Assessing Innova Lateral flow device in light of Liverpool field study evidence

Christos Maniatis *

December 2020

Abstract

Innova lateral flow device was introduced in public by the British government as a tool to contract SARS-CoV-2. Evaluation from PHE and Oxford university found an overall sensitivity and specificity of 76.8% and 99.6%. A later field study conducted in Liverpool found that this testing kit was able to identify only half of the infected people. In light of these evidence, we propose an alternative strategy using existing kits to improve sensitivity until a more reliable testing kit is adopted.

1 Introduction

The importance of testing and subsequent contact tracing to contain SARS-CoV-2 has been repeatedly highlighted from scientists, governments and international organizations like World Health organization (WHO) (LFD, b). The most reliable diagnostic procedure, reverse transcription polymerase chain reaction (RT-PCR), is performed in specialized laboratories and has high turnaround time. Furthermore, expanding existing facilities to cope with the increased testing demand is challenging. Lateral flow devices for covid-19 were introduced to the public as a fast alternative diagnostic test to assist in the safe opening up of communities Woloshin et al. (2020)

Early assessment of the Innova test by PHE found an overall sensitivity and specificity of 76.8% and 99.6% respectively (LFD, b). A later community testing pilot in Liverpool adjusted sensitivity to 48.89% (with 95% CI being $[33.70\% - 64.23\%]$) and confirmed the original specificity (99.93% with 95% CI being $[99.76\% - 99.99\%]$) (LFD, a). This sensitivity does not allow for conclusive results, as negative results does not imply an absence of infection, especially in areas with high Covid-19 prevalence Woloshin et al. (2020).

Despite the limited individual test sensitivity, it is possible to minimize infection uncertainty by performing several simultaneous or repeated tests, as suggested by Woloshin et al. (2020). In this document, we explore the proposed strategy assuming no interaction between tests (ie. consecutive tests are performed in short time interval on different swab samples). More precisely, we conduct a series of rapid tests and assume that the patient is infected if one or more tests return positive as suggested in Cornell (1978). For this strategy, we demonstrate the composite test sensitivity increases as a function of conducted tests. Then, using sequential sampling and Bayesian statistics, we demonstrate that under certain circumstances it is possible to reduce the number of required tests.

2 Assumptions

In this work, a series of assumptions are made to derive formulas and results. Working numbers for sensitivity and specificity are estimated under the assumption that PCR is the gold standard. PCR as every testing method is prone to error, which could potentially add to the existing uncertainty around estimates. The second assumption is that tests are independent. This means that for each swab a unique testing kit should be used and in a series of tests early outcomes do not affect following tests, unlike for example Hemoccult Test reported in Politser (1982). There is literature that allows for dependence between tests Galen et al. (1975) but it is beyond the scope of this study. It is also important to avoid exposure to the virus between repeated tests, for these results to apply. So, repeated rapid tests must be completed within a reasonable amount of time

*s1315538@sms.ed.ac.uk, Centre for Doctoral Training in Data Science, Bayes Centre, The University of Edinburgh

without coming in contact with potentially infected individuals. Hence, this might not apply in situations where tests are done days apart. Specificity and sensitivity estimates used here are tied to the lateral flow device provided by Innova as measured in (LFD, b). This does not apply to testing kits with different sensitivity and specificity levels. Finally, we do not take into account inconclusive test results.

3 Initial risk assessment

Accurately assessing the probability of being infected in asymptomatic people is challenging as it depends on various parameters like social contacts, location, and travel history. The social contacts significantly contribute to the infection risk, especially in case of close contact with an infected person. In such scenarios there are publications like Lelieveld et al. (2020) which suggest models that aid in the risk estimation. Furthermore, in some places, authorities have introduced contact tracing applications. Local and national authorities report percentage of infected people in the sample tests they conduct. This statistic tends to over-estimate Covid-19 prevalence as people do not test unless they experience symptoms, but this is not a problem, especially near the epicenter. For travelers recently arrived from other areas, assuming they have no symptoms, a reasonable estimate should be the highest Covid-19 prevalence among the transit stop locations adjusted by transmission probability tied to the traveling method and route. Again, work like Lelieveld et al. (2020) can aid in the risk estimate.

