

Application of pooled testing in screening and estimating the prevalence of COVID-19

Pritha Guha* Apratim Guha† Tathagata Bandyopadhyay‡

Abstract

The recent emergence of the COVID-19 pandemic has posed an unprecedented healthcare challenge and catastrophic economic and social consequences to the countries across the world. The situation is even worse for emerging economies like India. WHO recommends mass scale testing as one of the most effective ways to contain its spread and fight the pandemic. But, due to the high cost and shortage of test kits, specifically in India, the testing is restricted to only those who are symptomatic. In this context, pooled testing is recommended by some experts as a partial solution to overcome this problem. In this article, we explain the basic statistical theory behind the pooled testing procedure for screening as well as prevalence estimation. In real world situations, the tests are imperfect, and lead to false positive and false negative results. We provide theoretical explanation of the impact of these diagnostic errors on the performances of individual testing and pooled testing procedures. Finally, we study the effect of misspecification of sensitivity and specificity of tests on the estimate of prevalence, an issue, which is debated a lot among the scientists in the context of COVID-19. Our theoretical investigations lead to some interesting and precise understanding of some of these issues.

Keywords : Pooled testing, Prevalence Estimation, Imperfect test, COVID-19

1 Introduction

Since its detection in the Wuhan province of China at the end of December 2019, the COVID-19 pandemic, caused by the virus SARS-CoV-2, has created catastrophic social, economic and health consequences for countries around the world. For an emerging economy like India, it poses a formidable challenge to the policy makers to contain its spread at a level so that the health care infrastructure is not overwhelmed. At the time of writing this article, in India, the number infected is around 150,000 and more than 4300 deaths have occurred due to COVID-19. A large part of the country is under 'lockdown'. Given its unprecedented economic and social consequences, the policy makers are now desperately looking for an exit route to bring the economy back to normalcy. Effective implementation of social distancing, mass testing, contact tracing and quarantining is essential to contain the spread of the virus. Given that a large percentage of COVID-19 cases are asymptomatic, and are responsible for spreading the virus¹⁻³, it has been advised by the World Health Organization (WHO) that the most effective way to control the spread of the disease is to test as many people as possible. This was found to be effective in containing the spread of the virus in countries like South Korea, Singapore and China⁴⁻⁶. However, the high cost and

*XLRI-Xavier School of Management, Jamshedpur 831001, India. Email: pritha@xlri.ac.in

†XLRI-Xavier School of Management, Jamshedpur 831001, India. & Indian Institute of Management, Ahmedabad, Gujarat 380015, India. Email: apratim@xlri.ac.in

‡Indian Institute of Management, Ahmedabad, Gujarat 380015, India. Email: tathagata@iima.ac.in (Corresponding author)

the short supply of testing kits⁷ are the main impediments for a country like India in conducting mass testing. In order to ramp up the testing efforts, some experts^{8–11} are recommending the use of pooled testing technique, a technique originally proposed by Dorfman¹² to increase the speed, and reduce the cost, of screening the US army recruits for syphilis during the Second World War. Recently the Indian Council for Medical Research (ICMR) has given its approval¹³ and issued detailed guidelines¹⁴ for carrying out pooled testing for screening purposes.

It is interesting to observe that most experts recommend use of pooled testing only for the screening purpose. However, the data collected by pooled testing can also be effectively utilized for the estimation of prevalence of the virus, which is a later development in pooled testing research¹⁵. This fact is not as widely known as the use of pooled testing for screening purpose in the general scientific community. As observed by many^{16,17}, the estimation of prevalence of the SARS-CoV-2 virus is crucial for an accurate estimation of fatality rate. The initial estimate of fatality rate given by WHO¹⁸ was scandalously high, because it was based only on deaths among symptomatic patients. Later, it was revised downward after taking into consideration the fact that a significant percentage of the infected population is asymptomatic, and hence remain undetected. If 80% of the cases are asymptomatic, then the fatality rate computed only from the symptomatic cases would be five times higher than the true fatality rate. Such misinformation may create an unnecessary panic in the minds of the policy makers, which may lead to wrong decisions. Recently, a few planned serology surveys were carried out for the estimation of the prevalence of the SARS-CoV-2 virus (e.g., Sempos and Tian¹⁹), and the results suggest that the overestimation of fatality rate in the population may be between fifty to eighty five times of the original fatality rate. These results created waves in the scientific community. We will revisit these results in Section 4.

