Predictive Value
qRT-PCR	Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction
RTS	Regional Testing Site
TCT	Targeted Community Testing
TP	True Positive
UKHSA	UK Health Security Agency
VC	Viral Concentration
VoC	Variant of Concern

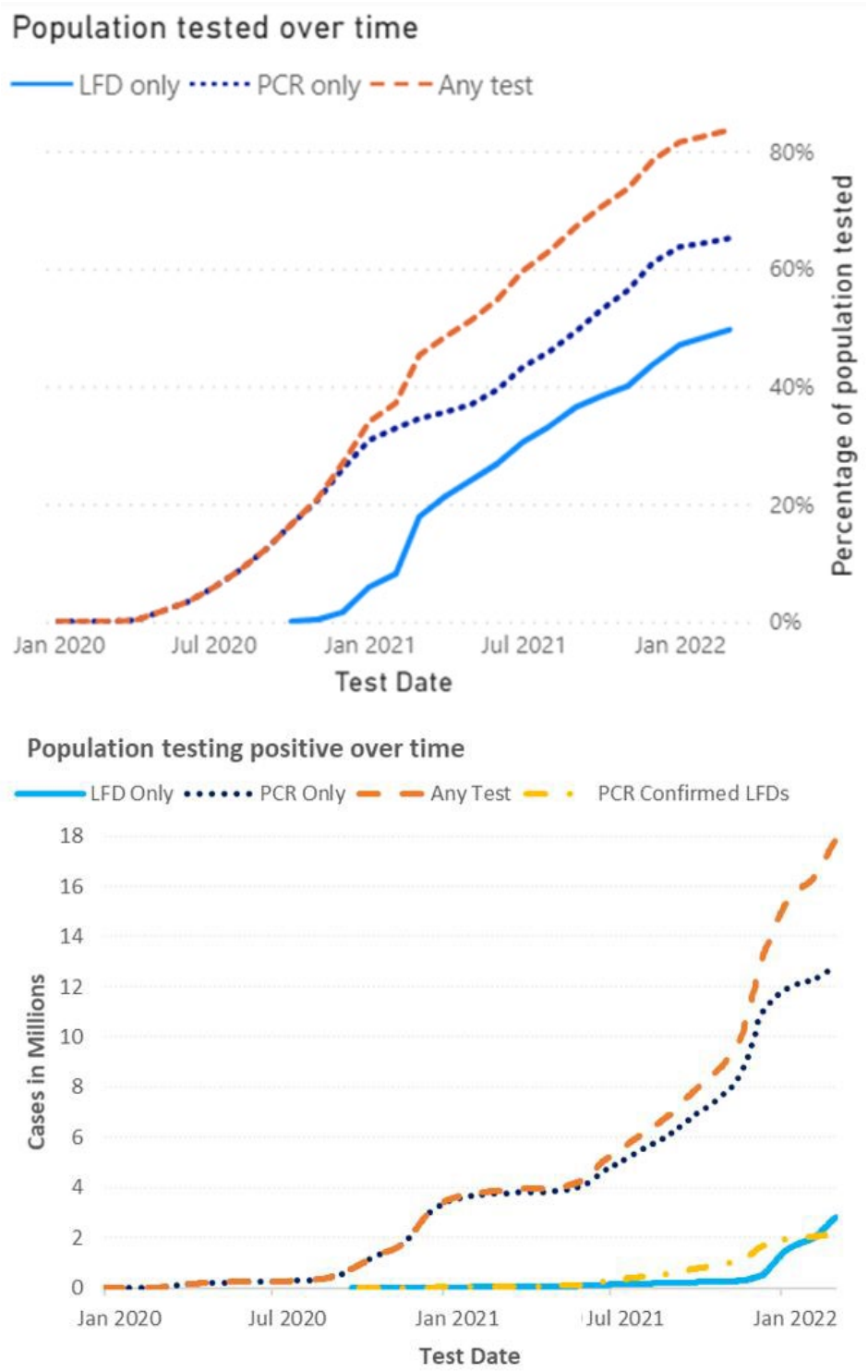
1. Introduction

1.1 Testing policy context

From Autumn 2020, NHS Test and Trace (now part of UKHSA) introduced Lateral Flow Device (LFD) antigen testing to identify cases of coronavirus (COVID-19). Ongoing Evaluation (OE) was undertaken to assess the performance of LFD kits used within the National Testing Programme (NTP) through the prospective collection of paired LFD-PCR samples (dual-tests). This report documents the results of the evaluation, commencing from when LFD testing began, through until provision of free testing for the general public ended in March 2022.

Figure 1 shows the percentage of the English population who have taken a COVID-19 test since the start of the pandemic and the percentage of those who tested positive.

Figure 1. The percentage of the English population who registered a COVID-19 test (PCR or LFD) (top)[2] from 1 Jan 2020 to 31 March 2022, and the number of reported positive cases in the same time period (bottom)[3]



[2] The data has been deduplicated. Three data sets were used: LFD only, PCR only, and any test (including LAMP). For each data set, the first/earliest test in Pillar 2 was included, then all subsequent tests from that individual were excluded. Individuals were identified using the EDGE-ERN code which takes into account people submitting tests with a slight change in details such as a change of address or use of a middle name.

[3] Source: Cases in England | Coronavirus in the UK (data.gov.uk). Data not deduplicated.

The structure of testing for coronavirus within the UK is separated under different operational Pillars. Patient testing, along with some testing of staff and some outbreak testing, has been delivered by the NHS under Pillar 1. Testing of the wider population in alignment with government policy has been delivered under Pillar 2.

Initially, LFD tests were used for asymptomatic testing only in the 4 main testing groups:

- Group 1 – repeated testing to detect positive cases amongst asymptomatic individuals (and remove them from circulation)
- Group 2 – testing prior to an activity to reduce risk (this may be one or more tests)
- Group 3 – asymptomatic testing where there is a signal of a potential outbreak (or where there has been an outbreak) to control infections, or where there is perceived to be a higher risk
- Group 4 – daily testing of contacts to identify positive cases early

Use of LFDs was extended to include symptomatic testing (previously predominated by PCR testing) in specific circumstances:

- concurrent testing, for example, as part of the testing regimes in place for the purposes of dispensing antiviral medication to eligible individuals with COVID-19
- ending of self-isolation early for individuals with COVID-19 in England testing negative on day 5 and day 6 of their self-isolation period
- staff who work in vulnerable settings such as the NHS and social care
- high risk settings such as homeless shelters and prisons

Testing with LFDs was through assisted-test (ASSIST) or self-test (SELF) delivery channels. In assisted testing, the user or a trained individual performed the swab, and the test result was interpreted by a trained individual. In self-testing, the user (or relevant person on behalf of the user such as a parent/guardian/carer) self-swabbed and interpreted the test themselves without assistance. LFD testing was initially offered through assisted testing at designated Asymptomatic Test Sites (ATS) and was followed by the introduction of home-based self-testing. As part of a test-trace-isolate approach, asymptomatic testing at scale allowed more people with transmissible virus to be detected and offered the potential to break chains of transmission.

The following LFD kits were utilised within the NTP during the period from 8 November 2020 to 21 March 2022:

- Innova SARS-CoV-2 Lateral Flow Antigen Test (Innova 25)
- DHSC COVID-19 Self-Test kit (DHSC 3&7)[4]
- Orient Gene COVID-19 Ag Rapid Test Cassette LFD antigen tests (Orient Gene)

[4] Innova 25 and DHSC 3&7 are the same LFD kit, except that with Innova 25 the bottle of buffer solution supplied is larger and the result is read by a trained professional.

- Acon Flowflex SARS-CoV-2 antigen rapid test (Self-Testing) kit (Acon)
- SureScreen SARS-CoV-2 Antigen Rapid Test Cassette V2 (SureScreen); excluded from statistical evaluation due to insufficient numbers of test results

1.2 Performance monitoring

To ensure devices used across testing services (asymptomatic (ASYMP) and symptomatic (SYMP)) were appropriate, safe, and operated as expected, a robust quality assurance system was put in place featuring both wet lab and field testing. Initial validation studies were carried out at UKHSA's Porton Down laboratories on all LFDs prior to use in the testing programme[5]. With the emergence of each new variant of concern, further validation studies have been conducted on those LFDs in use in the NTP. Post-deployment performance of devices has been continuously monitored through both evaluation of reporting data (Real World Data) and prospective evaluation known as Ongoing Evaluation (OE). Where UKHSA is the legal manufacturer of the LFD kit, evaluation outcome reports are provided to the Medicines and Healthcare products Regulatory Agency (MHRA) on a regular basis. The data for the OE reports previously submitted to MHRA was within the reporting period of this evaluation and included in the data set for this report.

In OE, as described in this report, LFD performance was evaluated using qRT-PCR as the reference test. The qRT-PCR samples were collected specifically for the purposes of evaluating LFD performance within the NTP.

qRT-PCR is the widely accepted comparator test and the only reference test available at the scale required for OE within the national Test and Trace infrastructure. However, it is not an ideal reference since the 2 tests assess distinct biological materials (viral RNA and viral antigen). qRT-PCR can detect very small amounts of viral RNA and gives an indication of how much of it is present, but it does not distinguish between virus that can replicate and therefore transmit, and virus that cannot. A guiding principle in the implementation of LFDs has been their ability to identify individuals with transmissible virus so that they could be directed to self-isolate in accordance with government policy. It is known that infectiousness increases at high viral concentrations (2, 9, 10) at which LFD performance is also at its highest (2, 6, 7, 8). In this evaluation, we therefore also looked to review the relationship with viral concentration and how it affected sensitivity of LFDs used within the NTP.

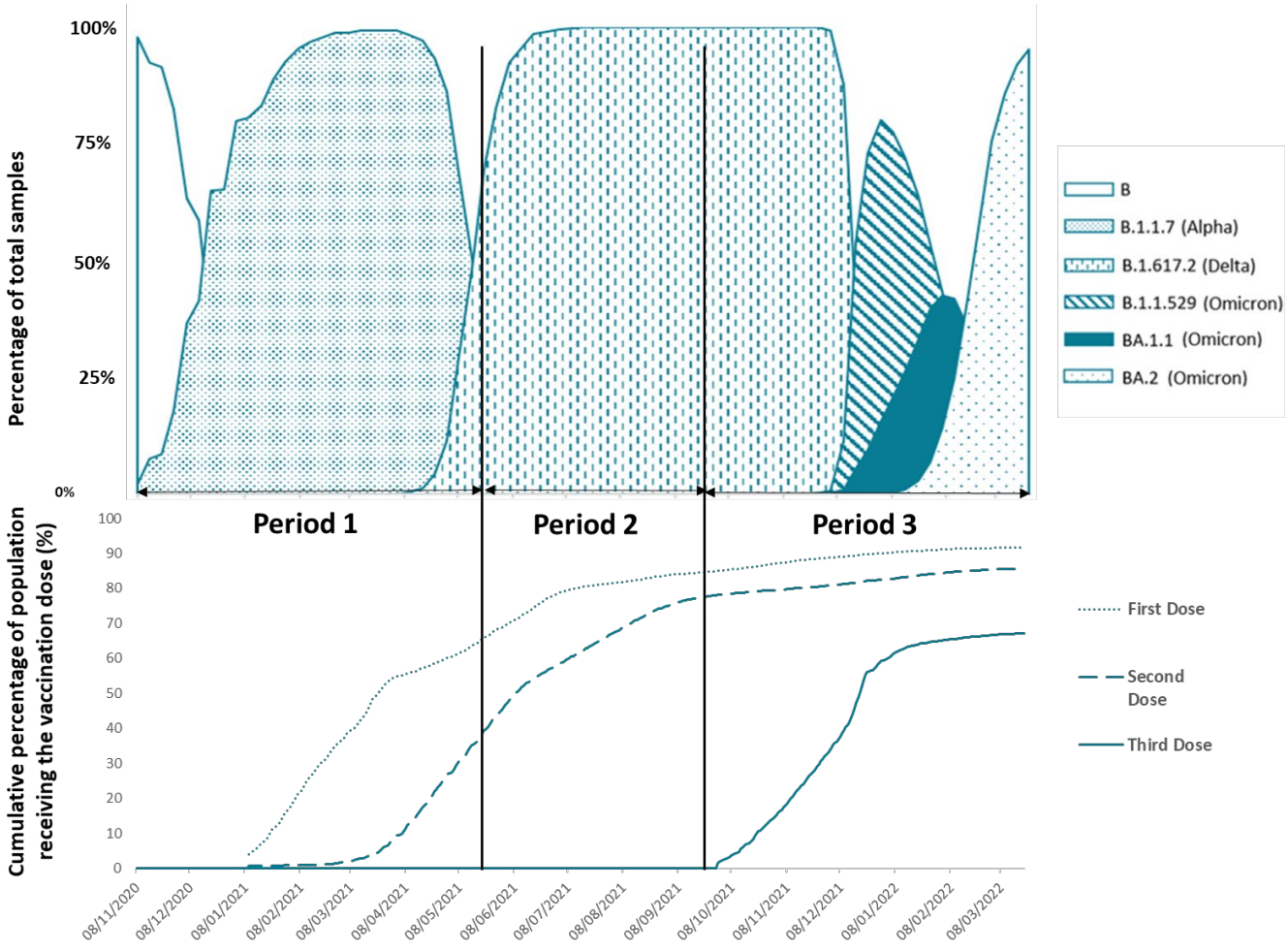
Prior to deployment by UKHSA, the performance of Biotime LFDs was evaluated using PCR and LFD dual-test results. Those evaluations came from field studies called LFD001 and LFD002 (see 2.5 Baseline comparisons for more details) and they set the benchmark upon which the MHRA Exceptional Use Authorisation (EUA) was granted for the use of Biotime LFDs. As such, they have been adopted for all subsequent assessments of safety and effectiveness of devices deployed by UKHSA and its NTP predecessors as the baselines against which performance is compared.

[5] Details of UKHSA's LFD validation process and its results can be found at Outcome of the evaluation of rapid diagnostic assays for specific SARS-CoV-2 antigens (lateral flow devices)

1.3 Epidemiological context of the evaluation

Testing has taken place in an evolving environment in terms of testing policy, disease prevalence, emergence of viral mutations and variants, and vaccination deployment. There was therefore a need to assess that lateral flow devices were working appropriately despite these changing factors. The overall time period covered in this report was further broken into 3 time periods (for certain analyses) in order to examine if performance changed or differed over time. Figure 2 shows vaccination uptake and variant distribution during the 3 time periods.

Figure 2: Graph showing the distribution of SARS-CoV-2 variants in England from 5 September 2020 to 21 March 2022[6] at the top, and the percentage of people in the UK who have received a SARS-CoV-2 vaccination, shown as a cumulative percentage of the population aged 12 and over[7]



[6] Wellcome Sanger Institute. COVID-19 Genomic Surveillance (viewed in April 2022)

[7] UK government. Vaccinations in the United Kingdom (viewed in April 2022)

1.4 Objectives of the evaluation of LFD performance

This report provides evidence of how LFDs performed in different test settings and services within the NTP during full scale deployment compared to original baseline evaluations.

The report is structured around the following 5 objectives:

Sensitivity

To compare the sensitivity of LFDs in post deployment with baseline results, using a non-inferiority margin of -10%; this is the primary outcome for this evaluation of LFD performance.

Impact of other factors on sensitivity

To evaluate the impact of factors such as viral concentration, testing environments, variants of concern (VoC), vaccination, and symptom status, using uni- and multifactorial logistic regression analysis.

Specificity

To derive the specificity of LFDs in post deployment.

PPV and NPV

To evaluate the post-deployment Positive Predictive Value (PPV) and Negative Predictive Value (NPV).

Void rate

To record the post-deployment void rate.

1.5 Hypothesis

The hypothesis was that performance of LFDs in post-deployment was non-inferior to observed baseline. It should be noted that non-inferiority was only assessed for the primary outcome of sensitivity.

2. Data

2.1 Testing channels and services

The NTP offered testing to people through different channels, including:

- community testing – asymptomatic testing offered to the general public, through
- local and regional test sites
- home testing
    - LFD Direct – LFD kits sent to the general public on request
    - pharmacies – LFD kits collected on request from pharmacies by the general public
- Citywide Testing (Liverpool) – asymptomatic testing of the general population in Liverpool at a time of high prevalence in the city
- Targeted Community Testing (TCT) – asymptomatic testing to disproportionately impacted and under-served groups who are likely to be suffering health inequalities and worse outcomes
- public and private sector industries – asymptomatic testing in workplaces in the public and private sector
- universities and schools (primary and secondary) – twice weekly asymptomatic testing of both students and staff
- for analysis purposes, universities were considered a separate testing channel to schools, and schools were broken into age 18 and under (presumed to be pupils), and over 18s (presumed to be staff)
- health – asymptomatic testing of staff in healthcare settings, including NHS, independent healthcare providers, frontline healthcare providers
- surge testing – increased testing aimed at asymptomatic subjects and enhanced contact tracing in specific locations
- In Person Testing Channel (IPTC) – onsite symptomatic testing at mobile testing units in Scotland
- regional or local test sites (RTS/LTS) – onsite symptomatic testing at regional or local test sites

2.2 Enrolment in Ongoing Evaluation

Ongoing Evaluation (OE) involved collecting dual-test samples (LFD and qRT-PCR) from individuals who were accessing testing through one of the testing channels listed above, for the specific purpose of prospectively evaluating the performance of LFDs in the NTP. Subjects were invited to participate at random either on arrival at a test site, or, in the case of home testing (such as LFD Direct), OE test bundles were sent directly to the individual's home on a 'push' basis.