4 Field study results

A field study was carried in Liverpool, one of the areas with the highest rates of Covid-19 in England, to test the accuracy of several types of SARS-CoV-2 tests including the Innova lateral flow swab antigen test (Iacobucci, 2020). Lateral flow device detects virus presence by applying a nose and throat swab to a special test kit. Results do not require laboratory processing and are available in around 30 minutes. Preliminary evaluation of the lateral flow tests from PHE Porton Down and University of Oxford found sensitivity and specificity around 76.8% and 99.6% respectively (Wise, 2020). Liverpool study presented the following more conservative estimates.

	Values	CI
p_{sen}	48.89%	$[33.70\% - 64.23\%]$
p_{sp}	99.93%	$[99.76\% - 99.99\%]$

Table 1: Sensitivity and specificity as measured in Liverpool field study (LFD, a)

This levels of sensitivity could give to infected people the false reassurance, turning them to superspreaders. An alternative could be to perform repeated test as suggested in Woloshin et al. (2020) and label one infected if at least one result comes back positive. With this approach, we can increase accuracy using less reliable testing kits.

5 Composite testing

Conducting multiple tests instead of a single one and labeling someone infected if at least one of them is positive has the potential to significantly increase the effectiveness of the testing procedure. Assuming that a test has initial sensitivity p_{sen}, the combined sensitivity becomes $p_s^c = 1 - (1 - p_{sen})^N$ after N tests. Using sensitivity and specificity values presented in (LFD, a) and Wise (2020) the combined sensitivity as a function of performed tests is plotted in figure 1. Hence to achieve RT-PCR sensitivity levels using the Lateral flow device, $3 - 7$ tests have to be done depending the test sensitivity compared to 2 that the original valuation required. If factory specifications for sensitivity is enough for the purpose of test and trace strategy, then 4 test can guarantee that target safety is met irrespective of where the actual specification falls in.

6 Sequential sampling

In practice, spending 7 tests to minimize the risk of one being infected will consume extremely fast all the available resources. Item by item sequential sampling (Wald, 1945) can reduce the