This article mainly serves two purposes. First, we present the statistical theory of the pooled testing procedure for screening, and also for the estimation of prevalence using basic probability theory²⁰. Next, we discuss some practical issues arising out of the fact that these tests are imperfect. In other words, the tests may yield false positive and/or false negative results. Naturally, it raises an important practical question: what is the impact of these diagnostic errors on screening as well as on prevalence estimation? We study these effects using some hypothetical scenarios and discuss its implications in the context of COVID-19 pandemic.

The rest of the article is divided as follows. In Section 2 we present the statistical theory of pooled testing. In Section 3 we discuss the consequences when the test is imperfect. In Section 4, we discuss the estimation of prevalence using pooled data, and compare it with the estimation procedure with individual data. Finally, we conclude with some remarks in Section 5.

2 Dorfman pooled testing technique

Pooled testing was originally introduced by Dorfman¹² in infectious disease studies to reduce the cost and increase the speed of data collection. Typically, for an infectious disease, a sample of blood or urine is tested for presence or absence of the disease. Rarer the disease, more effective is the pooled testing technique. Dorfman's technique runs as follows. Suppose n samples of blood are to be tested for presence or absence of a disease. A sample is either positive (disease present) or negative (disease absent). We attach a Bernoulli variable Y to each sample. We assume $Y = 1$ if the sample is positive, and $Y = 0$ if it is negative. Instead of testing the samples individually, the samples are pooled into J groups of sizes n_1, \dots, n_J . Let Y_{ij} denote the Y -value for the i -th sample in the j -th group. The samples in each group are pooled and tested. Instead of observing Y_{ij} , in the pooled testing set-up, we observe $Y_j^* = \max_{i=1, \dots, n_j} Y_{ij}$. It is further assumed that, $Y_j^* = 0$, if and only if all samples in the j -th group are negative, i.e., $Y_{ij} = 0, i = 1, \dots, n_j$ and $Y_j^* = 1$, if and only if at least one sample in the j -th group is positive, i.e., $Y_{ij} = 1$, for at least

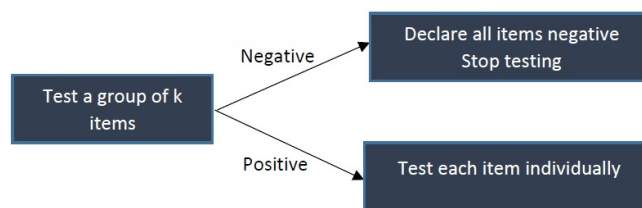


Figure 1: Process diagram of Dorfman's algorithm for pooled testing

one $i = 1, \dots, n_j$. If for a group, the test outcome is positive, then the samples in the group are individually tested to detect the positive samples, and hence the diseased individuals. If the result is negative for a group then no further testing is required. Thus, for a positive test outcome, the number of tests required is one more than the group size, and for a negative test outcome, a single test is enough. If the chance of a positive test outcome for a group is small, in other words, the prevalence of the disease is low, then the average number of tests required would be much smaller in a group testing set-up than individual testing. Thus, it leads to considerable savings in the cost and time of testing. The process diagram is presented in Figure 1 for a group of size k .

Now, suppose that the prevalence of a disease is p , and assume without loss of any generality $n_j = k$, i.e., all groups are of equal size, and consequently $n = Jk$. Notice that, $P(Y_j^* = 0) = (1 - p)^k$ and $P(Y_j^* = 1) = 1 - (1 - p)^k$. Further, suppose N_j denotes the number of tests for the j -th group. Clearly, it is equal to $k + 1$ if $Y_j^* = 1$, and 1 if $Y_j^* = 0$. Thus the expected number of tests for the j -th group is

$$E(N_j) = (k + 1) - k(1 - p)^k. \quad (1)$$

If p is small then $E(N_j)$ is close to 1 even for a sufficiently large group size k , and thus leading to a substantial saving in time and cost than individual testing. Notice that the technique is mainly useful when the prevalence is low.