Subjects were provided with a participant information sheet (PIS) including instructions on how to complete the additional test and consented if they wished to take part. They were provided with the supplementary self-test (either LFD or qRT-PCR depending on what the testing channel offered as standard) or were assisted with the supplementary test if that was the standard service.

The evaluations were carried out in live testing services, therefore standard testing was always prioritised over the supplementary one, meaning there was no randomisation of swabbing order in sample collection. If standard testing was LFD, subjects were asked to put the LFD to one side to develop while providing the PCR sample so that they were unaware of the result of either test at the point of agreeing to participate. Subjects or operatives were asked to interpret their LFD result either as positive, negative or void. Individuals assigned a positive result by LFD were categorised as LFD positive. They were unaware of the PCR result at the time of LFD testing.

Individuals assigned a positive result by the laboratory performing the qRT-PCR analysis were categorised as PCR positive. The laboratories were unaware of the LFD result.

2.3 Data processing

Paired LFD and qRT-PCR (dual-test) results from OE were collected into UKHSA's database directly as well as pulled from other databases and platforms such as EDGE (Environment for Data Gathering and Engineering – NHS Test and Trace research database) and were then collated into a master data workbook.

The data was cleaned by:

- removing duplicate LFD and PCR data
- ensuring consistency in coding of string variables
- removing outliers (clear anomalous results such as an age of 0 or 193) and inaccurate or invalid data (for example, incomplete barcodes or inaccurate postcodes)
- sourcing missing data by querying other source databases to complete the data set, for example, SGSS Datamart[8] was used to identify variant sequencing and genomic results associated with the qRT-PCR test
- validating the data with previous historical data which passed the quality threshold

[8] Information about SGSS (Second Generation Surveillance System) can be found at A guide for diagnostic laboratories

Figure 3 shows a CONSORT diagram of the number of dual-tests collected and the number excluded at different stages of data processing and analysis.

Figure 3. CONSORT diagram showing inclusion/exclusion of dual-tests in the evaluation

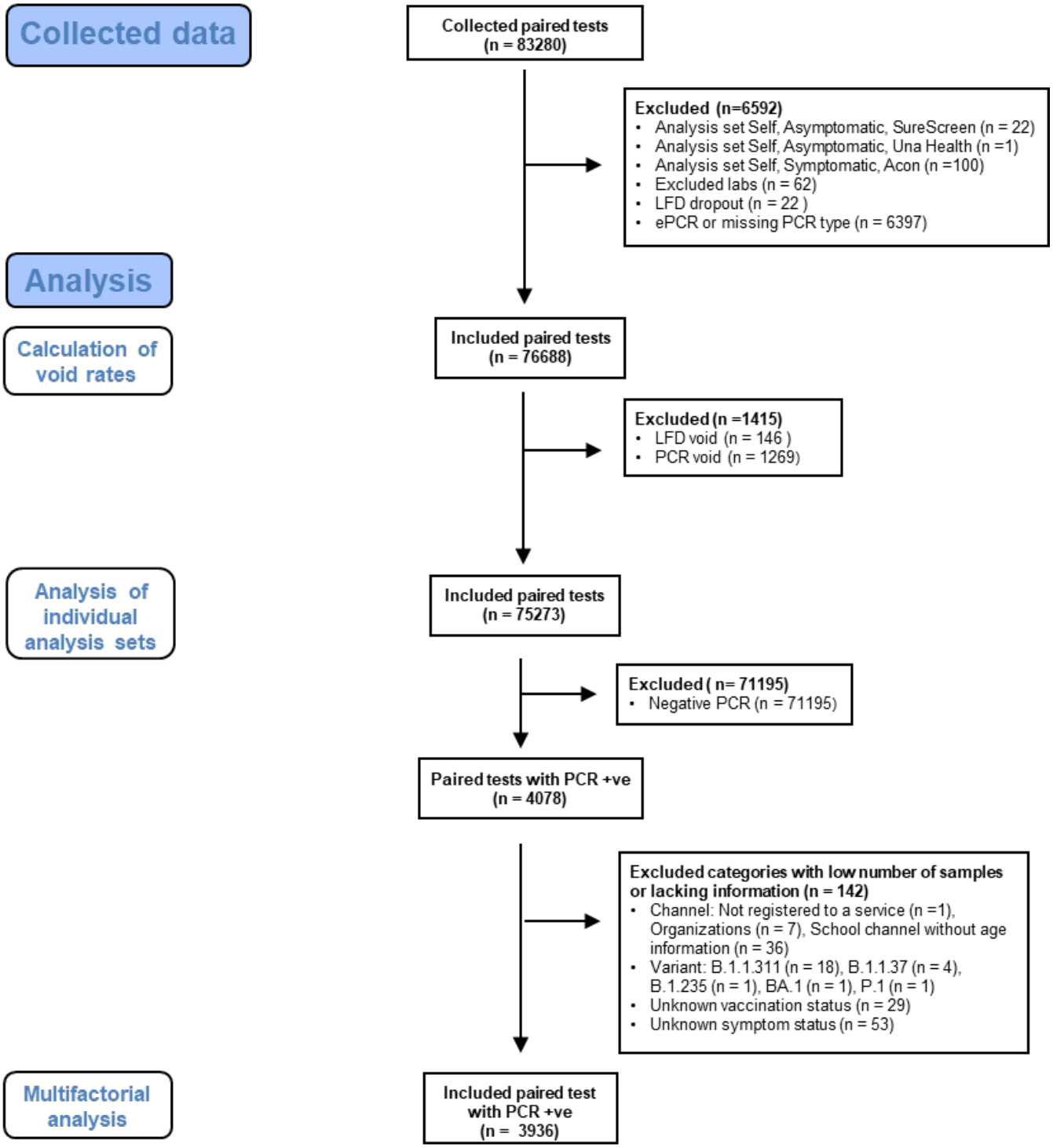


Figure 3. CONSORT diagram showing inclusion/exclusion of dual-tests in the evaluation – text equivalent

(On the left hand side)

Collected data

Analysis:

- Calculation of void rates
- Analysis of individual analysis sets
- Multifactorial analysis

(In the centre of page start of diagram)

Collected paired tests (n=83280)

(Arrow pointing down to 'Included paired tests (n=76688)' and arrow pointing right to 'Excluded (n-6592)')

Excluded (n-6592):

- Analysis set Self Asymptomatic SureScreen (n=22)
- Analysis set Self Asymptomatic Una Health (n=1)
- Analysis set Self Symptomatic Acon (n=100)
- Excluded labs (n=62)
- LFD drop out (n=22)
- ePCR or missing PCR type (n=6397)

Included paired tests (n=76688)

(Arrow pointing down to 'Included paired tests (n=75273)' and arrow pointing right to 'Excluded (n=1415)')

Excluded (n=1415):

- LFD void (n=146)
- PCR void (n=1269)

Included paired tests (n=75273)

(Arrow pointing down to **'**Paired tests with PCR +ve (n=4078)' and arrow pointing right to 'Excluded (n=71195)')

Excluded (n=71195):

- Negative PCR(n=71195)

Paired tests with PCR +ve (n=4078)

(Arrow pointing down to 'Included paired test with PCR +ve (n=3936)' and arrow pointing right to 'Excluded categories with low numbers if samples or lacking information (n=142)')

Excluded categories with low numbers if samples or lacking information (n=142):

- Channel – Not registered to a service (n=1), Organisation (n=7), Schools channel without age information (n=36)
- Variant – B1.1.311 (n=18), B1.1.37 (n=4), B1.235(n=1), BA.1 (n=1), P.1(n=1)
- Unknown vaccination status (n=29)

- Unknown symptom status (n=63)

Included paired test with PCR +ve (n=3936)

(End of diagram)

Once the fully compliant master workbook was created, only the relevant fields were passed onto the Biostatistics team for analysis. These fields included: reporting period, study name, analysis set, date of sample collection, LFD kit, LFD device, service, testing channel, LFD/PCR barcodes, LFD/PCR results, viral concentration[9], log viral concentration, viral concentration category, PCR type, age, sex, vaccination status, symptom status, date of onset of symptoms, and variant information (WHO name and Sanger lineage).

2.3.1 Variant assignment

The report stratified positive PCR samples by major lineages including B.1.177, B.1.1.7 (Alpha), B.1.617.2 (Delta), and B.1.1.529 (Omicron). These were identified via the following methods:

- sequencing – method used to identify major variant and exact sub-lineage
- genotyping – method used to identify major variants
- a geo-temporal method used to impute variant where sequencing or genotyping data was unavailable for the sample (see Appendix 5: Variant assignment method for more details)

2.3.2 Symptom status

Subjects were asked to describe their symptoms when submitting their results. Only subjects responding with at least one of the 3 cardinal symptoms were considered symptomatic:

- high temperature
- ageusia/anosmia (loss of taste or smell)
- dry continuous cough

2.4 Analysis sets

The performance of each LFD kit was analysed within the test setting (self-testing or assisted testing) and service (symptomatic or asymptomatic) in which they were used. Each ‘analysis set’ thereby consisted of test setting, service, and LFD kit. Table 10 in ‘Appendix 2: Analysis sets’ contains details of all the analysis sets. Those sets which met the sample size requirements (outlined in 2.6 Sample size) of having at least 154 PCR positive results, and were therefore analysed, were named as:

[9] See Appendix 4: qRT-PCR linearity data for the conversion formulae from Ct to VC.

- SELF-ASYMP-DHSC 3&7
- SELF-ASYMP-ACON
- SELF-ASYMP-ORIENT GENE
- ASSIST-ASYMP-INNOVA 25
- SELF-SYMP-DHSC 3&7
- SELF-SYMP-ORIENT GENE

Although separate services were set up to offer testing to either the symptomatic or asymptomatic, individuals were not prevented from testing if they did not fulfil the criteria. In asymptomatic services where individuals self-tested, between 4.5 and 16% of those tested reported symptoms (see Table 12). In asymptomatic services with assisted testing, fewer than 1% reported symptoms (see Table 13). In symptomatic services, approximately 50% of individuals who tested reported symptoms (see Table 12). The analysis set name refers to the service type being offered, rather than the symptomatic status of the subjects being tested. Symptoms were self-reported; people attending symptomatic testing sites were not examined for presence of symptoms; information on presence and nature of symptoms was accepted at face value.

Dual-testing results (paired LFD and PCR) were included in the data set for analysis only if the qRT-PCR samples were processed in laboratories where there was adequate cycle threshold (Ct) information available to convert the Ct value to a viral concentration (VC), and the test was performed within the reporting period.

Details of the dual-test results included in the data set for analysis are shown in Table 1.

Table 1. Samples within inclusion criteria for each analysis set

Analysis set	Date of first result	Date of last result	Number of partici-pants recruited	Number of dual-tests for analysis*	PCR Positive	PCR Negative	Dual-tests per analysis set**
Baseline: ASSIST	04/11/2020	18/12/2020	4,356	4,294	691	3,534	**4,225**
Baseline: SELF	23/11/2020	09/01/2021	2,582	2,560	424	2,049	**2,473**
SELF-ASYMP-DHSC 3&7	29/03/2021	21/03/2022	23,290	21,256	235	20,747	**20,982**
SELF-ASYMP-ACON	03/08/2021	21/03/2022	7,356	6,035	207	5,722	**5,929**
SELF-ASYMP-ORIENT GENE	31/07/2021	21/03/2022	5,898	5,102	159	4,871	**5,030**
ASSIST-ASYMP-INNOVA 25	08/11/2020	04/03/2022	29,851	27,788	568	26,549	**27,117**
SELF-SYMP-DHSC 3&7	18/02/2021	21/03/2022	8,658	8,586	1,353	7,119	**8,472**
SELF-SYMP-ORIENT GENE	08/01/2022	21/03/2022	1,166	1,067	441	604	**1,045**
Totals including baseline			**83,157**	**76,688**	**4,078**	**71,195**	**75,273**
Totals excluding baseline			**76,219**	**69,834**	**2,963**	**65,612**	**68,575**

*After exclusion of dropouts, samples processed by ePCR/without PCR information, PCR positives without viral concentration, PCRs processed in labs where Ct conversion formula were not available

** Voids removed

Records with voids were removed from the analysis sets other than for calculating the void rate itself. Most of the analysis reported is based on a ‘completers’ analysis (that is, those with a positive or negative qRT-PCR and LFD test outcome).

Full details of demographic breakdown and the clinical characteristics of participants in each analysis set are shown in Appendix 6: Participant demographics and clinical characteristics.

2.5 Baseline comparisons

The baseline service evaluations (LFD001 and LFD002) which took place prior to deployment of LFDs in the National Testing Programme were used both as a reference against which to evaluate performance and as a basis for adjustment of VC and COVID-19 prevalence in order to ensure a fair comparison, for sensitivity and predictive values, respectively.

Device performance of each analysis set was compared to device performance in the appropriate baseline evaluation study which was LFD001 for assisted testing and LFD002 for self-testing[10]. Both evaluations assessed the performance of Innova LFDs at symptomatic test sites. These baselines set the benchmark upon which the MHRA Exceptional Use Authorisation (EUA) was granted for use of Biotime LFDs and have been adopted for all subsequent assessments of safety and effectiveness of devices deployed by UKHSA and its NTP predecessors. The post-deployment results in this evaluation were compared to the baseline results shown in Table 2.

Comparing real-world (post-deployment) outcomes to baseline did not necessarily result in an unbiased (impartial) comparison because the differences in performance outcomes can be caused by variation of other factors (such as VC or prevalence of COVID-19). By taking these factors into account, a comparison between post-deployment and baseline outcomes might be fairer and less influenced by uncontrolled confounding factors. Besides observed values, we therefore present additional sensitivity metrics after adjustment for VC (see section 4.1) as well as positive and negative predictive values after adjustment for prevalence (see section 4.4).

For details of the baseline methodology see ‘Appendix 3: Methodology of the baseline studies’.

[10] Baseline study can be found in the reference list (1). In this, LFD001 equates to RTS Field Test 1. LFD002 equates to RTS Field Test 3 and 4 combined.

Table 2. Baseline device performance outcomes against which the analysis sets were compared

Outcome at baseline	LFD001 [assisted] (%, 95%CI)	LFD002 [self] (%, 95%CI)
Sensitivity	54.41 (50.62, 58.17)	50.00 (45.14, 54.86)
Specificity	99.58 (99.30, 99.76)	99.07 (98.56, 99.44)
Voids	0.40 (0.23, 0.63)	2.46 (1.90, 3.14)
PPV	96.16 (93.75, 97.84) at 16.36% prevalence	91.77 (87.45, 94.98) at 17.15% prevalence
NPV	91.78 (90.87, 92.63)	90.54 (89.26, 91.72)

2.6 Sample size

Power calculations were carried out to determine the necessary sample sizes to detect differences in sensitivity compared to the baseline evaluations. To this end, sample sizes were calculated based on formula 3.7 from Machin and others (5) with the sensitivity from LFD001 or LFD002 as given proportions. To detect an absolute difference in sensitivity of 10% with power of 80% and at a significance level of 5, we needed at least 154 PCR positive cases in an analysis set. A sufficient number of PCR positive cases were included in the analysis sets listed in Table 1 but not for the SELF-ASYMP-SURESCREEN analysis set (N=1), or for SELF-SYMP-ACON (N=46) shown in Table 10 and these were therefore removed from this performance report.

2.7 Ethics

The ongoing evaluations that have generated data included in this report do not constitute research and, as such, do not require research ethics approval. However, to hold the NTP to the highest level of accountability, UKHSA sought ethical review and approval of an umbrella framework and associated participant-facing materials for the prospective data collection elements of Service Evaluation and Ongoing Evaluation from Public Health England's Research Ethics and Governance Group (PHE REGG)[11]. This was reviewed and approved under REGG R and D 438.

[11] Following establishment of UKHSA this body is now UKHSA Research Ethics and Governance Group (UKHSA REGG).