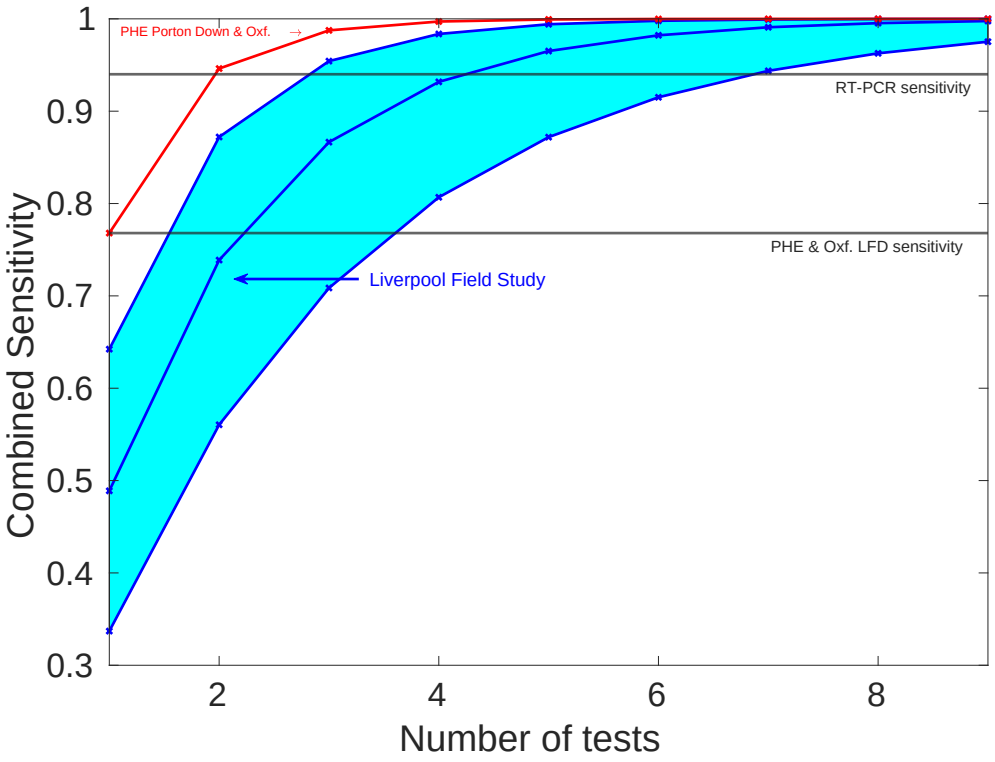

Figure 1: Composite test sensitivity as function of conducted tests for PHE Porton Down and University of Oxford initial valuation and field study

number of SARS-CoV-2 tests at the cost of increasing the risk of false reassurance to infected people. This approach works by conducting a series of tests until a definitive answer is produced. Depending on the number of tests done, tests specs and allowed risk, two variable thresholds define three regions for infection, safety and more testing. The number of negative results place the patient in of these regions. Assuming target miss-rate and fall-out 5% this strategy is neatly summarized in the plots of figure 2.

Looking at figure 2a and comparing it to 2b-2d it is evident that accurate and reliable tests can minimize the evidence required to determine the state of an asymptomatic person. Figures 2b-2d can help determine the number of tests before getting conclusive results based on the sensitivity/specificity estimates of the Liverpool field study (LFD, a). Irrespective of where in the confidence interval true sensitivity and specificity fall, if three rapid tests are used and at least two are positive, then it is safe to say that the tested is infected.

7 Decision making under uncertain results

In decision-making, it is vital to assess the probability of being infected before doing a test, otherwise results might be misleading. If prior infection probability is close to certainty (ie. 0 or 100%), test results do not make any difference. Starting from an initial infection probability assessment and adjusting beliefs using evidence and Bayes rule. Then by setting probability thresholds for treating someone as healthy or infected can aid in the decision-making process. In this work, we set 2 safety levels for assuming that someone is healthy, one at 1% and another one at 5%. If posterior infection probability falls below these thresholds, we could say that the tested person is healthy. Threshold at 5% could be adequate for people working/studying from home. For care-workers and medical personnel, the extra safety provided by the 1% threshold is probably preferable due to the damaging consequences associated with transmission. On the other hand, someone is assumed to be infected if the probability of infection given test outcomes is 90% or more.

In the presence of at least one positive result in less than five tests, the probability of infection rises above 90% for a starting infection probability around 7% or more. Hence, if 3 tests were conducted sequentially the first positive result would be enough to stop the sequence and treat