The simplicity of the idea, and its easy implementation, has led to its wide applicability in different fields (see Roy and Banerjee¹⁵ and the references therein). Its use in the screening of sexually transmitted disease like chlamydia, gonorrhoea²¹ and HIV^{22,23} is worth mentioning. American Red Cross and the European Blood Alliance routinely use it to screen donated blood for infectious diseases^{24,25}. Given the huge shortage of test kits around the world, researchers worldwide are recommending the use of pooled testing for COVID-19^{8,26-28}.

3 Pooled testing for screening when the test is imperfect

Dorfman¹² proposed the pooled testing technique assuming the test to be perfect, i.e., the chance of a false positive or a false negative test result to be zero. However, in reality, tests are far from perfect. In testing for the SARS-CoV-2 virus, two types of tests are mainly in use. "The first is a reverse-transcription polymerase chain reaction test, or RT-PCR. This is the most common diagnostic test used to identify people currently infected with SARS-CoV-2. It works by detecting viral RNA in a person's cells – most often collected from their nose. The second test being used is called a serological or antibody test. This test looks at a person's blood to see if they have produced antibodies for the SARS-CoV-2 virus. If a test finds these antibodies, it means a person was infected and made antibodies in response"²⁹. How accurate are these tests? The accuracy of a test is measured by two numbers: sensitivity and specificity attached to it. If a test has 5% chance

of a false negative (positive) outcome, its sensitivity (specificity) is 95%. Although in laboratory setting it is observed that RT-PCR test has high sensitivity and specificity, in the real world testing condition the sensitivity and the specificity are usually much lower because the testing conditions, the method of sample collection, and the sample preservation technique are far from perfect. For example, in real world testing condition the sensitivity of RT-PCR test ranges between 66% to 80%²⁹. Thus, out of three infected people on the average one may test negative. The sensitivity and specificity of the antibody tests are also found to be very high in laboratory setting, but, then again, in real world condition the accuracy is bound to suffer.

Using the same notations as above, we now find $E(N_j)$ assuming that the sensitivity and specificity of the test are S_e and S_p , respectively. Notice that for the j -th group, we now observe \hat{Y}_j^* instead of Y_j^* , a surrogate (or proxy) for Y_j^* , where $P(\hat{Y}_j^* = 1|Y_j^* = 1) = S_e$ and $P(\hat{Y}_j^* = 0|Y_j^* = 0) = S_p$. A simple probability calculation then yields $P(\hat{Y}_j^* = 1) = P(Y_j^* = 1)S_e + P(Y_j^* = 0)(1 - S_p) = S_e + (1 - p)^k(1 - S_p - S_e)$, and hence,

$$E(N_j) = 1 + k[S_e + (1 - p)^k(1 - S_p - S_e)]. \quad (2)$$

The expected number of tests for n people would then be $J \times E(N_j)$, and hence compared to the individual testing, the reduction in the number of tests achieved by using group testing is $J(k - E(N_j))$. This reduction is a measure of the *efficiency* of the pooled testing procedure. Also, notice that for $S_p = S_e = 1$, Equation (2) reduces to (1).

For a perfect test, there is no chance of misclassification of an infected as non-infected or vice versa, and hence the efficiency of pooled testing procedure is measured only by $E(N_j)$. However, for an imperfect test, $E(N_j)$ is not enough to judge the efficiency of the pooled testing procedure. One also needs to measure the sensitivity and specificity of the pooled testing procedure, and most importantly, the probabilities of the diagnostic errors.

Let T^+ and T^- denote the events that an individual is tested positive and negative, respectively, by the pooled testing procedure. Further, let I (I^c) denote the events that an individual is infected (uninfected). Then one can easily check that the sensitivity, say, S_e^D , and specificity, say, S_p^D , of the Dorfman's pooled testing procedure, are given by

$$S_e^D = P(T^+|I) = S_e^2, \quad (3)$$

and

$$S_p^D = P(T^-|I^c) = 1 - (1 - S_p)[S_e + (1 - S_p - S_e)(1 - p)^{k-1}], \quad (4)$$

respectively. The proofs of (3)-(4) are given in the appendix.