3. Statistical methods

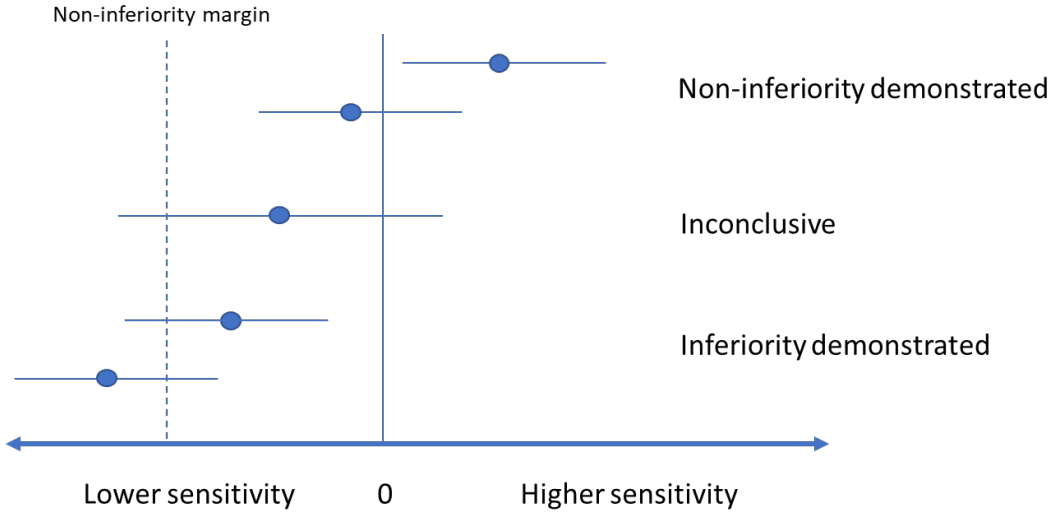
3.1 Objective 1: Sensitivity

The primary outcome was sensitivity defined as Sensitivity Se = TP/(TP+FN) with TP (true positive) as the number of positive LFD results confirmed by a positive PCR result and FN (false negative) as the number of negative LFD results corresponding to a positive PCR.

Non-inferiority was assessed based on 95% confidence intervals for differences in sensitivity between baseline and post-deployment. In this test setting, non-inferiority is demonstrated when the lower CI is within a chosen non-inferiority margin, while inferiority is indicated when the upper CI of the difference is smaller than 0. The outcome is regarded as inconclusive when both the non-inferiority margin and the zero difference are included in the CI. Figure 4 illustrates the different outcomes.

CI for differences in sensitivity were calculated using the Wald method. The non-inferiority margin was set to -10% to be consistent with previous service evaluations. Given the size of our analysis sets and the sample calculations (2.6 Sample size), would expect to detect with at least 80% probability (at a significance level of 5%) if the sensitivity decreased by 10% or more.

Figure 4. Illustration of the different outcomes of non-inferiority testing



Sensitivity was additionally presented as adjusted for VC[12]. More specifically, VC adjusted sensitivity was calculated by weighting the stratified sensitivity by the proportion of cases observed in the corresponding category in the baseline data set. This was achieved by summing the product of the proportion of PCR positive samples at each VC band (>10M, 1M-

[12] VC describes the amount of virus in the transport medium in RNA copies/mL and is thus only an indirect indicator of the amount of virus present in a person's nasopharynx. To calculate the VC of each sample, qRT-PCR cycle threshold values were converted to VC using laboratory specific and gene specific conversion formulae (see Appendix 4: qRT-PCR linearity data).

10M, 100K-1M, 10K -100K, 1K-10K, 100-1K, <100) in the relevant baseline study, and the sensitivity observed at that corresponding band in the analysis set. Confidence intervals for the VC adjusted sensitivity were derived by bootstrapping.

3.2 Objective 2: Impact of other factors on sensitivity

To assess the impact of additional factors on the performance of LFDs, logistic regression models were constructed and fitted to the data. Subsequently, they can be used to predict the probability of a TP (that is, the probability that a PCR positive subject will also be LFD positive) for specific values of VC or other covariates (factors). The probability of a TP can be interpreted here as a proxy for the expected sensitivity under defined conditions and to assess the impact of different factors on the performance of the LFD kits.

3.2.1 Unifactorial analysis[13]

Besides VC as an independent factor, the following covariates were individually included in the model and their statistical significance assessed after fitting:

- vaccination status (one dose, 2 doses, 3 doses)
- symptom status (symptomatic/asymptomatic)
- variant type: B.1.1.7 (Alpha), B.1.617.2 (Delta), B.1.1.529 (Omicron), B.1.177 (Spanish variant)
- channel of testing (grouped as outlined in 2.1 Testing channels and services)
- time period (with 3 periods of time: Period 1 (8 November 2020 to 21 May 2021), Period 2 (22 May 2021 to 21 September 2021), Period 3 (22 September 2021 to 21 March 2022)
- time lag between symptom onset and LFD test (split into 2 quantiles: 0 to 1 days, 2+days[14])

The above covariates were included in the logistic regression analyses using TP as the outcome. The likelihood ratio test (LRT) was used as a global test to determine if a given factor was statistically associated with the chance of a TP. As LRT were carried out for all analysis sets and all covariates, the resulting p-values were Bonferroni adjusted for multiple testing. An adjusted $p<0.05$ was set as significance threshold. Additionally, the odds ratios were reported to provide an estimate of the relative impact of a given covariate on the chance of a true positive outcome, along with the corresponding (Wald) 95% confidence intervals and p-values.

[13] Although the underlying logistic regression models contained 2 independent factors (VC and an additional covariate), we refer to the analysis here as unifactorial analysis, as only the significance of the additional covariate was assessed.

[14] The data was not adequate to split into quartiles as suggested in the statistical analysis plan, hence 2 quantiles used instead.

3.2.2 Multifactorial analysis

To investigate the simultaneous impact of multiple factors on LFD performance, a multifactorial logistic regression model was constructed and fitted to the data. VC, type of device, test setting (self or assisted), testing service (symptomatic or asymptomatic), symptom status, testing channel, vaccination status and variant were included as covariates. DHSC 3&7s, Innova 25s and pre-deployment (baseline) kits were all classified as Biotime devices for multifactorial logistic regression analysis (the physical LFD cartridge was the same, the kits simply contained a different buffer bottle and a different number of LFDs supplied in a pack) to avoid collinearity with the testing channel factor. For variant type and testing channel the most common strata were chosen as the comparator against which others were compared.

Testing over time (the time period of testing) was not included as a high degree of association (correlation) with vaccination status and variant was observed which would result in an inflated error margin for effect size estimates. Interaction terms between independent factors were also not included, as their inclusion did not decrease Bayesian Information Criterion (data not reported).

3.3 Objective 3: Specificity

The second outcome was specificity defined as Specificity Sp = TN/(TN + FP) with TN (true negative) as the number of LFD negative results confirmed by a PCR negative result and FP (false positive) as the number of positive LFD results corresponding to a PCR negative. Specificity was summarized using descriptive statistics and 95% confidence intervals (Clopper-Pearson).

Objective 4: PPV and NPV

Positive predictive value (PPV) = TP/(TP+FP)

Modelled positive predictive value adjusted for prevalence:

$PPV_{modelled} = \frac{Se * Pre}{Se * Pre + (1 - Sp) * (1 - Pre)}$ with Se as Sensitivity, Sp as Specificity of an analysis set and Pre as PCR-based prevalence (Pre = (TP+FN)/(TP+FN+FP+TN)) from baseline study.

Text version of equation: Positive predictive value equals true positives divided by (true positives plus false positives).

Negative predictive value (NPV) = TN/(TN+FN)

$NPV_{modelled} = \frac{Sp * (1 - Pre)}{Sp * (1 - Pre) + (1 - Se) * Pre}$ with Se as Sensitivity, Sp as Specificity of an analysis set, and Pre as PCR-based prevalence (Pre = (TP+FN)/(TP+FN+FP+TN)) from baseline study.

Text version of equation: Negative predictive value equals true negatives divided by (true negatives plus false negatives).

These outcomes were summarised using descriptive statistics and 95% confidence intervals (Clopper-Pearson).

3.5 Objective 5: Void rate

Void rate defined as the percentage of LFD results which do not show a control line. This outcome was summarized using descriptive statistics and 95% confidence intervals (Clopper-Pearson).

4. Results and discussion

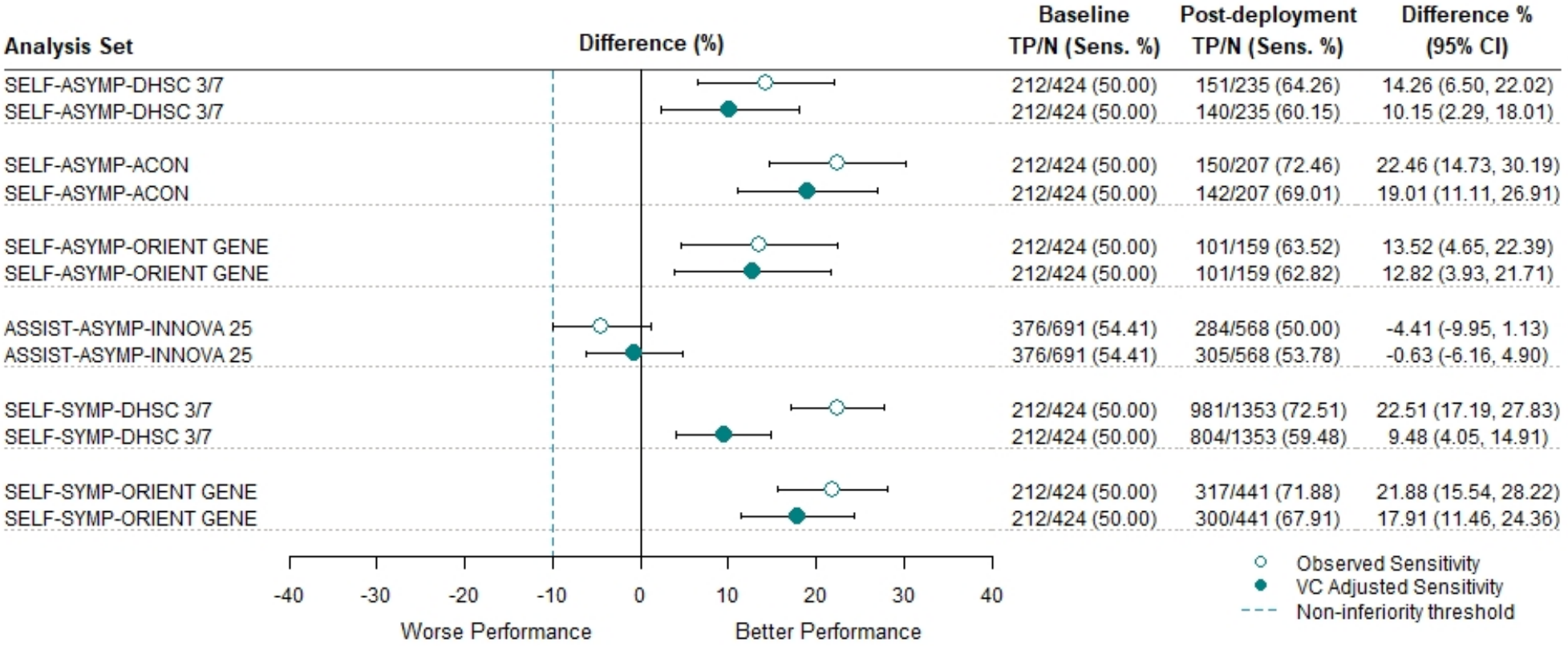
4.1 Objective 1: Sensitivity

4.1.1 Overall sensitivity results

For all analysis sets, the post-deployment performance in terms of sensitivity was non-inferior compared to baseline (Figure 5 and Table 14) as defined by the margin of -10%. As such, all LFDs were considered to have performed within an acceptable range. Notably, the majority (apart from assisted testing in asymptomatic settings using Innova 25s) had higher sensitivity than at baseline.

The observed sensitivity in the ASSIST-ASYMP-INNOVA 25 analysis set was lower compared to all other analysis sets (despite still demonstrating non-inferiority). The results in section 4.2 show that being symptomatic was a significant predictor of a TP. Nearly 91% of the PCR positives in ASSIST-ASYMP-INNOVA 25 were asymptomatic which could be contributing to the lower observed sensitivity in this analysis set (all other analysis sets contained more symptomatic than asymptomatic PCR positives samples with a mean of 30.57% asymptomatic, see Table 15 in ‘Appendix 7: Analysis outputs’).

Figure 5. Forest plot of differences in overall sensitivity between post-deployment (observed and VC adjusted) compared to baseline – dashed line indicates non-inferiority margin



4.1.1.1 Adjustment for viral concentration

The conclusion that all analysis sets were non-inferior to baseline remained valid when the sensitivity in post-deployment was adjusted for the distribution of VC seen at baseline (Figure 5).

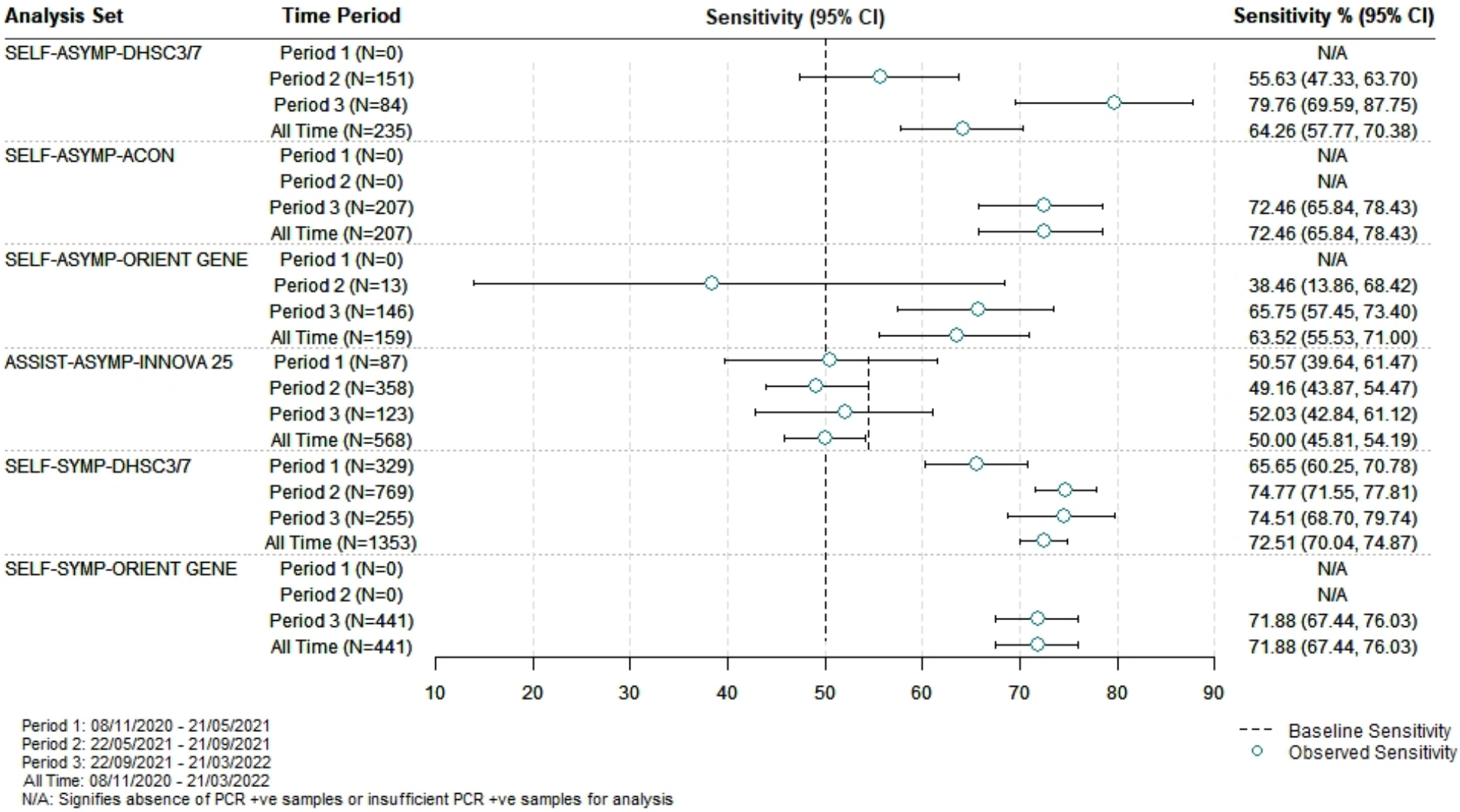
In one analysis set, the post-deployment VC-adjusted sensitivity was found to be lower than the baseline sensitivity, though still comfortably within the 10% non-inferiority allowance. This was ASSIST-ASYMP-INNOVA 25, where sensitivity was 53.78% (95%CI: 49.12, 57.57) compared to the baseline sensitivity of 54.41% with a difference of -0.63% (95%CI: -6.16, 4.90).

4.1.2 Sensitivity by strata

4.1.2.1 Sensitivity during different time periods

To assess if sensitivity changed over time and to check whether it remained within acceptable limits at different times, the all-time period was split into 3 time periods and sensitivity was calculated for each one. Figure 6 shows the results.

Figure 6. Observed sensitivity in different time periods



Period 1: 08/11/2020 - 21/05/2021
Period 2: 22/05/2021 - 21/09/2021
Period 3: 22/09/2021 - 21/03/2022
All Time: 08/11/2020 - 21/03/2022
N/A: Signifies absence of PCR +ve samples or insufficient PCR +ve samples for analysis

Sensitivity did vary over time though no consistent pattern appeared. All sensitivity results for each time period were higher in post-deployment compared to baseline, except for SELF-ASYMP-ORIENT GENE in time period 2. However, the sample size was small (n=13) leading to a large CI and preventing any firm statistical conclusions.