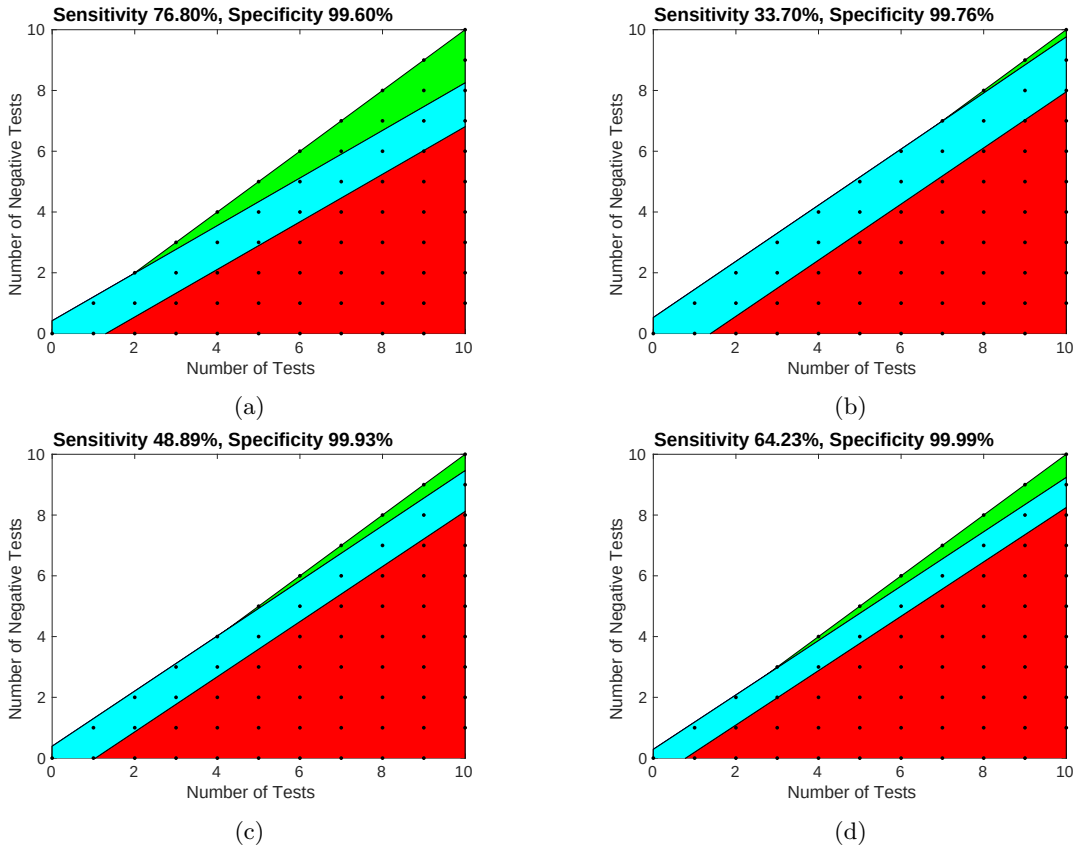

Figure 2: Decision regions using different sensitivity and specificity for sequential sampling. Red area is where there is strong evidence in favor of infection, blue area is where more testing need to take place and green area is where the tested person probably healthy.

the tested as infected as we can see from figures 4b-4d. On the contrary, repeated negative results make infection less likely, as the diagnostic test is not accurate enough. To minimize infection risk below 5% requires at least 2 negative results and prior infection risk less than 20% (figure 3c) or at least 3 tests assuming the worst-case sensitivity/specificity (figure 3b). For the stricter threshold, the minimum number of tests has to double for an average sensitivity and specificity of 48.89% and 99.93% respectively. If test specs are closer to the ones presented in figure 3b, then this test kit is only fit for testing care-workers with low exposure risk to the virus. So, it is probably better to seek more accurate testing methods.

8 Methods

Suppose we have a test for SARS-CoV-2 with sensitivity p_{sen} and specificity p_{sp} and use it to a person N times. Under the assumption that a person is infected if at least one or more tests give positive results Cornell (1978), the sensitivity and specificity of the composite tests are $p^c_{sen} = 1 - (1 - p_{sen})^N$ and $p^c_{sp} = p^N_{sp}$ (Lau, 1989). An interesting property derived from these equations is that sensitivity increases with more tests, but specificity drops. This means that we manage to identify more from the infected group but we also mislabel some as having the virus. The number of false positives is not a significant problem, as there is the safeguard of the molecular RT-PCR test.

In some cases, it might not be feasible to use many tests, due to limited time or supply. Sequential sampling can be used to reduce the number of tests. To determine if test results can safely conclude the infection status of a patient, a graph with an acceptance and rejection line is plotted. Given miss-rate p_1, specificity p_2, target miss-rate $1 - a$ on the composite test and target