Notice that S_e^D depends only on S_e while S_p^D depends on all three parameters S_e , S_p and p . As stated above, S_e^D and S_p^D provide useful information about the performance of the pooled testing procedure. However, the posterior or inverse probabilities $P(I^c|T^+)$ and $P(I|T^-)$, called the False Positive Predictive Value (FPPV) and the False Negative Predictive Value (FNPV) respectively, are often critical for assessing the performance of the test in the real life situation. These are measures of diagnostic errors. The *FPPV* and the *FNPV* represent the proportion of misclassified individuals among those who are tested positive, and among those who are tested negative, respectively. In the context of testing for sera samples of HIV virus, Litvak (²²) proposed that the five key elements to be used to capture the overall performance of a pooled testing procedure should be: $E(N_j)$, S_e^D , S_p^D , *FPPV* and *FNPV*.

Now, using Bayes' Theorem (²⁰), one can easily show that

$$FPPV = \frac{(1-p)(1-S_p^D)}{(1-p)(1-S_p^D) + pS_e^D}, FNPV = \frac{p(1-S_e^D)}{p(1-S_e^D) + (1-p)S_p^D}. \quad (5)$$

Notice that, for individual testing $FPPV$ and $FNPV$, denoted henceforth $FPPV_I$ and $FNPV_I$ respectively, are obtained by substituting S_e^D and S_p^D by S_e and S_p , respectively, in (5).

In Tables 1a and 1b, we furnish the values of $E(N_j)$, S_e^D , S_p^D , $FPPV$, $FNPV$, $FPPV_I$ and $FNPV_I$ for some appropriately chosen values of the sensitivity (S_e), the specificity (S_p) and the prevalence (p). The "Eff" (efficiency) column gives the percentage reduction in the expected number of tests achieved by using the pooled testing than the individual testing. Evidently, with increase in p , and decrease in S_e and S_p , the efficiency reduces. The value of $S_e^D (= S_e^2)$ is always less than, or equal to, S_e . However, for a given set of values of p , S_e and S_p , the value of S_p^D is always more than S_p , and for smaller values of S_p this effect is substantial. Thus, the pooled testing procedure may lead to a significant improvement of the specificity, especially when the specificity of the individual test is low. Further, it is interesting to observe that even a slight decrease in S_p leads to a substantial increase in the value of $FPPV$ for any given values of p and S_e . On the other hand, a change in the value of S_e has negligible effect on the value of $FPPV$ for any given values of p and S_p . Also, with increase in p , $FPPV$ decreases for any given values of S_e and S_p . Most importantly, compared to the individual testing, the pooled testing leads to a substantial reduction in the value of $FPPV$, which is extremely important from the point of view of its application in practice. Still, $FPPV$ can take very high values even for pooled testing for a low prevalence disease with low specificity (S_p). While this is undesirable, this compares favourably to individual testing. In contrast, for the range of values considered in the tables for p , S_e and S_p , the impact of changes in the values of these parameters has little effect on $FNPV$.

Notice that in case both the sensitivity and the specificity of the test are low, say, for example, $S_e = 0.8$ and $S_p = 0.7$, even with 10% prevalence, 68% (77%) of the pooled (individual) testing results would be false positive, which is extremely high. A recent meta-study by a group of medical professionals from the Johns Hopkins University³⁰ demonstrated that over a different varieties of RT-PCR tests, which are the most commonly used tests for SARS-CoV-2, a best case scenario is an $S_e \approx 0.8$. Another meta-study³¹ has looked at thirty seven different external quality assessment studies of different medical assays, and found that in some cases, the specificity was as low as 0.83.