ASSIST-ASYMP-INNOVA 25 showed lower sensitivity in all 3 time periods compared to baseline. However, post-hoc analysis of difference shown in Table 3 indicated that post-deployment sensitivity was not statistically different to baseline in any of the time periods. Non-inferiority testing was inconclusive in all 3 time periods.

Table 3. Difference in sensitivity for ASSIST-ASYMP-INNOVA 25 compared to baseline for each time period

Time period	Baseline sensitivity % (95%CI)	Sensitivity % (95%CI)	Difference % (95%CI)
Period 1	54.41 (50.62, 58.17)	50.57 (39.64, 61.47)	-3.84 (-14.98, 7.30)
Period 2	54.41 (50.62, 58.17)	49.16 (43.87, 54.47)	-5.25 (-11.62, 1.12)
Period 3	54.41 (50.62, 58.17)	52.03 (42.84, 61.12)	-2.38 (-11.96, 7.20)

4.1.2.2 Sensitivity on/after day of symptom onset

To assess if sensitivity depended on the time delay between symptoms starting and taking the LFD test, sensitivity was calculated for different time delays (see Table 4).

Broadly speaking sensitivity remained high across the analysis sets averaging well above 50% over the 7 days. This was contrasted by a strong decrease in the average VC over the same time period (Table 4). This suggests that the decrease in levels of antigen lags behind the decrease in VC although caution is warranted due to the small sample sizes. For example, for SELF-ASYMP-ACON analysis set, the mean sensitivity up to a 7-day delay was around 66%, well above baseline performance levels (based on sample sizes between 3 and 11). This is contrasted by a more than 400-fold decrease in mean VC in the corresponding qRT-PCR samples.

As there is a move towards using LFDs in symptomatic settings, this result provides confidence that LFDs can be used at all stages of the infection cycle including days zero and day one of symptom onset.

Table 4. Sensitivity after different time delays starting from the day of symptom onset (in addition, the mean VCs of the corresponding PCR tests are shown)

Analysis set	Measure	Day0&1	Day2+	Day3+	Day4+	Day5+	Day6+	Day7+
SELF-ASYMP-DHSC 3&7	Sensitivity %, (95%CI)	76.92 (56.35, 91.03)	96.43 (81.65, 99.91)	95.00 (75.13, 99.87)	100.00 (73.54, 100.00)	100.00 (63.06, 100.00)	100.00 (29.24, 100.00)	100.00 (29.24, 100.00)
	N	26	28	20	12	8	3	3
	mean VC (RNA copies/ml)	7,337,933	8,113,057	2,882,011	2,071,284	1,835,584	12,353	12,353
SELF-ASYMP-ACON	Sensitivity % (95%CI)	93.33 (68.05, 99.83)	68.75 (41.34, 88.98)	63.64 (30.79, 89.07)	55.56 (21.20, 86.30)	60.00 (14.66, 94.73)	66.67 (9.43, 99.16)	50.00 (1.26, 98.74)
	N	15	16	11	9	5	3	2
	mean VC (RNA copies/ml)	4,193,291	8,063,547	349,899	390,022	178,054	8,994	11,223
SELF-ASYMP-ORIENTGENE	Sensitivity % (95%CI)	92.86 (66.13, 99.82)	80.95 (58.09, 94.55)	72.73 (39.03, 93.98)	75.00 (34.91, 96.81)	66.67 (22.28, 95.67)	50.00 (1.26, 98.74)	50.00 (1.26, 98.74)
	N	14	21	11	8	6	2	2
	mean VC (RNA copies/ml)	2,079,833	1,398,041	606,162	550,289	685,133	10,096	10,096
ASSIST-ASYMP-INNOVA 25	Sensitivity % (95%CI)	75.00 (47.62, 92.73)	66.67 (40.99, 86.66)	66.67 (34.89, 90.08)	85.71 (42.13, 99.64)	80.00 (28.36, 99.49)	75.00 (19.41, 99.37)	66.67 (9.43, 99.16)
	N	16	18	12	7	5	4	3
	mean VC (RNA copies/ml)	3,568,053	6,830,580	8,620,329	3,318,544	3,055,900	3,800,997	5,067,440

Analysis set	Measure	Day0&1	Day2+	Day3+	Day4+	Day5+	Day6+	Day7+
SELF-SYMP-DHSC 3&7	Sensitivity % (95%CI)	79.20 (71.03, 85.94)	77.42 (67.58, 85.45)	76.74 (61.37, 88.24)	74.07 (53.72, 88.89)	66.67 (29.93, 92.51)	40.00 (5.27, 85.34)	66.67 (9.43, 99.16)
	N	125	93	43	27	9	5	3
	mean VC (RNA copies/ml)	1,687,839	863,839	633,553	471,573	374,357	8,243	12,921
SELF-SYMP-ORIENT GENE	Sensitivity % (95%CI)	73.39 (67.01, 79.13)	81.56 (74.16, 87.59)	79.69 (67.77, 88.72)	82.14 (63.11, 93.94)	76.47 (50.10, 93.19)	70.00 (34.75, 93.33)	75.00 (19.41, 99.37)
	N	218	141	64	28	17	10	4
	mean VC (RNA copies/ml)	1,725,924	1,428,992	1,090,474	1,192,775	559,863	229,727	18,944

Note: Day 0 & 1 included LFDs on those respective days. Day2+ denotes LFD tests taken at day 2 after symptom onset or later. Similar denotations for the other columns.

4.2 Objective 2: Impact of other factors on sensitivity

4.2.1 Unifactorial analysis

To explore whether other factors besides VC influenced LFD sensitivity, logistic regression models were fitted to data of the analysis sets. The unifactorial models included the VC and an additional covariate (such as symptom status) as predictor variables.

The unifactorial analysis indicated cases where individual factors have a significant impact on the predicted probability of a TP (Table 18 in Appendix 7: Analysis outputs). Symptom status was found to be a statistically significant predictor of a TP in 3 out of 6 analysis sets; vaccination status was found to be a significant predictor of a TP in one analysis set; time period had a significant impact on sensitivity in one out of the 4 analysis sets in which it could be assessed; and testing channel had a significant impact on sensitivity in one out of the 5 analysis sets in which the analysis could be performed.

However, caution in the interpretation of these results needs to be taken as the unifactorial analysis did not take potential confounding by other variables (besides VC) into account. To overcome this limitation, multifactorial analysis was carried out to assess the impact of multiple covariates on performance of LFDs to identify any trends and better distinguish between

confounding factors. To obtain sufficient statistical power for such approach, the different analysis sets were combined and categories with low numbers of samples were removed (see Figure 3).

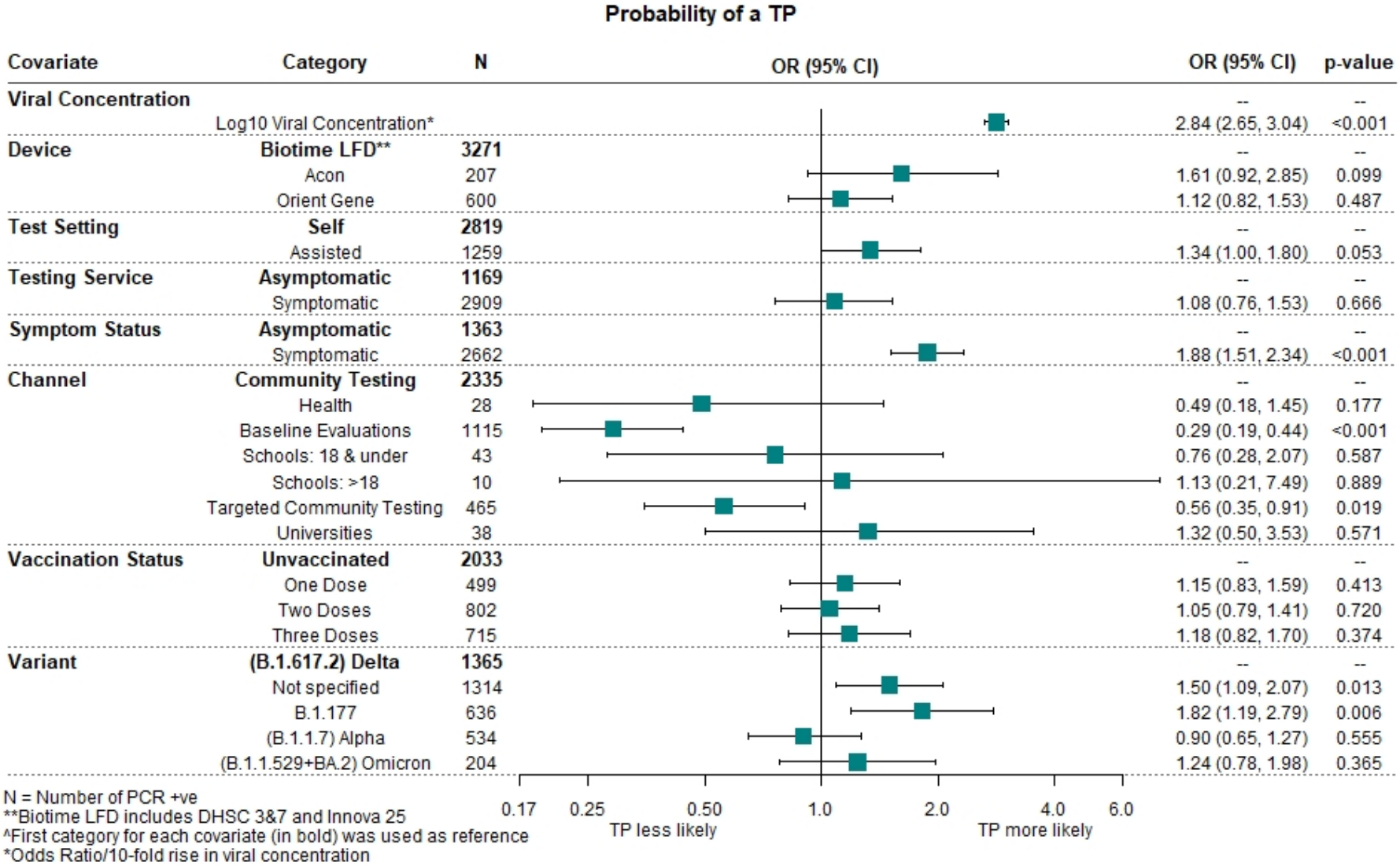
4.2.2 Multifactorial analysis

The results of the multifactorial analysis showed that both VC and symptom status (symptomatic disease) were significant predictors of an LFD returning a TP, see Figure 7 and Table 5. As VC increased the chances of a TP significantly increased, with an odds ratio of 2.84 (95%CI: 2.65, 3.04) for each unit increase in $\log_{10}$ VC (see Figure 7).

On average, over the observed VC range, symptomatic disease increased the chances of a TP compared to the asymptomatic with an odds ratio of 1.88 (95%CI: 1.51, 2.34) (see Figure 7).

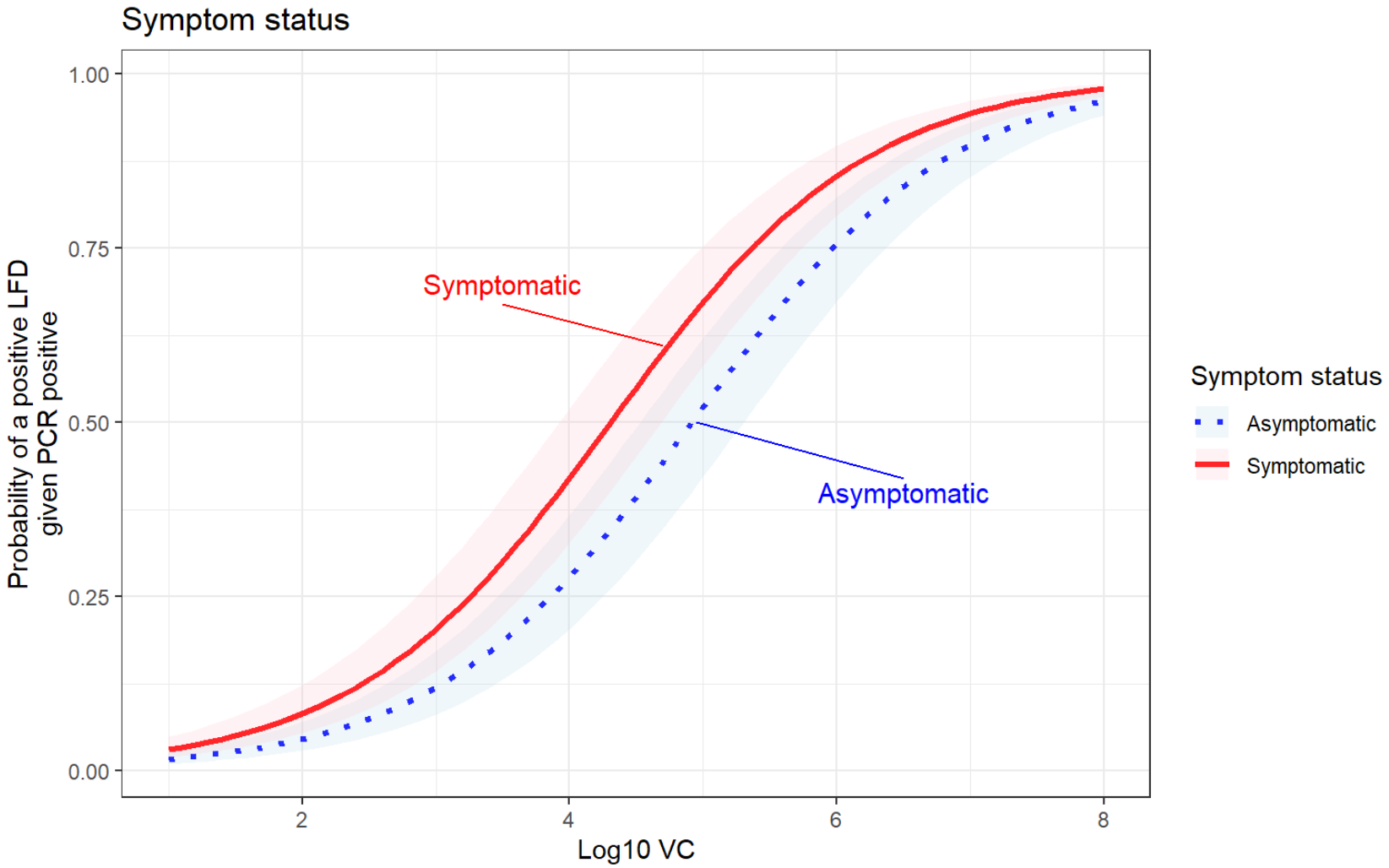
Figure 8 shows the results of the modelling for VC and symptom status. It illustrates the relationship between VC and sensitivity showing that at high VCs LFDs perform with highest sensitivity. This is the case in both the asymptomatic and symptomatic. The figure demonstrates that the symptomatic were significantly more likely than the asymptomatic to return a TP over the whole range of VC.

Figure 7. The effect of covariates on the probability of an LFD returning a TP



N = Number of PCR +ve
**Biotime LFD includes DHSC 3&7 and Innova 25
^First category for each covariate (in bold) was used as reference
*Odds Ratio/10-fold rise in viral concentration

Figure 8. Probability of a TP by VC and symptom status



Ribbons indicate 95% confidence intervals. Reference levels used for covariates: Device, Biotime LFD; Test setting, Self; Testing service, Asymptomatic; Symptom status, Asymptomatic; Channel, Community Testing; Vaccination status, Unvaccinated; Variant, B.1.617.2 (Delta).

As LFDs are now being used in symptomatic testing, this result provides confidence that LFDs perform well in symptomatic testing. To better characterise performance, sensitivity stratified by symptom status was calculated for each analysis set (see Table 16 in 'Appendix 7: Analysis outputs'). In all analysis sets, sensitivity in symptomatic cases was higher than in asymptomatic cases (see Figure 9). This could have been due to differences in VC, as VC tends to be higher in symptomatic individuals (see Figure 122 in Appendix 7: Analysis outputs which shows the distribution of VC by symptom status). Even after sensitivity was adjusted for VC however, this difference remained though it was smaller (see Figure 10). This suggests the result was not simply a consequence of different VCs in the 2 symptom status groups. It potentially indicates biological differences between the symptomatic and asymptomatic, with a higher level of viral antigen relative to the concentration of viral RNA in symptomatic samples.