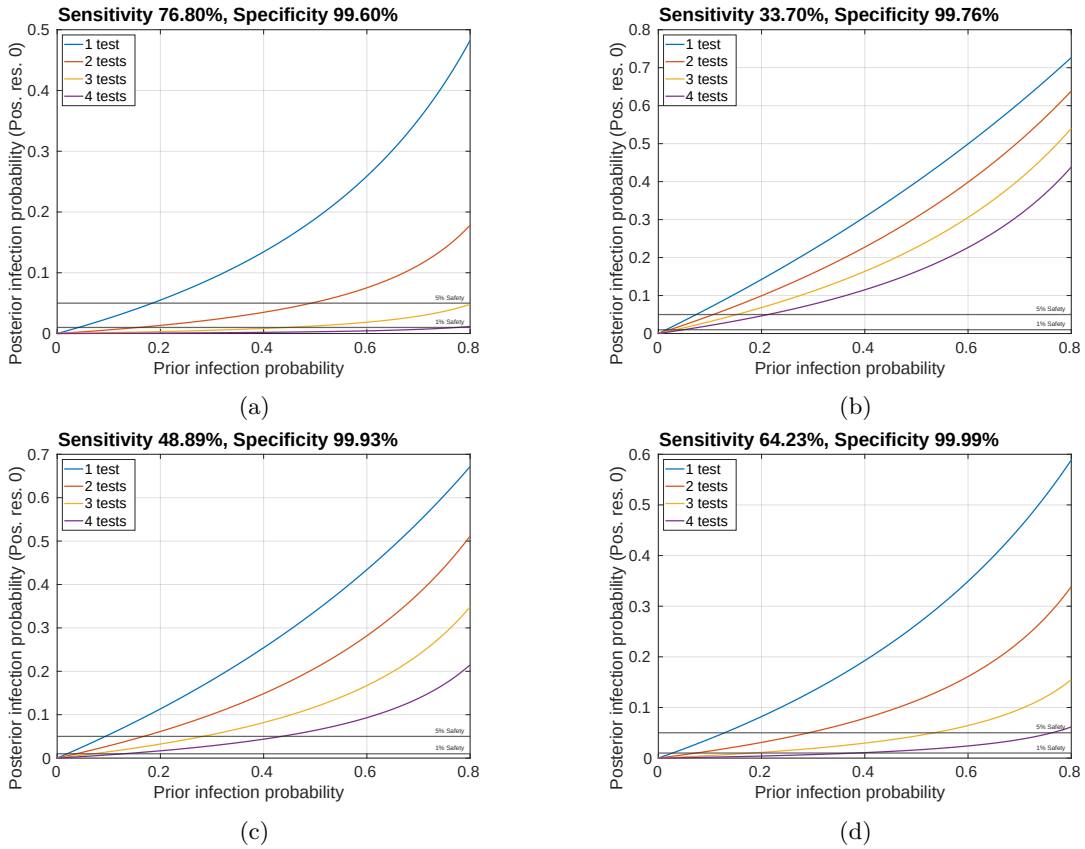

Figure 3: Posterior infection probabilities as a function of prior when all test results are negative.