For testing for the SARS-CoV-2 virus, the above observations suggest that at the initial phase of the pandemic when the prevalence is low, (less than 10%,) pooled testing may lead to a substantially lower rate of false positive cases than individual testing, especially if a low specificity test is used. Admittedly, pooled testing leads to a slightly higher rate of false negative cases compared to individual testing, but as observed from Tables 1a and 1b, that effect is usually negligible.

Finally, it may be noted here that the impact of the two diagnostic errors are asymmetric. In case of more false positive cases, more people are to be quarantined leading to a lot of economic, social and emotional turmoil. On the other hand, in case of more false negative results, more infected people will be released in the population causing the disease to spread faster. Considering the costs of the errors, the policy planner has to make a trade-off.

4 Estimation of prevalence using pooled testing data

In this section, first we present the theory of estimation of prevalence of a disease from the test data using basic probability theory²⁰. Next, we study the impact of misspecification of sensitivity and specificity on the estimate of prevalence.

An accurate estimation of the prevalence of SARS-CoV-2-virus, being critical for assessing the lethality of the pandemic, is crucial for the policy makers to make strategic decisions. Recently, Ioannidis³², a Stanford medicine professor, lamented in a highly critical opinion piece on COVID-19: “Three months after the outbreak emerged, most countries, including the U.S., lack the ability to test a large number of people and no countries have reliable data on the prevalence of the virus in a representative random sample of the general population.” Ioannidis argued that in the absence of such data, the estimate of fatality rate is bound to be a substantial overestimate of the true fatality rate if a significant percentage of COVID-19 cases are undetected. It is now well known that a significant percentage of COVID-19 cases are asymptomatic. Raman R Gankhedkar, chief epidemiologist, ICMR, reported “Of 100 people with infection, 80 do not have symptoms,”³³. A lab in Iceland³⁴ has suggested that 50% of the infected are asymptomatic. In a Boston homeless shelter, out of 400 guests staying there, 146 tested positive for COVID-19, but all were reported to be asymptomatic³⁵. If, for example, 80% of the cases are really asymptomatic, then the estimate of fatality rate would naturally be five times the true fatality rate. While the reported percentage of asymptomatic cases vary from place to place, all agree that it is significant. Considering the possibility of substantial underestimation of infected cases in the absence of test data, Ioannidis has contended that the virus could be less deadly than people think, and destroying the economy in the effort to fight with the virus could be a “once-in-a-century-evidence-fiasco”.

A recent research paper published by a team of Stanford medical scientists¹⁶ lends support to Ioannidis’ contention. By estimating the prevalence of SARS-CoV-2 virus among the residents of Santa Clara County of California using antibody test data of a “properly selected” sample of 3300 residents, it predicted that between 50 to 85 times more residents of the county were actually infected than what appeared in the official tallies³⁶. The authors claimed that “their data helps prove ... if undetected infections are as widespread as they think, then the death rate in the county may be less than 0.2%, about a fifth to a tenth other estimates.” The publication of this research paper immediately created waves in social media, in press, and in policy circles; if indeed, this number is not far from the truth, then the COVID-19 fatality rate is not very different from common flu. Realizing the findings’ serious policy implications, especially its lending support to the view of lifting the lock down, some scientists immediately issued a caveat about the accuracy of the estimate. They expressed concerns about the flaws in the sample selection method (through Facebook advertisements), the statistical analysis carried out, and the reliability of the antibody test^{37,38}.

We now discuss the methodology used by the Stanford group¹⁶ for prevalence estimation based on individual serological testing data. Let us denote the probability of a positive test result for an arbitrarily chosen individual by π . Clearly, for a sample of size n , the estimate of π is simply the proportion of people being identified as positive out of the sample. But we know from the previous section that

$$\pi = P(\text{Positive result}) = (1 - p)(1 - S_p) + pS_e. \quad (6)$$

Plugging in the proportion of positive test results obtained from the sample, say $\hat{\pi}$, for π in (6), we obtain

$$\hat{p} = \frac{\hat{\pi} + S_p - 1}{S_e + S_p - 1}. \quad (7)$$

However, for estimating prevalence using Dorfman's algorithm, we simply need to replace S_e and S_p in (6) by S_e^D and S_p^D respectively (cf. (Equations 3)-(4)). Thus, we have

$$\pi = (1 - S_p)(1 - S_p - S_e)(1 - p)^k + (1 - S_p)S_e(1 - p) + pS_e^2, \quad (8)$$

or, equivalently,

$$\pi = (1 - S_p - S_e)[(1 - S_p)(1 - p)^k - pS_e] + (1 - S_p)S_e. \quad (9)$$

Hence, plugging in an estimate of π obtained from the test data to (8) and then solving it numerically for p we get an estimate of p , say \hat{p}^D . This estimate is consistent for p (i.e., close to p as the sample size n gets very large).