In 3 analysis sets, VC adjusted sensitivity in the asymptomatic was lower than baseline (SELF-ASYMP-ACON, SELF-ASYMP-ORIENT GENE, and ASSIST-ASYMP-INNOVA). The first 2 of these had small sample sizes for the asymptomatic (n= 47 and n=39 respectively) which led to a less precise estimation of the sensitivity (with wide confidence intervals). When the difference between VC sensitivity in the asymptomatic and baseline was calculated for these analysis sets (see Table 17 in 'Appendix 7: Analysis outputs'), the lower CI was below the non-inferiority margin in 2 analysis sets and, in all analysis sets, the upper CI was above 0. As such, non-inferiority testing was inconclusive for SELF-ASYMP-ACON AND SELF-ASYMP-ORIENT GENE, and non-inferior for ASSIST-ASYMP-INNOVA 25.

Figure 9. Observed sensitivity of the symptomatic and asymptomatic within each analysis set (dashed vertical line represents the baseline)

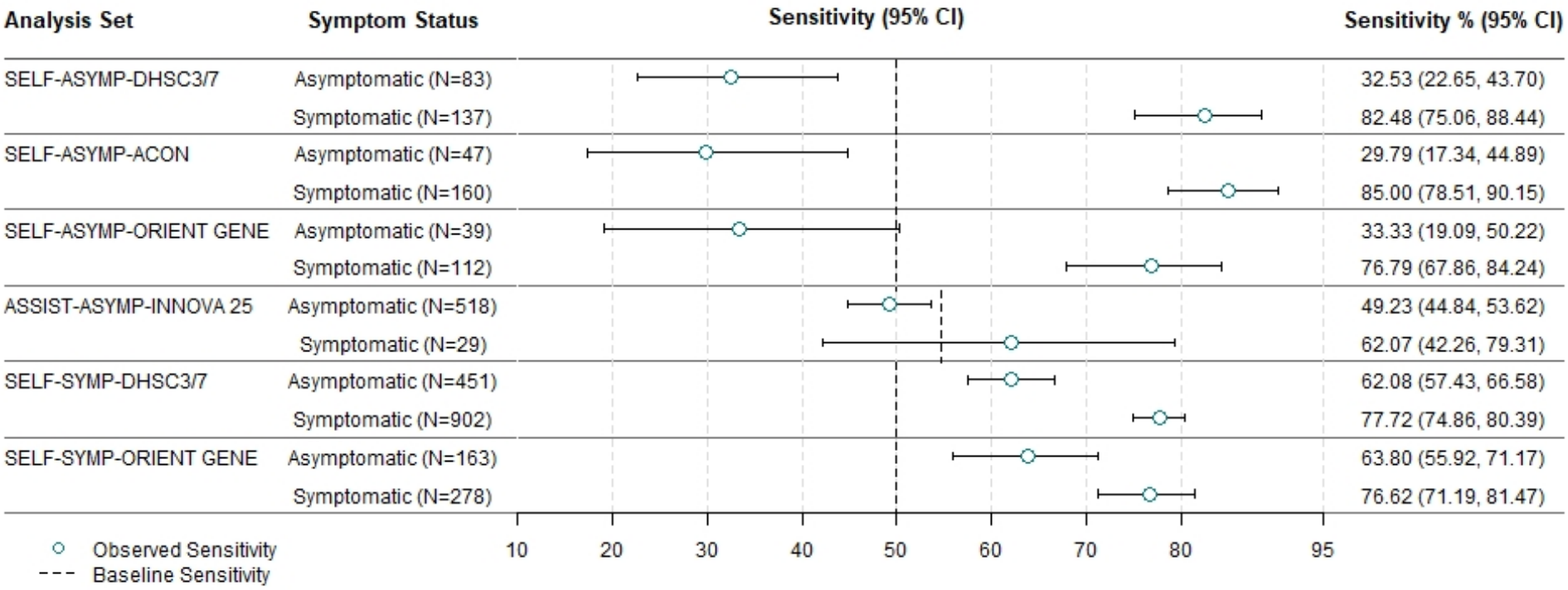
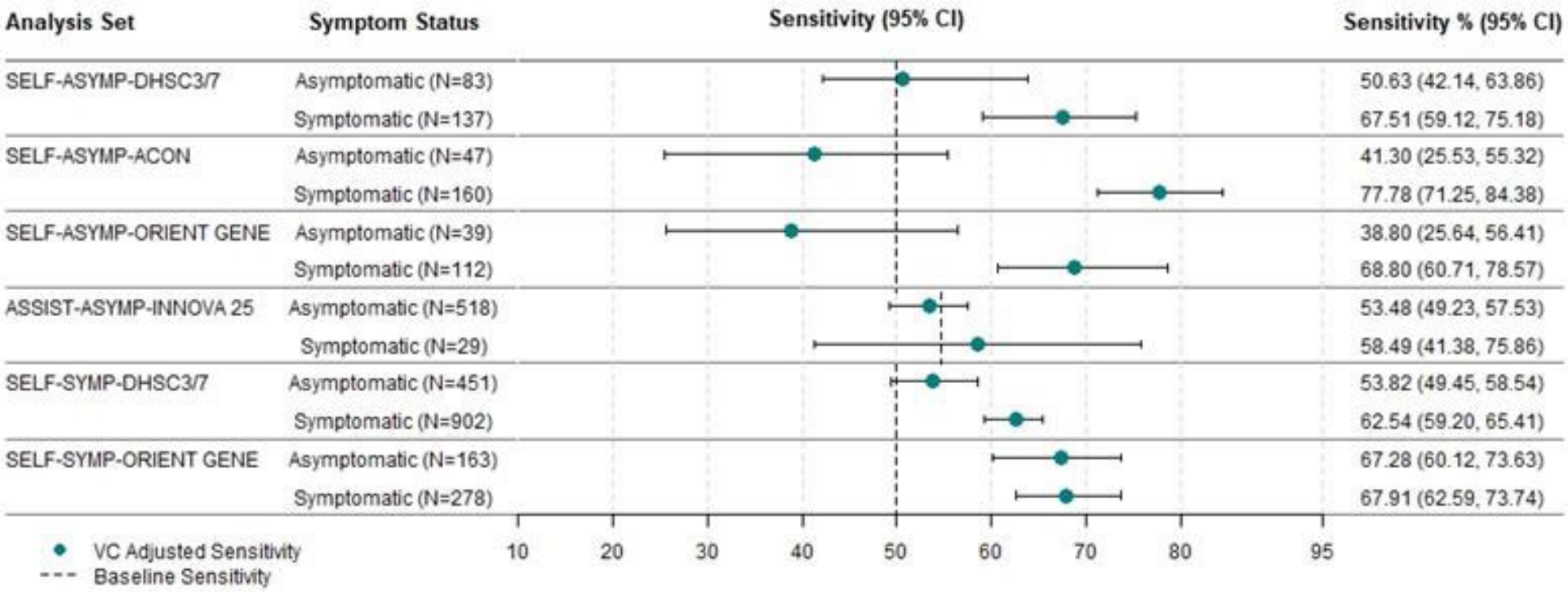


Figure 10. VC adjusted sensitivity of the symptomatic and asymptomatic within each analysis set (dashed vertical line represents the baseline)



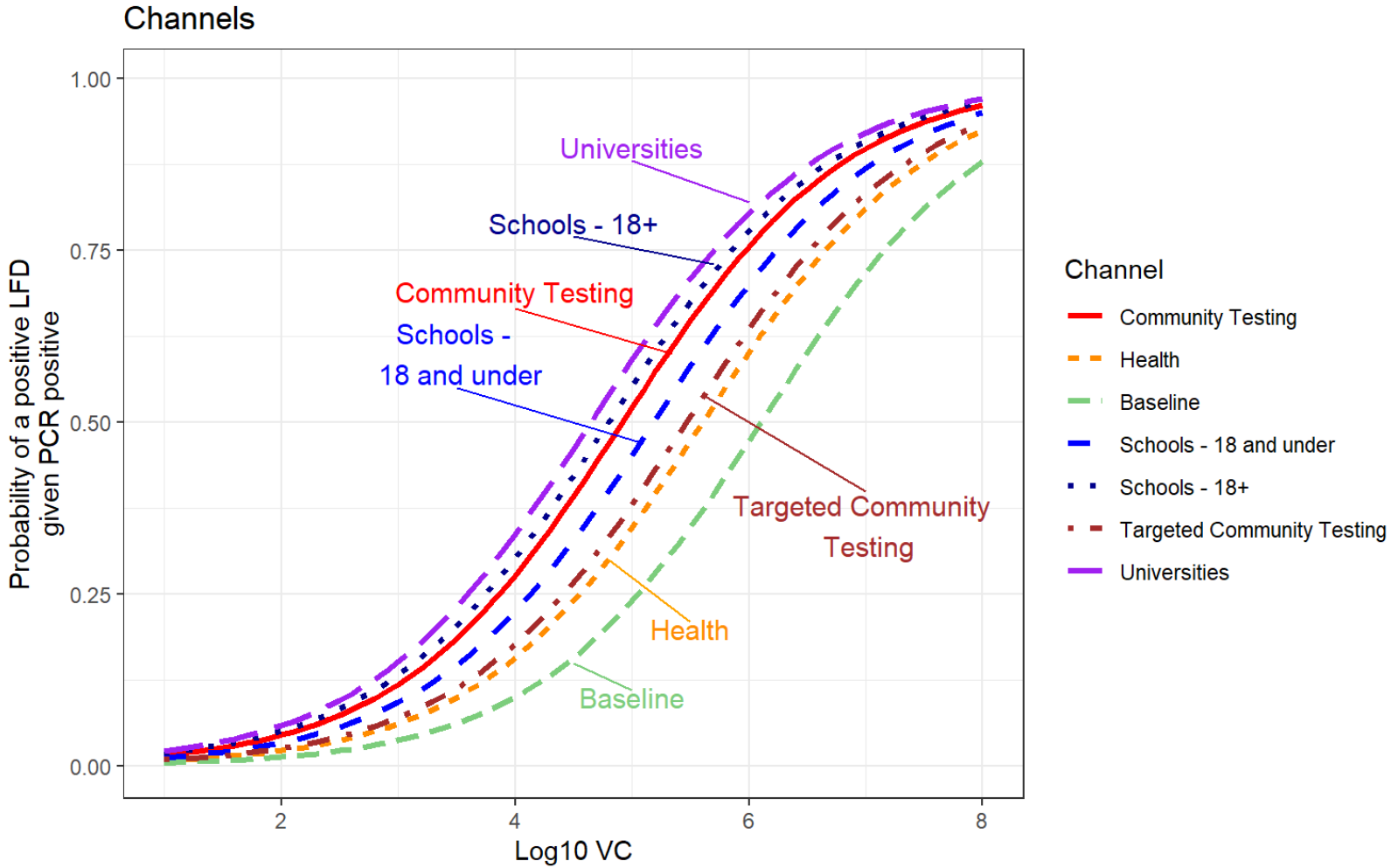
Overall, there was no strong evidence to suggest that vaccination or variant significantly affected the chances of an LFD returning a TP when VC was accounted for. The only variant which significantly increased the chances of a TP compared to Delta was B.1.177 (seen during the baseline period) (OR: 1.82 (95%CI: 1.19, 2.79)), however all other variants showed no statistically significant difference.

The analysis showed that community testing (the largest testing channel and which spanned the all-time period) had significantly higher performance than the baseline. Other testing channels showed no statistically significant difference to community testing, apart from Targeted

Community Testing (TCT) which showed a statistically significant reduction in the chances of an LFD returning a TP (OR: 0.56 (95%CI: 0.35, 0.91)). Figure 11 shows the results of the modelling for testing channels.

Although the chances of a TP in TCT was significantly lower than in community testing, sensitivity in this channel was non-inferior to baseline (difference = -2.75% (95%CI: -8.64, 3.14)) therefore performance in TCT was considered acceptable. The significant difference may suggest though that these groups need specific targeted support to ensure LFD testing performs optimally.

Figure 11. Probability of a TP for the different testing channels



Reference levels used for covariates: Device, Biotime LFD; Test setting, Self; Testing service, Asymptomatic; Symptom status, Asymptomatic; Channel, Community Testing; Vaccination status, Unvaccinated; Variant, B.1.617.2 (Delta).

The multifactorial analysis found no strong evidence that LFD kit type, test setting (self-test or assisted-test) or testing service (symptomatic or asymptomatic testing), impacted the sensitivity of LFDs and the chances of the test returning a TP, after taking into account symptom status and the other covariates (see Table 5).

Table 5. Multifactorial logistic regression results (the first listed stratum of each covariate was used as the comparator for the others)

Characteristic	Odds Ratio (OR) (95% CI)	p-value
Viral concentration	2.84 (2.65, 3.04)	**<0.001**
LFD		
Biotime (DHSC 3&7 and Innova 25s)		
Acon	1.61 (0.92, 2.85)	0.099
Orient Gene	1.12 (0.82, 1.53)	0.487
Test setting		
Self		
Assisted	1.34 (1.00, 1.80)	0.053
Testing service		
Asymptomatic		
Symptomatic	1.08 (0.76, 1.53)	0.666
Symptom status		
Asymptomatic		
Symptomatic	1.88 (1.51, 2.34)	**<0.001**
Testing channel		
Community testing		
Health	0.49 (0.18, 1.45)	0.177
Baseline evaluations	0.29 (0.19, 0.44)	**<0.001**
Schools: 18 and under	0.76 (0.28, 2.07)	0.587
Schools: >18	1.13 (0.21, 7.49)	0.889
Targeted community testing	0.56 (0.35, 0.91)	**0.019**

Characteristic	Odds Ratio (OR) (95% CI)	p-value
Universities	1.32 (0.50, 3.53)	0.571
Vaccination status		
Unvaccinated		
One dose	1.15 (0.83, 1.59)	0.413
Two doses	1.05 (0.79, 1.41)	0.720
Three doses	1.18 (0.82, 1.70)	0.374
Variant		
B.1.617.2 (Delta)		
Not specified	1.50 (1.09, 2.07)	0.013
B.1.177 (B.1.177)	1.82 (1.19, 2.79)	**0.006**
B.1.1.7 (Alpha)	0.90 (0.65, 1.27)	0.555
B.1.1.529+BA.2 (Omicron)	1.24 (0.78, 1.98)	0.365

4.3 Objective 3: Specificity

Specificity was higher than or similar to baseline in all analysis sets.

Table 6. Specificity results

Analysis set	Baseline % (95% CI)	Post-deployment % (95% CI)
SELF-ASYMP-DHSC 3&7	99.07 (98.56, 99.44)	99.79 (99.72, 99.85)
SELF-ASYMP-ACON	99.07 (98.56, 99.44)	99.25 (98.99, 99.46)
SELF-ASYMP-ORIENT GENE	99.07 (98.56, 99.44)	99.86 (99.70, 99.94)
ASSIST-ASYMP-INNOVA 25	99.58 (99.30, 99.76)	99.90 (99.86, 99.94)
SELF-SYMP-DHSC 3&7	99.07 (98.56, 99.44)	99.26 (99.03, 99.44)
SELF-SYMP-ORIENT GENE	99.07 (98.56, 99.44)	99.34 (98.31, 99.82)

4.4 Objective 4: PPV and NPV

PPV in all analysis sets was higher than, or similar to, baseline in 3 analysis sets, and lower than baseline in SELF-ASYMP-DHSC 3&7, SELF-ASYMP-ACON and ASSIST-ASYMP-INNOVA 25 (see Table 7). PPV was affected by prevalence. For the same sensitivity and specificity, PPV increases with prevalence. To make the derived PPV comparable between different prevalence settings, an adjustment is necessary (see 3.4 Objective 4: PPV and NPV). When post-deployment PPV was adjusted to the prevalence seen at baseline, these adjusted results showed a PPV in post-deployment that was higher than, or similar to, baseline for all analysis sets.

Table 7. PPV results

Analysis set	Baseline % (95% CI)	Baseline PCR-based prevalence[15] %	Post-deployment % (95% CI)	Post-deployment PCR-based prevalence %	Post-deployment adjusted for baseline prevalence % (95%CI)[16]
SELF-ASYMP-DHSC 3&7	91.77 (87.45, 94.98)	17.15	77.84 (71.33, 83.47)	1.12	98.45 (97.71, 98.98)
SELF-ASYMP-ACON	91.77 (87.45, 94.98)	17.15	77.72 (71.19, 83.38)	3.49	95.24 (93.10, 96.78)
SELF-ASYMP-ORIENT GENE	91.77 (87.45, 94.98)	17.15	93.52 (87.10, 97.35)	3.16	98.95 (97.46, 99.59)
ASSIST-ASYMP-INNOVA 25	96.16 (93.75, 97.84)	16.36	91.61 (87.95, 94.45)	2.09	98.99 (98.46, 99.44)
SELF-SYMP-DHSC 3&7	91.77 (87.45, 94.98)	17.15	94.87 (93.35, 96.14)	15.97	95.30 (93.73, 96.51)
SELF-SYMP-ORIENT GENE	91.77 (87.45, 94.98)	17.15	98.75 (96.84, 99.66)	42.20	95.75 (89.20, 98.87)

All analysis sets showed an NPV in post-deployment that was higher than at baseline, except for SELF-SYMP-ORIENT GENE which showed a lower result of 82.87% compared to a baseline of 90.54% (see Table 8). NPV decreases with an increase in prevalence. As with PPV, an

[15] PCR-based prevalence used to adjust both PPV and NPV is the percentage of positive PCR results in the whole analysis set sample, calculated as PCR+/all results.