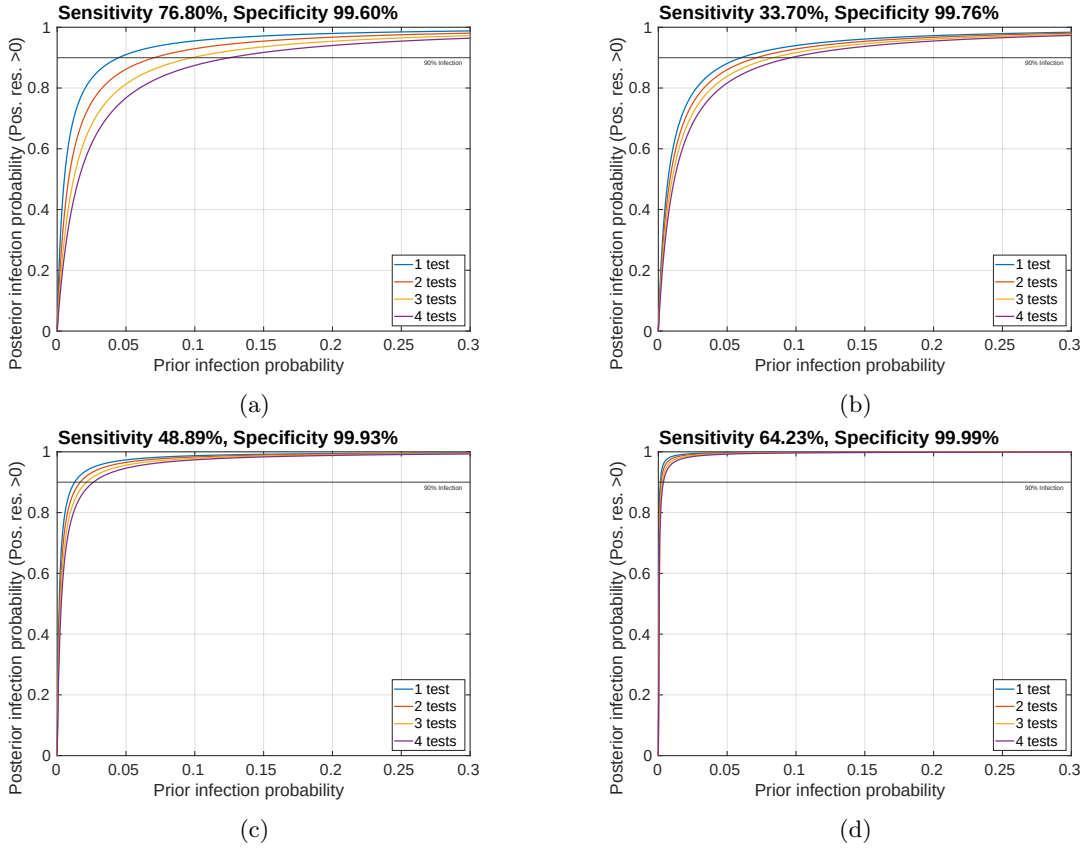

Figure 4: Posterior infection probabilities as a function of prior when at least one test is positive.

fall-out b on the composite test the limit lines given by Montogomery (2009) are

$$x_a = -h_1 + sn \text{ (acceptance line)}$$
$$x_r = h_2 + sn \text{ (rejection line)},$$
$$k = \ln\left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]$$

After each result, a point with abscissa equal to the number of conducted tests and ordinate equal to the negative results is placed on the same plot (Lau, 1989). If the point is above the rejection line, the tested person is labeled SARS-CoV-2 free. When the point is in the region defined by acceptance and rejection line, then more tests need to be done. In case the point falls below the acceptance line, the tested person is considered infected and a more reliable clinical test has to be used.

Understanding the characteristics of a test can provide useful insights. However, attention needs to be paid to avoid a common fallacy called base-rate neglect (Watson et al., 2020). That happens when people fail to account for pre-test probability of being infected and only respond to a piece of new information (Watson et al., 2020). If the pre-test probability is high, testing loses its value, as results cannot lower the posterior probability below safety standards (Woloshin et al., 2020). To estimate the risk of being infected given no positive test in N conducted tests to a particular person, we use the following formula.

$$\mathbb{P}\left(D^{+}|NT^{-}\right) = \frac{\mathbb{P}\left(NT^{-}|D^{+}\right)\mathbb{P}\left(D^{+}\right)}{\mathbb{P}\left(NT^{-}\right)} \tag{2}$$

Where D^{+} denotes SARS-CoV-2 infection. D^{-} denotes infection free state, T^{-}/NT^{-} negative test outcome/ N negative tests outcome and T^{+} positive test. To estimate the risk of being infected given one or more positive results in N conducted tests we use the following formula.

$$\mathbb{P}\left(D^{+}||T^{+}|\geq 1\right) = \frac{\mathbb{P}\left(|T^{+}|\geq 1|D^{+}\right)\mathbb{P}\left(D^{+}\right)}{\mathbb{P}\left(|T^{+}|\geq 1\right)} \tag{3}$$

Where $|T^{+}|$ is the number of positive results in N tests. As we can see from equation 2 the probability of being infected after N negative tests is the product of initial risk estimate $\mathbb{P}\left(D^{+}\right)$, how likely is to observe N negative N conducted tests adjusted by the probability of observing N negative tests irrespective of infection status (This term takes into account all performance metrics related to the test presented in table 2).