As noted in Section 3, both RT-PCR and antibody tests have high sensitivity and specificity in lab settings, but in a real world situation, these values may be substantially lower than those obtained in lab settings. Suppose S_e^P and S_p^P are respectively the sensitivity and the specificity values as perceived, and used by the scientist for estimating the prevalence from Equation (7) or (8), whereas S_e^T and S_p^T are respectively the true sensitivity and specificity values in the field. In real-life situations, it is often the case that S_e^T (S_p^T) is substantially less than S_e^P (S_p^P), because the scientist's perceived values are usually influenced by the values under the lab setting. Thus, given the true value of π , the solution to Equation (7) or (8) for p is equal to the true prevalence p_T if S_e and S_p are replaced by S_e^T and S_p^T . On the other hand, given the true value of π , replacing S_e and S_p by S_e^P and S_p^P in (7) and (8), would yield a solution p_P which is expected to deviate from the true prevalence p_T . We study the effect of the misspecification of sensitivity and specificity values on the estimate of prevalence by evaluating the bias ($p_T - p_P$).

In Tables 3a-3b, we report the values of the bias ($p_T - p_P$) resulting from the individual testing as well as Dorfman's pooled testing, denoted by $bias_I$ and $bias_D$, respectively, for different values of p , S_e^P , S_p^P , S_e^T and S_p^T . The following patterns are visible from the table:

(i) With increase in the prevalence p , the bias is decreasing. Initially, it reduces from positive to zero and then becomes negative. We report the bias for four different values of p : 0.01, 0.02, 0.05 and 0.1.

(ii) The effect of misspecification of S_e and S_p are asymmetric in nature. Misspecification of S_e has a little effect on the bias. However, the misspecification of S_p has a significantly large effect on the bias. Also, more is the deviation from the true values more is the effect on bias.

(iii) The most interesting observation is, compared to the individual testing, pooled testing reduces the bias due to misspecification of S_e and S_p substantially. This is extremely important to know given the fact that misspecification of S_e and S_p are quite common.

In Table 2, we report the values of π , the probability of testing positive, for the individual testing as well as for the Dorfman testing procedure, for different values of p , S_e and S_p . Let π_I (cf. Equation (6)) and π_D (cf. Equation (8)) denote the π value corresponding to the individual and Dorfman testing procedure, respectively. Notice from Table 2 that, like the observation (ii) made above, the value of π is significantly influenced by S_p , but not so by S_e . Also, it is important to note that, for a given value of S_p (say, 0.9) reduction in S_e (say, from 1 to 0.7) results in reduction of π , though not a significant reduction. Thus, we may conclude that misspecification of S_p (S_e) has a large (negligible) effect on the value of π , and consequently, on the estimates of prevalence obtained by solving Equations (6) or (8), as the case may be. So, for accurate estimation of prevalence it is very important to specify the value of S_p correctly, while misspecification of S_e has little effect.

Finally, we explain a subtle connection between the numbers reported in Tables 3a-3b and Table 2 with a specific example. Consider the case when prevalence p is equal to 0.01, and the

true sensitivity (S_e^T) and specificity (S_p^T) are both equal to 0.9. From Table 2, we observe that the corresponding value of π (cf. Equation (6)) for individual testing is equal to 0.11. Suppose now the perceived sensitivity (S_e^P) and specificity (S_p^P) are equal to 1. To obtain the biased prevalence p corresponding to the perceived sensitivity and specificity, we plug in 0.11 for π on the left hand side of Equation (6), and $S_e = S_p = 1$ on the right hand side. The solution can be obtained from Table 2 by observing the value of p corresponding to $S_e = S_p = 1$ and $\pi_I = 0.11$, which is clearly 0.11. Thus the resulting bias due to misspecification is $0.11 - 0.01 = 0.10$, which is reported in Table 3a. By using a similar argument we can find the bias for pooled testing which is 0.01 using Table 2.