[16] Baseline prevalence was used for the adjustment.

adjustment is necessary to enable comparison. When post-deployment NPV was adjusted to baseline prevalence, the adjusted NPV was 94.46% which was higher than the baseline NPV.

Table 8. NPV results

Analysis set	Baseline % (95% CI)	Baseline PCR-based prevalence %[9]	Post-deployment % (95% CI)	Post-deployment PCR-based prevalence %	Post-deployment adjusted for baseline prevalence % (95%CI)[10]
SELF-ASYMP-DHSC 3&7	90.54 (89.26, 91.72)	17.15	99.60 (99.50, 99.68)	1.12	93.10 (91.94, 94.21)
SELF-ASYMP-ACON	90.54 (89.26, 91.72)	17.15	99.01 (98.71, 99.25)	3.49	94.57 (93.33, 95.70)
SELF-ASYMP-ORIENT GENE	90.54 (89.26, 91.72)	17.15	98.82 (98.48, 99.10)	3.16	92.97 (91.55, 94.33)
ASSIST-ASYMP-INNOVA 25	91.78 (90.87, 92.63)	16.36	98.94 (98.81, 99.06)	2.09	91.08 (90.40, 91.77)
SELF-SYMP-DHSC 3&7	90.54 (89.26, 91.72)	17.15	95.00 (94.48, 95.48)	15.97	94.58 (94.11, 95.03)
SELF-SYMP-ORIENT GENE	90.54 (89.26, 91.72)	17.15	82.87 (79.93, 85.55)	42.20	94.46 (93.58, 95.26)

4.5 Objective 5: Void rate

All analysis sets show a void rate in post-deployment that was lower than at baseline, see Table 9.

Table 9. Void rate results

Analysis set	Baseline % (95% CI)	Post-deployment % (95% CI)
SELF-ASYMP-DHSC 3&7	2.46 (1.90, 3.14)	0.04 (0.02, 0.08)
SELF-ASYMP-ACON	2.46 (1.90, 3.14)	0.08 (0.03, 0.19)
SELF-ASYMP-ORIENT GENE	2.46 (1.90, 3.14)	0.02 (0.00, 0.11)
ASSIST-ASYMP-INNOVA 25	0.40 (0.23, 0.63)	0.16 (0.12, 0.22)
SELF-SYMP-DHSC 3&7	2.46 (1.90, 3.14)	0.05 (0.01, 0.12)
SELF-SYMP-ORIENT GENE	2.46 (1.90, 3.14)	0.19 (0.02, 0.68)

5. Conclusions

UK government policy on the use of LFDs has been driven by their ability to minimise onward transmission by detecting the most infectious cases so that they could be directed to self-isolate. This evaluation considered the performance of LFDs compared to PCRs as the reference test. PCRs detect the presence of viral RNA which alone does not indicate infectiousness. Although this evaluation cannot assess how well LFDs have achieved the stated aim of preventing onward transmission, it is known that LFDs perform best at high viral concentrations (2, 6, 7, 8) and that high viral concentration is related to infectiousness (2, 9, 10). When high viral concentration is taken as a proxy for infectiousness, a study found that LFDs typically detected over 83% of the cases that were likely to have transmitted to other people (2). To account for this relationship, UKHSA historically presented sensitivity stratified by viral concentration categories. We have further developed our approach to account for the dependency of LFD sensitivity and VC by developing methods to adjust sensitivity across different evaluations and by treating VC as a continuous variable in statistical models.

This evaluation provides evidence of how well LFDs have performed compared to expectations set prior to deployment within the NTP. From the data available, we have established that sensitivity in all analysis sets was non-inferior to baseline, and that sensitivity was higher in post-deployment than at baseline in all self-test settings. Furthermore, the relationship between viral concentration and sensitivity continued to be seen in all settings and services in which LFDs were used in the NTP and which were considered in this evaluation. Likewise, symptomatic disease independently increased the sensitivity of LFDs. Figure 8 illustrates that sensitivity increased as VC increased for both symptomatic and asymptomatic individuals. This relationship between VC and sensitivity has been a continuous feature of LFDs and can provide confidence that LFDs have detected the most infectious cases tested within the NTP. Testing behaviours changed over time with more people testing symptomatically using LFDs. Even though guidance was that PCRs should be used in such cases, the data here suggests LFDs performed well for those with symptomatic disease. Despite best efforts which enabled recruitment of over 83,000 participants to this evaluation, in certain services there were insufficient numbers of PCR+ asymptomatic cases to draw conclusions on the performance of LFDs in asymptomatic disease. Further research is warranted to elucidate differences in performance in asymptomatic and symptomatic usage.

Testing channel was found to significantly affect the chance of an LFD returning a TP for Targeted Community Testing compared to Community Testing, reinforcing the need to design the service with users in mind.

Over the course of the COVID-19 pandemic, there were substantial changes in the epidemiological context. Vaccination and variant were assessed for their impact on LFD performance and statistical analysis suggests that in general they have minor impact on the performance of LFDs. This can provide reassurance that LFDs work regardless of vaccination status and for all variants so-far detected.

No strong evidence was found that LFD kit type, test setting (self- or assisted-testing), or testing service (symptomatic or asymptomatic) impacted the sensitivity of LFDs.

Besides sensitivity, other performance outcomes were inspected. Specificity in post-deployment was higher than, or not different to, baseline in all analysis sets. All other outcomes (PPV, NPV and void rates) showed similar or improved performance than at baseline.

The evidence generated by Ongoing Evaluation shows that the LFD kits used in the National Testing Programme have been performing at or above expected levels within the settings and services in which they have been utilised. As such, the conclusion is drawn that they are robust diagnostic devices (meaning different settings, variants, vaccination statuses did not strongly impact performance) and provide sufficient diagnostic performance for use as part of the public health intervention for the COVID-19 pandemic.

5.1 Evaluation limitations

There were some limiting factors to this evaluation.

5.1.1 Data collection

The analysis was based on data collected from multiple ongoing evaluation studies with unequal numbers of subjects for the different covariates resulting in an unbalanced design and loss of statistical power.

Several factors assessed in this evaluation were correlated making it difficult to distinguish between their individual impact on LFD performance.

Some of this data, such as symptom status, was self-reported and could not be independently substantiated, and there were a number of differences between the baseline and post-deployment evaluation data collection methods. For example, baseline evaluations were conducted at symptomatic test sites meaning that the comparison with some of the asymptomatic analysis sets was between a baseline population containing 87% symptomatic cases vs a post-deployment population containing just 0.5% symptomatic cases (see Table 13).

5.1.2 Context of the pandemic

Over the course of the pandemic, the number of symptoms and types of symptoms experienced changed as new variants emerged, for example, the symptoms most commonly experienced with Omicron were different to those experienced with Alpha and Delta (4). For consistency, when categorising the data, only subjects responding with at least one of the 3 cardinal symptoms (high temperature, ageusia/anosmia (loss of taste or smell), dry continuous cough) were categorised as symptomatic. However, 20% of those who tested PCR+ and were categorised as asymptomatic reported other symptoms.

5.1.3 Reference test

Finally, there were also limitations inherent to the reference test. PCR is an imperfect comparator since it detects viral RNA rather than the antigens that LFDs were designed to detect. Additionally, the participating laboratories each used their own formula to convert CT value to VC due to differences in their respective assays, so there is a reliance on the accuracy of the formulae (see Appendix 4: qRT-PCR linearity data). It was assumed that converting Ct into VC made the quantity of virus present in the sample comparable between labs.

5.2 Implications for future planning

Lateral Flow Devices offer some crucial advantages over PCR: they do not require a laboratory or highly trained operatives, results are produced within minutes, and they are much cheaper. The results presented in this report agree with previous studies that have shown LFD performance increases as viral concentration increases, suggesting LFDs are good at identifying the most infectious cases (2, 9, 10).

Giving members of the public the ability to quickly and easily identify that they are likely to be infectious with a high degree of confidence is a powerful public health tool. LFD deployment enables these individuals to self-isolate and break the chain of transmission, assisting in the public health response to the SARS-CoV-2 pandemic. Findings suggest that implementation has been relatively robust in most settings and that, for the most part, individuals have been able to successfully self-test. When planning future interventions, it is appropriate to make the assumption that people can use these tests in different settings with different instructions competently. The findings showed lower performance in some disproportionately impacted and under-served groups which suggest that in some instances a more tailored implementation strategy may be required.

Prior to deploying LFDs as part of the National Testing Programme, in vitro testing took place in UKHSA Porton Down laboratories. The findings here suggest that testing and quality assurance activities have been broadly effective in identifying LFDs which met the requirements for deployment and that this process can be relied upon in future.

6. References

1. Department of Health and Social Care. 'Asymptomatic testing for SARSCoV-2 using antigen-detecting lateral flow devices: Evidence from performance data October 2020 – May 2021' Department of Health and Social Care 2021

2. Lee LYW, Rozmanowski S, Pang M, Charlett A, Anderson C and others. 'SARS-CoV-2 infectivity by viral load, S gene variants and demographic factors and the utility of lateral flow devices to prevent transmission' Clinical Infectious Disease: Volume 74, Issue 3

3. Pickering S, Batra MD, Merrick B, Snell LB, Nebbia G and others. 'Comparative performance of SARS-CoV-2 lateral flow antigen tests and association with detection of infectious virus in clinical specimens: a single-centre laboratory evaluation study' The Lancet 2021: Volume 2, Issue 9, E461-E471

4. Whitaker M, Elliott J, Bodinier B, Barclay W, Ward H and others. 'Variant-specific symptoms of COVID-19 among 1,542,510 people in England' medRxiv 23 May 2022

5. Machin D, Campbell MJ, Tan SB, Tan SH. 'Sample Size Tables for Clinical Studies' 2008, 3rd edition, London, Wiley Blackwell

6. Scohy A, Anantharajah A, Bodéus M, Kabamba-Mukadi B, Verroken A, Rodriguez-Villalobos H. 'Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis' Journal of Clinical Virology 2020: Volume 129, 104455

7. Lambert-Niclot S, Cuffel A, Le Pape S, Vauloup-Fellous C, Morand-Joubert L and others. 'Evaluation of a Rapid Diagnostic Assay for Detection of SARS-CoV-2 Antigen in Nasopharyngeal Swabs' Journal of Clinical Microbiology 2020: Volume 58

8. Mertens P, De Vos N, Martiny D, Jassoy C, Mirazimi A and others. 'Development and Potential Usefulness of the COVID-19 Ag Respi-Strip Diagnostic Assay in a Pandemic Context' Frontiers in Medicine 2020: volume 7

9. La Scola B, Le Bideau M, Andreani J, Hoang VT, Grimaldier C and others. 'Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards' European Journal of Clinical. Microbiology and Infectious Disease 2020: volume 39

10. Van Kampen JJ, van de Vijver DA, Fraaij PL, Haagmans BL, Lamers MM and others. 'Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (COVID-19): Duration and key determinants' medRxiv 2020

Appendices

Appendix 1: Glossary

Evaluation and analysis terms

Analysis set

The performance of LFD kits was evaluated in the different test settings (self-test and assisted test) and services (symptomatic and asymptomatic) in which they were used. These are referred to as analysis sets. See Table 10 for the list of analysis sets.

Baseline

The evaluation studies which took place prior to the deployment of LFDs in real-world settings. These studies were provided as part of the technical document that was submitted as part of the original EUA application for the DHSC 3&7 self-test kit and against which the post-deployment observed results in this evaluation are compared. The baseline studies included LFD001 and LFD002. In LFD001, a trained individual swabbed the participant and interpreted the test (assisted testing). In LFD002, users self-swabbed and then performed and interpreted the test without assistance (self-testing). LFD001 was considered the baseline performance for assisted testing (analysis set ASSIST-ASYMP-INNOVA 25), and LFD002 was considered the baseline performance for self-testing (all other analysis sets). See Appendix 3: Methodology of the baseline studies for details of the baseline service evaluation methodology.

Cycle threshold

The number of cycles required to detect the presence of viral RNA by qRT-PCR.

Dual-test

An LFD and qRT-PCR taken concurrently, and collected specifically for the purposes of ongoing evaluation

Ongoing Evaluation (OE)

Ongoing Evaluation is the approach used to monitor the performance of testing services overseen by UKHSA to ensure that they continue to be safe, effective, and appropriate as epidemiological factors change over time.

Post-deployment

Post-deployment refers to the time period covered in this report during which LFDs were being used in real-world settings. This time period can be further broken down into the reporting periods below.

Sensitivity

Observed sensitivity: the sensitivity observed in the post-deployment period reported here.

VC adjusted sensitivity: the sensitivity in the post-deployment period after adjustment for VC. Specifically, the distribution of viral concentration seen in post-deployment samples was adjusted to match the distribution of viral concentration seen at baseline (using viral concentration as a categorical variable as outlined in 3.1). In this way, the effect of VC on difference in sensitivity between post-deployment and baseline was minimised and the 2 could be compared.

Service evaluation

Field test to assess the suitability of devices in the settings in which they are to be used. The results of 2 of these Service Evaluations (LFD001 and LFD002) set the benchmark upon which the MHRA Exceptional Use Authorisation (EUA) was granted for the use of Biotime LFDs.

Services

Testing for SARS-CoV-2 within NHS Test and Trace falls into asymptomatic and symptomatic testing categories, referred to in this report as “services”.

Settings

Refers to whether the method of testing was self or assisted. See 1.1 Testing policy context for details.

Symptom status

Subjects self-reported their symptoms and were categorised as either symptomatic or asymptomatic in the following ways:

- symptomatic included any of the 3 cardinal symptoms – high temperature, ageusia/anosmia (loss of taste or smell), dry continuous cough
- asymptomatic included no symptoms or any other symptoms not listed above

Reporting periods

Report periods relate to discrete time periods during which participants were recruited and analysis has been performed. For the purposes of this report, these were:

- Period 1 – 8 November 2020 to 21 May 2021
- Period 2 – 22 May 2021 to 21 September 2021
- Period 3 – 22 September 2021 to 21 March 2022
- All Time – 8 November 2020 to 21 March 2022

Testing channel

Community testing – asymptomatic testing offered to the general public through:

- home testing
- LFD Direct – LFD kits sent to the general public on request
- pharmacies – LFD kits collected on request from pharmacies by the general public
- Citywide Testing (Liverpool) – asymptomatic testing of the general population in Liverpool at a time of high prevalence in the city

Targeted Community Testing (TCT) – asymptomatic testing of the general population in areas of high prevalence.

Public and private industry – asymptomatic testing in workplaces in the public and private sector.

Universities and schools (primary and secondary) – twice weekly asymptomatic testing of both students and staff. For analysis purposes, universities were considered a separate testing channel to schools, and schools were broken into age 18 and under (presumed to be pupils), and over 18s (presumed to be staff).

Health – asymptomatic testing of staff in healthcare settings, including NHS, independent healthcare providers, frontline healthcare providers.

Surge testing – increased testing aimed at asymptomatic subjects and enhanced contact tracing in specific locations.

In Person Testing Channel (IPTC) – onsite symptomatic testing at mobile testing units in Scotland.

Regional or local test sites (RTS/LTS) – onsite symptomatic testing.

Variant

Variants assessed within this evaluation refer to major lineages of SARS-CoV-2 during the time period of this report:

- B.1.1.7 (Alpha) variant – Pango lineage B.1.1.7 first documented in the United Kingdom in September 2020 and designated 18 December 2020
- B.1.617.2 (Delta) variant – Pango lineage B.1.617.2 first documented in India in October 2020 and designated 4 April 2021
- B.1.177 (no WHO name) – Pango lineage B.1.177 first documented in Spain in June 2020 and designated December 2020
- B.1.1.529 (Omicron) – Pango lineage B.1.1.529 first documented in South Africa in November 2021 and designated 26 November 2021

Viral Concentration (VC)

The amount of virus present in the viral transport medium calculated by converting the Ct value from qRT-PCR into a viral concentration using the laboratory's specific conversion formula (see 'Appendix 4: qRT-PCR linearity data'). This is a proxy for the amount of virus present in a person's nasal/oral cavity rather than a direct measure. It depends both on the quality of the swabbing technique and the efficiency of the release of the virus from the swab into the transport medium.

Statistical definitions

Likelihood Ratio Test (LRT)

The Likelihood-Ratio test (sometimes called the likelihood-ratio chi-squared test) is a test which compares how well 2 alternative statistical models fit the observed data. In this report, LRT was used to compare logistic regression models with and without a specific covariate. If the likelihood of the model with the covariate is greater and the LRT shows significance ($p<0.05$), it indicates that the model with the covariate fits the data better and thus the covariate helps to predict whether an LFD is a true positive or a false negative.

Odds Ratio (OR)

The OR represents the odds that an outcome will occur given a particular condition (such as the LFD test was carried out in a specific test channel or setting) compared to the odds of the outcome occurring in the absence of that condition.