$$\mathbb{P}\left(D^{+}|NT^{-}\right) = \frac{(1-p_{sen})^{N}\mathbb{P}\left(D^{+}\right)}{(1-p_{sen})^{N}\mathbb{P}\left(D^{+}\right) + p_{sp}^{N}\mathbb{P}\left(D^{-}\right)} \tag{4}$$

Similarly, the probability of being infected after 1 or more positive results in N tests is

$$\mathbb{P}\left(D^{+}||T^{+}|\geq 1\right) = \frac{[1-(1-p_{sen})^{N}]\mathbb{P}\left(D^{+}\right)}{[1-(1-p_{sen})^{N}]\mathbb{P}\left(D^{+}\right) + [1-p_{sp}^{N}]\mathbb{P}\left(D^{-}\right)} \tag{5}$$

Qunatity	Explanation
Sensitivity	Probability of getting a positive test result given Covid-19 infection
Miss rate	Probability of getting a negative test result given Covid-19 infection
Specificity	Probability of getting a negative test result given no Covid-19 infection
Fall-out	Probability of getting a positive test result given no Covid-19 infection

Table 2: Terminology explanation

9 Conclusion

In this study, we investigated the viability of using repeated testing kits to assess whether a patient is infected as suggested by Woloshin et al. (2020) using a strategy proposed in Cornell (1978) and specificity/sensitivity results provided in Liverpool field study. Depending where sensitivity/specificity fall in the confidence interval, results will vary. Despite that, we have determined that for the general population, a sequence of 3 tests can determine if a patient is healthy or infected with small error probability. If all three results are negative in test sequence, then the tested person is probably healthy. Such people can return to their daily routine, respecting social distancing measures as a small chance of being infected still exists. If one or more tests in the sequence returns positive, then the tested should get a RT-PCR test to remove the possibility of false positive. For care-workers and medical personnel, to conclude whether the test is usable or not depends on test's specs. Using the confidence interval provided by (LFD, a) a sequence of $3-7$ of tests have to be used or seek a more reliable test to assess infection status.

References

Liverpool field study. https://www.gov.uk/government/publications/innova-lateral-flow-sars-cov-2-antigen-test-accuracy-in-liverpool-pilot-preliminary-data-26-no a. Accessed: 2020-12-29.

Phe innova leteral flow devise evaluation. https://www.ox.ac.uk/news/2020-11-11-oxford-university-and-phe-confirm-lateral-flow-tests-show-high-specificity-and-are, b. Accessed: 2020-12-29.

Richard G Cornell. Sequence length for repeated screening tests. *Journal of chronic diseases*, 31 (8):539–545, 1978.

RS Galen, SR Gambino, et al. The predictive value and efficiency of medical diagnoses. *J. Willey & Sons, New York*, 1975.

Gareth Iacobucci. Covid-19: Mass population testing is rolled out in liverpool, 2020.

Tai Shing Lau. On repeated screening tests. *Biometrics*, pages 891–898, 1989.

Jos Lelieveld, Frank Helleis, Stephan Borrmann, Yafang Cheng, Frank Drewnick, Gerald Haug, Thomas Klimach, Jean Sciare, Hang Su, and Ulrich Pöschl. Model calculations of aerosol transmission and infection risk of covid-19 in indoor environments. *International Journal of Environmental Research and Public Health*, 17(21):8114, 2020.

DC Montogomery. Introduction to statistical quality control 6th edition. *Arizona state University*, 2009.

Peter Politser. Reliability, decision rules, and the value of repeated tests. *Medical Decision Making*, 2(1):47–69, 1982.

Abraham Wald. Sequential tests of statistical hypotheses. *The annals of mathematical statistics*, 16(2):117–186, 1945.

Jessica Watson, Penny F Whiting, and John E Brush. Interpreting a covid-19 test result. *Bmj*, 369, 2020.

Jacqui Wise. Covid-19: Lateral flow tests miss over half of cases, liverpool pilot data show, 2020.

Steven Woloshin, Neeraj Patel, and Aaron S Kesselheim. False negative tests for sars-cov-2 infection—challenges and implications. *New England Journal of Medicine*, 2020.