5 Concluding Remarks

In this article, we have discussed the statistical theory behind Dorfman's pooled testing technique used for screening. We have explained the method of estimation of prevalence from individual testing and pooled testing data. Most importantly, we have provided theoretical insights into the practical issues that are being discussed in the scientific community, arising out of the fact that the tests for the SARS-CoV-2 virus are imperfect. The theoretical results show that pooled testing is not only preferable for reducing time and cost of screening, it also helps in significant reduction of misclassification among those who are tested positive. Our results also show that for prevalence estimation, pooled testing is always preferable to individual testing, especially when the specificity of the test is low. It helps in reducing the bias of the prevalence estimate significantly. Finally, it is worth mentioning that our observations are valid for low values of prevalence, possibly less than or equal to 10%.

In this connection, as a final remark, we may conclude that for estimating the prevalence of SARS-CoV-2 of Santa Clara County, California¹⁶, the Stanford scientists collected data by conducting individual antibody tests on the 3300 subjects. Instead, had they used pooled testing technique, it would have been possible to collect the test data from a much larger sample for a similar cost and time. Also, as mentioned in the paper, given that the prevalence is low (estimated as 2.8%), adopting pooled testing technique would have made all the more sense in the context of their study.

Appendix

A Computation of S_e^D and S_p^D

In the context of the Dorfman test, S_e^D , the probability that a positive person can come positive is the the probability that both the pooled test as well as the individual test comes correct, which is S_e^2 .

To compute the misclassification probability that a person is non-diseased and is found to be diseased, we first note that a non-diseased person, say Person j , can either be a part of a pool constituted entirely of non-diseased individuals (Case A) or a pool containing some diseased individuals (Case B). The probability of Case A is $(1-p)^k$, and the probability of Case B is $(1-p) - (1-p)^k$.

Now, for Person j to be found to be diseased in Case A, the pooled test, and subsequently the individual's test, will both erroneously give positive result, the probability of which is $(1 - S_p)^2$ for the original Dorfman algorithm. On the other hand, in Case B, the pooled test will be positive, but this time correctly, while the individual's test will still be erroneously positive, the probability

of which is $S_e(1 - S_p)$. Putting all together, the probability that a non-diseased person is found to be diseased is $[(1 - S_p)^2(1 - p)^k + S_e(1 - S_p)((1 - p) - (1 - p)^k)]/(1 - p) = (1 - S_p)S_e + (1 - S_p)(1 - S_p - S_e)(1 - p)^{k-1}$. Hence, S_p^D is $1 - (1 - S_p)S_e + (1 - S_p)(1 - S_p - S_e)(1 - p)^{k-1}$.