Appendix 2: Analysis sets

Table 10. Overview of the 8 analysis sets of which 6 contained enough samples for analysis to be conducted (greyed rows show the analysis sets that did not contain enough samples for analysis)

Analysis set	Test setting (service and test type)	LFD kits used in this setting	Example of testing services covered	Baseline study
SELF-ASYMP-DHSC 3&7	Self-testing in asymptomatic services	DHSC 3&7	LFD home testing (for example LFD Direct or pharmacies), schools, health	LFD002
SELF-ASYMP-ACON	Self-testing in asymptomatic services	Acon	LFD home testing (for example LFD Direct or pharmacies), schools, health	LFD002
SELF-ASYMP-SURESCREEN (Did not contain enough samples for analysis)	Self-testing in asymptomatic services	Surescreen	LFD home testing (for example LFD Direct or pharmacies), schools, health	LFD002
SELF-ASYMP-ORIENT GENE	Self-testing in asymptomatic services	Orient Gene	LFD home testing (for example LFD Direct or pharmacies), schools, health	LFD002
ASSIST-ASYMP-INNOVA 25	Assisted testing in asymptomatic services	Innova 25	Citywide, targeted community, public and private industry, universities	LFD001
SELF-SYMP-DHSC 3&7	Self-testing in symptomatic services and surge testing	DHSC 3&7	LTS/RTS, IPTC	LFD002
SELF-SYMP-ACON (Did not contain enough	Self-testing in symptomatic services and surge testing	Acon	LTS/RTS, IPTC	LFD002

Analysis set	Test setting (service and test type)	LFD kits used in this setting	Example of testing services covered	Baseline study
samples for analysis)				
SELF-SYMP-ORIENT GENE	Self-testing in symptomatic services and surge testing	Orient Gene	LTS/RTS, IPTC	LFD002

Appendix 3: Methodology of the baseline studies

In LFD001, symptomatic participants arrived at a Regional or Local Test Site (RTS/LTS) for the purposes of receiving a standard of care diagnostic PCR test. On arrival, they were invited to take part in the study and if they consented to participate, consented for their data to be used, understood that their LFD result was indicative, and were over 18, then they were enrolled in the study and their details were registered. To perform the PCR test, a trained individual swabbed the participants throat and nose. The barcode was then recorded, and the test was sent for processing. To perform the LFD test, a trained individual swabbed the participant's throat and nose. The sample was then transferred to a separate location and applied to the LFD cassette according to the manufacturer's Instructions for Use (IFU). The trained individual then read the result and positive results were confirmed by a second trained individual. The result was then logged on the system. After both swabs were taken, the subject was free to leave the site and received a text with their LFD result. The PCR result was communicated to the participant as per standard operating procedures at that time.

For LFD002, symptomatic participants arrived at Regional or Local Test Site (RTS/LTS) in order to receive a diagnostic test. On arrival, they were invited to take part in the study and if they consented to participate, consented to their data to be used, understood that their LFD result was indicative, and were willing to perform the test themselves and wait for the result, then they were enrolled in the study and their details were registered. For children under 12, the parent or guardian administered the swab on their behalf. Upon enrolment into the study, participants were expected to read and follow the LFD instructions and then perform the test themselves, swabbing the throat and a single nostril. Observers did not provide guidance as this would not be representative of 'real-world' deployment. After performing the LFD test, the staff member and participant started a timer for 30 minutes checking after 5 minutes that the control line for the LFD had appeared. Whilst the timer was running, the participant swabbed their nose and throat for the PCR test which was sent to a laboratory for processing. After exactly 30 minutes, the participant visually inspected the LFD and read the result. The staff member took a photo of the LFD and barcode for quality assurance. The PCR result was communicated to the participant as per standard operating procedures at that time.

Appendix 4: qRT PCR linearity data

qRT-PCR assays and protocols differed between processing laboratories. Ct values were therefore not directly comparable between labs and were converted to VC to be comparable. Each laboratory provided a bespoke conversion formula that was specific to their assay. The conversion formulae are presented in Table 11 below.

Table 11:qRT-PCR processing lab and gene target specific cycle threshold to viral concentration conversion formulae

Lab	Ct to viral concentration conversion formulae				
	ORF1ab	N-Gene	S-Gene	E-Gene	RdRp
Randox Assay AQP1-A	y = -0.3065x + 12.477	-	-	y = -0.3103x + 12.85	-
Randox Assay PE	y = -0.2909x + 11.921	y = -0.3355x + 13.334	-	-	-
Alderly Park	y = -0.3035x + 11.599	y = -0.312x + 11.881	-	-	-
Glasgow	y = -0.305x + 11.372	y = -0.3096x + 11.449	y = -0.2894x + 11.221	-	-
Milton Keynes	y = -0.3181x + 11.859	y = -0.3241x + 12.119	y = -0.3641x +13.372	-	-
HSL UCL	-	y = -0.2915x + 14.049	-	-	-
Newcastle	y= -0.2831x + 11.354	y = -0.3161x + 11.882	y = -0.3857x + 14.003	-	-
Plymouth	y = -0.2971x + 12.649	y = -0.3441x + 14.637	-	-	-

Appendix 5: Variant assignment method

Variant major lineages were assigned to positive samples using the steps below:

- sequencing – for samples that were sequenced the major and sub-lineage were assigned through sequencing
- genotyping – for samples that were genotyped the major lineage was assigned through genotyping
- positive samples that had not been sequenced or genotyped had the major lineage imputed using a variant assignment algorithm which uses the following GeoTemporal (GeoTemp) method for samples from England:
    - GeoTemp method uses the postcode of the participant and their PCR test date to query all samples within the lower tier local authority (LTLA) that week that have been sequenced (accessed via the Sanger Genomics Surveillance lineages by LTLA and week). This provides the most likely major lineage by LTLA for any given week. If more than 50% of samples are identified as being from a major lineage that major lineage is assigned. If the percentage is below 50% the major lineage is recorded as ‘unknown’

For Scotland, Wales and Northern Ireland, COG-UK partner (The Centre for Genomic Pathogen Surveillance) maintain a Microreact website which permits continuous evaluation of the lineages circulating in the UK. This was integrated within the Open Source Intelligence (OSINT) tool to extract the major lineage and sub-lineage for the samples in the same way as detailed above.

At the 50% threshold detailed above, a 97.72% (95%CI: 97.69, 97.75) concordance rate was found between the algorithm predictions and Second-Generation Surveillance System (SGSS) Provisionally Genotyped data.

Appendix 6: Participant demographics and clinical characteristics

The tables below show the demographic breakdown and clinical characteristics of each analysis set and the SARS-CoV-2 strain detected by PCR:

Table 12. Participant demographics and clinical characteristics of self-test analysis sets

Variable	Baseline	SELF-ASYMP-DHSC 3&7	SELF-ASYMP-ACON	SELF-ASYMP-ORIENT GENE	SELF-SYMP-DHSC 3&7		SELF-SYMP-ORIENT GENE
Age					**RTS / LTS**	**IPTC**	
Mean	37.66	40.98	38.35	30.21	38.03	40.06	39.45
Median	36	41	39	25	37	40.79	38.45
Mode	N/A	N/A	N/A	N/A	N/A	50-54	40-44
Std	13.44	16.20	17.40	18.49	12.83	16.73	16.65
Count	2,357	10,551	2,088	2,307	7,954	511	1,045
Range[17]	12 - 80	1 - 88	1 - 82	2 - 85	1 - 87	2 - 92	2 - 92
Missing	88	10,373	3,566	2,697	0	33	69
Unknown	28	58	275	26	7	N/A	N/A
Gender							
Female	1,312 (53.05%)	6,569 (31.31%)	1,399 (23.6%)	1,366 (27.16%)	4,968 (58.64%)		605 (57.89%)
Male	1,080 (43.67%)	3,957 (18.86%)	895 (15.10%)	934 (18.57%)	3,461 (40.85%)		369 (35.31%)
No Information	81 (3.28%)	10,456 (49.83%)	3,635 (61.31%)	2,730 (54.27%)	43 (0.51%)		71 (6.79%)

[17] Age range starting at 1 or 2 is likely a database/transcription error.

Variable	Baseline	SELF-ASYMP-DHSC 3&7	SELF-ASYMP-ACON	SELF-ASYMP-ORIENT GENE	SELF-SYMP-DHSC 3&7	SELF-SYMP-ORIENT GENE
Symptoms						
Asymptomatic	293 (11.85%)	19,825 (94.49%)	4,988 (84.13%)	3,618 (71.93%)	4,168 (49.20%)	519 (49.67%)
Symptomatic	2,092 (84.59%)	951 (4.53%)	941 (15.87%)	674 (13.40%)	4,304 (50.80%)	526 (50.33%)
No information	88 (3.56%)	206 (0.98%)	0	738 (14.67%)	0	0
Variant						
B.1.1.7 (Alpha)	19 (4.48%)	1 (0.43%)	0	0	326 (24.09%)	0
B.1.177 (B.1.177)	321 (75.71%)	0	0	0	0	0
B.1.617.2 (Delta)	0	122 (51.91%)	32 (15.46%)	48 (30.19%)	767 (56.69%)	N/A
B.1.1.529+BA.2 (Omicron)	0	14 (5.96%)	36 (17.39%)	42 (26.42%)	14 (1.03%)	29 (6.58%)
No information	84 (19.81%)	98 (41.70%)	139 (67.15%)	69 (43.40%)	245 (18.11%)	412 (93.42%)
P.1 (Gamma)	0	N/A	N/A	N/A	1 (0.07%)	N/A
Vaccination						
Unvaccinated	2,473 (100%)	3,002 (14.31%)	497 (8.38%)	341 (6.78%)	3,335 (39.36%)	78 (7.46%)
One dose	0	4,606 (21.95%)	361 (6.09%)	322 (6.40%)	2,431 (28.69%)	22 (2.11%)
Two doses	0	12,767 (60.85%)	2,709 (45.69%)	3,111 (61.85%)	2,285 (26.97%)	130 (12.44%)

Variable	Baseline	SELF-ASYMP-DHSC 3&7	SELF-ASYMP-ACON	SELF-ASYMP-ORIENT GENE	SELF-SYMP-DHSC 3&7	SELF-SYMP-ORIENT GENE
Three doses	0	463 (2.21%)	2,343 (39.52%)	511 (10.16%)	418 (4.93%)	814 (77.89%)
No Information	0	144 (0.69%)	19 (0.32%)	745 (14.81%)	3 (0.04%)	1 (0.10%)
Testing channel						
Pre-Deployment Evaluation	2,473 (100%)	0	0	0	0	0
Health	0	815 (3.88%)	396 (6.68%)	376 (7.47%)	N/A	N/A
LFD Direct	0	9,538 (45.45%)	4,282 (72.22%)	2,943 (58.51%)	N/A	N/A
Not Registered to a Service	0	12 (0.06%)	15 (0.25%)	19 (0.38%)	N/A	N/A
Pharmacy	0	32 (0.15%)	68 (1.15%)	258 (5.13%)	N/A	N/A
Schools – Primary	0	2,125 (10.13%)	15 (0.25%)	3 (0.06%)	N/A	N/A
Schools – Secondary	0	8,011 (38.18%)	891 (15.03%)	1,352 (26.88%)	N/A	N/A
Targeted Community Testing (TCT)	0	15 (0.07%)	43 (0.73%)	45 (0.89%)	N/A	N/A
Universities	0	434 (2.07%)	219 (3.69%)	34 (0.68%)	N/A	N/A
IPTC	N/A	N/A	N/A	N/A	511 (6.03%)	1045 (100%)
RTS/LTS LFD	N/A	N/A	N/A	N/A	7,961 (93.97%)	N/A

Variable	Baseline	SELF-ASYMP-DHSC 3&7	SELF-ASYMP-ACON	SELF-ASYMP-ORIENT GENE	SELF-SYMP-DHSC 3&7	SELF-SYMP-ORIENT GENE
Viral concentration category (dc/ml)						
>10M	22 (5.19%)	34 (14.47%)	47 (22.71%)	20 (12.58%)	225 (16.63%)	12 (2.72%)
1M-10M	96 (22.64%)	61 (25.96%)	42 (20.29%)	35 (22.01%)	454 (33.56%)	95 (21.54%)
100K-1M	104 (24.53%)	44 (18.72%)	38 (18.36%)	38 (23.90%)	295 (21.8%)	150 (34.01%)
10K-100K	85 (20.05%)	24 (10.21%)	30 (14.49%)	24 (15.09%)	160 (11.83%)	88 (19.95%)
1K-10K	50 (11.79%)	37 (15.74%)	15 (7.25%)	12 (7.55%)	97 (7.17%)	48 (10.88%)
100-1K	36 (8.49%)	18 (7.66%)	19 (9.18%)	18 (11.32%)	71 (5.52%)	30 (6.8%)
<100	31 (7.31%)	17 (7.23%)	16 (7.73%)	12 (7.55%)	51 (3.77%)	18 (4.08%)

Table 13. ASSIST-ASYMP-INNOVA 25 – participant demographics and clinical characteristics

Variable	Baseline	Post-deployment
Age		
Mean	37.90	42.73
Median	36	41
Std	12.48	17.58
Count	4,224	26,421
Range[18]	1 - 86	1 - 90
Missing	0	680
Unknown	1	16
Gender		
Female	2,321 (54.93%)	12,326 (45.45%)
Male	1,904 (45.07%)	14,131 (52.11%)
No Information	0	660 (2.43%)
Symptoms		
Asymptomatic	524 (12.4%)	24,319 (89.68%)
Symptomatic	3,696 (87.48%)	143 (0.53%)
No information	5 (0.12%)	2,655 (9.79%)
Variant		
B.1.1.7 (Alpha)	176 (25.47%)	12 (2.11%)
B.1.177 (B.1.177)	315 (45.59%)	0
B.1.617.2 (Delta)	0	396 (69.72%)
B.1.1.529 (Omicron)	0	70 (12.32%)
No information	178 (25.76%)	89 (15.67%)
B.1.235 (B.1.235)	0	1 (0.18%)

[18] Age range starting at 1 is likely a database/transcription error.

Variable	Baseline	Post-deployment
B.1.1.311 (B.1.1.311)	18 (2.6%)	0
B.1.1.37 (B.1.1.37)	4 (0.58%)	0
Vaccination		
Unvaccinated	4,225 (100%)	12,775 (47.11%)
One dose	0	5,046 (18.61%)
Two doses	0	8,079 (29.79%)
Three doses	0	680 (2.51%)
No information	0	537 (1.98%)
Testing channel		
Pre-deployment evaluation	4,225 (100%)	0
City-wide testing	0	5,534 (20.41%)
Private industry	0	3,720 (13.72%)
Public industry	0	1,980 (7.30%)
Targeted Community Testing (TCT)	0	10,114 (37.30%)
Universities	0	5,769 (21.27%)
Viral concentration category (dc/ml)		
>10M	86 (12.45%)	63 (11.09%)
1M-10M	147 (21.27%)	114 (20.07%)
100K-1M	151 (21.85%)	110 (19.37%)
10K-100K	125 (18.09%)	83 (14.61%)
1K-10K	72 (10.42%)	84 (14.79%)
100-1K	45 (6.51%)	66 (11.62%)
<100	65 (9.41%)	48 (8.45%)

Appendix 7: Analysis outputs

Sensitivity results

Table 14. Sensitivity of each analysis set (observed and VC adjusted)

Analysis set	Baseline % (95% CI)	Post-deployment sensitivity type	Post-deployment sensitivity % (95%CI)	Difference (post-deployment minus baseline) % (95%CI)	Conclusion
SELF-ASYMP-DHSC 3&7	50.00 (45.14, 54.86)	Observed	64.26 (57.77, 70.38)	14.26 (6.50, 22.02)	Non-inferior
		VC adjusted	60.15 (53.19, 65.96)	10.15 (2.29, 18.01)	Non-inferior
SELF-ASYMP-ACON	50.00 (45.14, 54.86)	Observed	72.46 (65.84, 78.43)	22.46 (14.73, 30.19)	Non-inferior
		VC adjusted	69.01 (62.31, 74.88)	19.01 (11.11, 26.91)	Non-inferior
SELF-ASYMP-ORIENT GENE	50.00 (45.14, 54.86)	Observed	63.52 (55.53, 71.00)	13.52 (4.65, 22.39)	Non-inferior
		VC adjusted	62.82 (55.97, 70.44)	12.82 (3.93, 21.71)	Non-inferior
ASSIST-ASYMP-INNOVA 25	54.41 (50.62, 58.17)	Observed	50.00 (45.81, 54.19)	-4.41 (-9.95, 1.13)	Non-inferior
		VC adjusted	53.78 (49.12, 57.57)	-0.63, (-6.16, 4.90)	Non-inferior
SELF-SYMP-DHSC 3&7	50.00 (45.14, 54.86)	Observed	72.51 (70.04, 74.87)	22.51 (17.19, 27.83)	Non-inferior
		VC adjusted	59.48	9.48	Non-inferior

Analysis set	Baseline % (95% CI)	Post-deployment sensitivity type	Post-deployment sensitivity % (95%CI)	Difference (post-deployment minus baseline) % (95%CI)	Conclusion
			(56.76, 61.94)	(4.05, 14.91)	
SELF-SYMP-ORIENT GENE	50.00 (45.14, 54.86)	Observed	71.88 (67.44, 76.03)	21.88 (15.54, 28.22)	Non-inferior
		VC adjusted	67.91 (63.49, 72.56)	17.91 (11.46, 24.36)	Non-inferior

Symptom status results

Table 15. Count of asymptomatic and symptomatic individuals who tested PCR+ within each analysis set, and the proportion of each analysis set that tested PCR+ that were asymptomatic

Analysis set	Symptom status	Number of PCR+ results	Proportion of PCR+ that are asymptomatic %
SELF-ASYMP-DHSC 3&7	Asymptomatic	83	35.32
	Symptomatic	137	
	No Information	15	
SELF-ASYMP-ACON	Asymptomatic	47	22.71
	Symptomatic	160	
SELF-ASYMP-ORIENT GENE	Asymptomatic	39	24.53
	Symptomatic	112	
	No Information	8	
ASSIST-ASYMP-INNOVA 25	Asymptomatic	518	91.20
	Symptomatic	29	
	No Information	21	

Analysis set	Symptom status	Number of PCR+ results	Proportion of PCR+ that are asymptomatic %
SELF-SYMP-DHSC 3&7	Asymptomatic	451	33.33
	Symptomatic	902	
SELF-SYMP-ORIENT GENE	Asymptomatic	163	36.96
	Symptomatic	278	

Figure 12. Distribution of VC by symptom status for all samples used in the multifactorial analysis

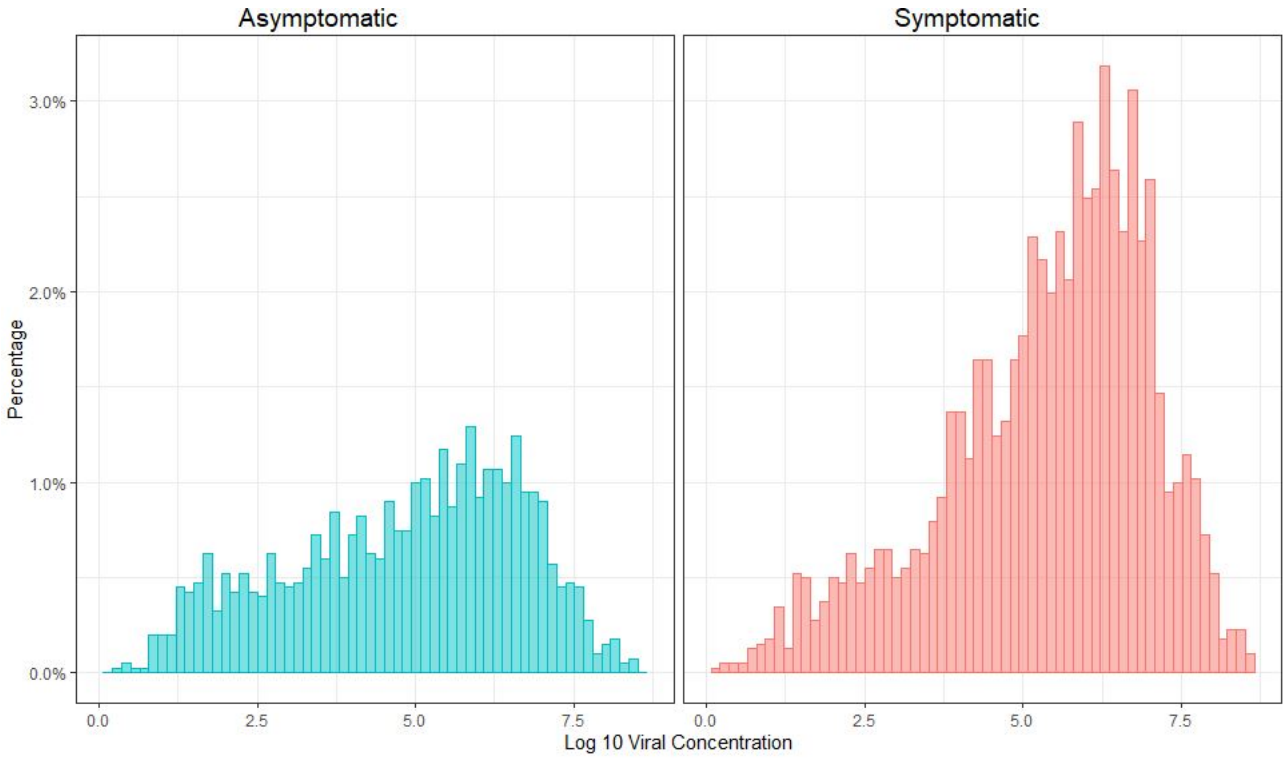


Table 16. Sensitivity of the symptomatic and asymptomatic within each analysis set

Analysis set	Overall baseline (%, 95%CI)	Symptom status	Baseline (%, 95%CI)	Observed sensitivity (%, 95%CI)	VC adjusted sensitivity (%, 95%CI)
SELF-ASYMP-DHSC 3&7	50.00 (45.14, 54.86)	Asymptomatic (n=83)	34.78 (16.38, 57.27)	32.53 (22.65, 43.70)	50.63 (42.14, 63.86)
		Symptomatic (n=137)	51.53 (46.46, 56.58)	82.48 (75.06, 88.44)	67.51 (59.12, 75.18)
SELF-ASYMP-ACON	50.00 (45.14, 54.86)	Asymptomatic (n=47)	34.78 (16.38, 57.27)	29.79 (17.34, 44.89)	41.30 (25.53, 55.32)
		Symptomatic (n=160)	51.53 (46.46, 56.58)	85.00 (78.51, 90.15)	77.78 (71.25, 84.38)
SELF-ASYMP-ORIENT GENE	50.00 (45.14, 54.86)	Asymptomatic (n=39)	34.78 (16.38, 57.27)	33.33 (19.09, 50.22)	38.80 (25.64, 56.41)
		Symptomatic (n=112)	51.53 (46.46, 56.58)	76.79 (67.86, 84.24)	68.80 (60.71, 78.57)
ASSIST-ASYMP-INNOVA 25	54.41 (50.62, 58.17)	Asymptomatic (n=518)	41.03 (25.57, 57.90)	49.23 (44.84, 53.62)	53.48 (49.23, 57.53)
		Symptomatic (n= 29)	55.21 (51.30, 59.08)	62.07 (42.26, 79.31)	58.49 (41.38, 75.86)
SELF-SYMP-DHSC 3&7	50.00 (45.14, 54.86)	Asymptomatic (n= 451)	34.78 (16.38, 57.27)	62.08 (57.43, 66.58)	53.82 (49.22 - 58.31)
		Symptomatic (n=902)	51.53 (46.46, 56.58)	77.72 (74.86, 80.39)	62.54 (59.20 - 65.19)

Analysis set	Overall baseline (%, 95%CI)	Symptom status	Baseline (%, 95%CI)	Observed sensitivity (%, 95%CI)	VC adjusted sensitivity (%, 95%CI)
SELF-SYMP-ORIENT GENE	50.00 (45.14, 54.86)	Asymptomatic (n=163)	34.78 (16.38, 57.27)	63.80 (55.92, 71.17)	67.28 (60.12 - 74.23)
		Symptomatic (n=278)	51.53 (46.46, 56.58)	76.62 (71.19, 81.47)	67.91 (62.59 - 73.38)

Table 17. Difference between VC adjusted sensitivity for asymptomatic samples in post-deployment and overall sensitivity at baseline (for the analysis sets in which VC adjusted sensitivity is below baseline)

Analysis set	Baseline sensitivity % (95%CI)	Post-deployment VC adjusted asymptomatic sensitivity % (95%CI)	Difference % (95%CI)
SELF-ASYMP-ACON	50.00 (45.14, 54.86)	41.30 (25.53, 55.32)	-8.70 (-23.56, 6.16)
SELF-ASYMP-ORIENT GENE	50.00 (45.14, 54.86)	38.80 (25.64, 56.41)	-11.20 (-27.22, 4.82)
ASSIST-ASYMP-INNOVA 25	54.41 (50.62, 58.17)	53.48 (49.23, 57.53)	-0.93 (-6.61, 4.75)

Probability of a True Positive

Table 18. Unifactorial analysis: probability of a TP (logistic regression included VC as an independent variable)

Covariate (Reference level)[19]	Analysis set	LRT significance (adjusted p value)	Contrast[20]	Odds Ratio (95%CI; p-value)
Vaccination status (Unvaccinated)	SELF-ASYMP-DHSC 3&7	1	1 dose vs unvaccinated	0.89 (0.30, 2.65) p = 0.836
			2 doses vs unvaccinated	1.45 (0.60, 3.57) p = 0.413
			3 doses vs unvaccinated	4.38 (1.04, 23.99) p = 0.06
	SELF-ASYMP-ACON	1	1 dose vs unvaccinated	3.31 (0.28, 90.7) p = 0.391
			2 doses vs unvaccinated	0.79 (0.27, 2.22) p = 0.656
			3 doses vs unvaccinated	1.07 (0.36, 3.08) p = 0.904
	SELF-ASYMP-ORIENT GENE	1	1 dose vs unvaccinated	0.66 (0.08, 7.16) p = 0.714
			2 doses vs unvaccinated	0.38 (0.12, 1.08) p = 0.079
			3 doses vs unvaccinated	0.17 (0.04, 0.66) p = 0.013
	ASSIST-ASYMP-INNOVA 25	1	1 dose vs unvaccinated	0.93 (0.51, 1.72) p = 0.827
			2 doses vs unvaccinated	0.87 (0.53, 1.42) p = 0.581

[19] If listed reference level did not exist in an analysis set, an alternative reference level was chosen

[20] Contrasts were only derived for covariate levels with 10 or more samples.

			3 doses vs unvaccinated	1.14 (0.40, 3.39) p = 0.814
	SELF-SYMP-DHSC 3&7	<0.001	1 dose vs unvaccinated	1.24 (0.82, 1.89) p = 0.316
			2 doses vs unvaccinated	1.43 (0.94, 2.18) p = 0.097
			3 doses vs unvaccinated	4.09 (2.45, 6.97) p <0.001
	SELF-SYMP-ORIENT GENE	1	1 dose vs unvaccinated	2.97 (0.39, 34.75) p = 0.329
			2 doses vs unvaccinated	0.57 (0.16, 1.87) p = 0.37
			3 doses vs unvaccinated	0.43 (0.14, 1.16) p = 0.117
Symptom status (Asymptomatic)	SELF-ASYMP-DHSC 3&7	0.134	Symptomatic vs asymptomatic	3.27 (1.47, 7.31) p = 0.004
	SELF-ASYMP-ACON	0.004	Symptomatic vs asymptomatic	7.22 (3.01, 18.11) p = <0.001
	SELF-ASYMP-ORIENT GENE	0.026	Symptomatic vs asymptomatic	5.02 (2.05, 12.81) p <0.001
	ASSIST-ASYMP-INNOVA 25	1	Symptomatic vs asymptomatic	1.77 (0.66, 4.96) p = 0.263
	SELF-SYMP-DHSC 3&7	0.002	Symptomatic vs asymptomatic	1.96 (1.42, 2.73) p <0.001
	SELF-SYMP-ORIENT GENE	1	Symptomatic vs asymptomatic	1.10 (0.64, 1.88) p = 0.726
Variant	SELF-ASYMP-DHSC 3&7	0.857	B.1.1.529 (Omicron) vs. B.1.617.2 (Delta)	6.19 (1.25, 47.35) p = 0.041
	SELF-ASYMP-ACON	1	B.1.1.529 (Omicron) vs. B.1.617.2 (Delta)	3.27 (0.79, 15.39) p = 0.110

(B.1.1.7 (Alpha) or B.1.617.2 (Delta)[21])	SELF-ASYMP-ORIENT GENE	1	B.1.1.529 (Omicron) vs. B.1.617.2 (Delta)	0.99 (0.33, 2.97) p = 0.984
	ASSIST-ASYMP-INNOVA 25	1	B.1.617.2 (Delta) vs B.1.1.7 (Alpha)	0.33 (0.08, 1.50) p = 0.135
			B.1.1.529 (Omicron) vs B.1.1.7 (Alpha)	0.23 (0.05, 1.12) p = 0.06
	SELF-SYMP-DHSC 3&7	1	B.1.617.2 (Delta) vs B.1.1.7 (Alpha)	1.16 (0.79, 1.69) p = 0.46
			B.1.1.529 (Omicron) vs B.1.1.7 (Alpha)	5.04 (1.09, 36.61) p = 0.06
	SELF-SYMP-ORIENT GENE	N/A	N/A	N/A
			N/A	N/A
Channel of testing (Community Testing)	SELF-ASYMP-DHSC 3&7	1	Health vs community testing	0.81 (0.11, 8.22) p =0.839
			Schools (over 18) vs community testing	0.55 (0.10, 3.79) p = 0.512
			Schools (18 and under) vs community testing	0.31 (0.09, 1.12) p =0.070
	SELF-ASYMP-ACON	1	Health vs community testing	0.58 (0.12, 3.43) p = 0.522
			Schools (18 and under) vs community testing	0.19 (0.04, 0.74) p = 0.024
	SELF-ASYMP-ORIENT GENE	1	Schools (over 18) vs community testing	1.00 (0.30, 3.49) p = 0.995
	ASSIST-ASYMP-INNOVA 25	0.726	TCT vs community testing	0.44 (0.23, 0.83) p = 0.012

[21] B.1.1.7 (Alpha) was used as a comparison where there were sufficient samples when this was not the case B.1.617.2 (Delta) was used as the comparison

			Universities vs community testing	0.87 (0.29, 2.67) p = 0.811
	SELF-SYMP-DHSC 3&7	0.002	IPTC vs LTS/RTS	2.93 (1.93, 4.52) p <0.001
	SELF-SYMP-ORIENT GENE	N/A	N/A	N/A
Time period (1st time period)[22]	SELF-ASYMP-DHSC 3&7	0.355	Period 3 vs period 2	2.97 (1.30, 7.21) p = 0.013
	SELF-ASYMP-ACON	N/A	N/A	N/A
			N/A	N/A
	SELF-ASYMP-ORIENT GENE	1	Period 3 vs period 2	1.38 (0.32, 6.02) p = 0.660
	ASSIST-ASYMP-INNOVA 25	0.211	Period 2 vs period 1	0.47 (0.25, 0.87) p = 0.017
			Period 3 vs period 1	0.31 (0.15, 0.64) p =0.002
	SELF-SYMP-DHSC 3&7	<0.001	Period 2 vs period 1	1.17 (0.80, 1.72) p = 0.41
			Period 3 vs period 1	3.26 (2.00, 5.39) p = 0
	SELF-SYMP-ORIENT GENE	N/A	N/A	N/A
			N/A	N/A
Time lag between symptom onset and testing (1st Quantile)	SELF-ASYMP-DHSC 3&7	0.799	2nd Quantile vs 1st Quantile	9.83 (1.34, 214.93) p = 0.055
	SELF-ASYMP-ACON	1	2nd Quantile vs 1st Quantile	1.15 (0.07, 26.04) p = 0.924
	SELF-ASYMP-ORIENT GENE	1	2nd Quantile vs 1st Quantile	0.43 (0.02, 4.64) p = 0.512

[22] Where there were no samples in period 1, period 2 was used as the comparison.

	ASSIST-ASYMP-INNOVA 25	1	2nd Quantile vs 1st Quantile	1.01 (0.09, 13.20) p = 0.990
	SELF-SYMP-DHSC 3&7	1	2nd Quantile vs 1st Quantile	0.74 (0.32, 1.70) p = 0.48
	SELF-SYMP-ORIENT GENE	1	2nd Quantile vs 1st Quantile	1.69 (0.90, 3.24) p = 0.106

N/A signifies that the value could not be calculated because there were either no samples, not enough samples (<10), or that there was no comparator within the analysis set against which to compare.

As post-hoc analysis, time period was broken into quarters and months to check if the way of stratifying time made a difference to the result. Regardless of the way time was broken down, the only analysis set in which time was a significant predictor of a TP was SELF-SYMP-DHSC 3&7. This analysis set is made up of 2 distinct testing channels (IPTC and RTS/LTS) and the significance of this result might point to a difference between testing channels, rather than an effect of time period itself. Although there was a statistically significant difference between these testing channels in their chances of returning a TP, both channels showed sensitivity above baseline levels.

About the UK Health Security Agency

UKHSA is responsible for protecting every member of every community from the impact of infectious diseases, chemical, biological, radiological and nuclear incidents and other health threats. We provide intellectual, scientific and operational leadership at national and local level, as well as on the global stage, to make the nation health secure.

UKHSA is an executive agency, sponsored by the Department of Health and Social Care.

Published: December 2022
Publishing reference: GOV-13369

UKHSA supports the
Sustainable Development Goals