References

1. Aguilar, J., Faust, J., Westafer, L., Gutierrez, J.(2020). Investigating the Impact of Asymptomatic Carriers on COVID-19 Transmission. *medRxiv* doi: <https://doi.org/10.1101/2020.03.18.20037994>
2. Asymptomatic spread of COVID-19 makes it harder to contain (2020 May 13). Retrieved from <https://medicalxpress.com/news/2020-05-asymptomatic-covid-harder.html>
3. Yu, Y., Liu, Y., Luo, F., Tu, W., Zhan, D., Yu, G., and Zhou, Z. (2020). COVID-19 Asymptomatic Infection Estimation. *medRxiv* doi: <https://doi.org/10.1101/2020.04.19.20068072>
4. Cheung, H.(2020, March 21). Coronavirus: What could the West learn from Asia? *BBC News*. Retrieved from <https://www.bbc.com/news/world-asia-51970379>
5. Normile, D. (2020, March 17). Coronavirus cases have dropped sharply in South Korea. What's the secret to its success. *Science*. Retrieved from <https://www.sciencemag.org/news/2020/03/coronavirus-cases-have-dropped-sharply-south-korea-whats-secret-its-success#>
6. Rolander, N.,(2020, April 20). Sweden says controversial COVID-19 strategy is proving effective. *The Economic Times*. Retrieved from <https://economictimes.indiatimes.com/news/international/world-news/sweden-says-controversial-covid-19-strategy-is-proving-effective/articleshow/75242644.cms?from=mdr>
7. Alluri, A., and Pathi, K.(2020, April 20).India coronavirus: Should people pay for their own COVID-19 tests? *BBC News*. Retrieved from <https://www.bbc.com/news/world-asia-india-52322559>
8. Bilder, C., Iwen, P., Abdalhamid, B., Tebbs J. and McMahan C. (2020, April 9) Increasing testing capacity for SARS-CoV-2 by pooling specimens. *Significance*. Retrieved from <https://www.significancemagazine.com/science/651-increasing-testing-capacity-for-SARS-CoV-2-by-pooling-specimens>.
9. Hanel, R., and Thurner, S. (2020). Boosting test-efficiency by pooled testing strategies for Sars-CoV-2. *arXiv preprint arXiv:2003.09944*.
10. Lakdawalla, D., Keeler, E., Goldman, D. and Trish, E. (2020). Getting Americans back to work (and School) with pooled testing. *White paper*, Leonard D. Schaeffer Center for Health Policy & Economics, University of Southern California.
11. Pouwels, K. B., Roope, L., Barnett, A., Hunter, D. J., Nolan, T. M., & Clarke, P. M. (2020). Group Testing for Sars-CoV-2: Forward to the Past?. *PharmacoEconomics* - open, 1–4. Advance online publication. <https://doi.org/10.1007/s41669-020-00217-8>
12. Dorfman, R. (1943). The detection of defective members of large populations. *The Annals of Mathematical Statistics*,14, pp 436–440.
13. SNS (2020, April 25) ICMR approves IIT-Delhi's COVID-19 test, can significantly reduce cost. *The Statesman*. Retrieved from <https://www.thestatesman.com/coronavirus/icmr-approves-iit-delhis-covid-19-test-can-significantly-reduce-cost-1502880185.html>

14. ICMR (2020). Advisory on feasibility of using pooled samples for molecular testing of COVID-19. (accessed 8 May 2020) <https://www.mohfw.gov.in/pdf/letterregguidanceonpoolingsamplesfortesting001.pdf>
15. Roy, S., and Banerjee, T. (2019). Estimation of log-odds ratio from group testing data using Firth correction. *Biometrical Journal*, 61, pp. 714–728.
16. Bendavid, E., Mulaney, B., Sood, N., Shah, S., Ling, E., Bromley-Dulfano, R., Lai, C., Weissberg, Z., Saavedra, R., Tedrow, J. and Tversky, D. (2020). COVID-19 Antibody Seroprevalence in Santa Clara County, California. *medRxiv*.
17. Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490), pp. 489–493.
18. World Health Organization. (2020). WHO Director-General’s opening remarks at the media briefing on COVID-19-11 March 2020. Geneva, Switzerland. Retrieved from <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---3-march-2020>
19. Sempos, C., and Tian, L. (2020). Adjusting Coronavirus prevalence estimates for laboratory test kit error. *medRxiv*.
20. Feller, W. (1968). An introduction to probability theory and its applications. Vol. 1. John Wiley & Sons.
21. Lewis, J. L., Lockary, V. M., and Kobic, S. (2012). Cost savings and increased efficiency using a stratified specimen pooling strategy for Chlamydia trachomatis and Neisseria gonorrhoeae. *Sexually transmitted diseases*, 39, pp. 46–48.
22. Litvak, E., Tu, X. M., and Pagano, M. (1994). Screening for the presence of a disease by pooling sera samples. *Journal of the American Statistical Association*, 89, 424-434.
23. Wein, L.M. and Zenios, S.A. (1996) Pooled testing for HIV screening: capturing the dilution effect. *Operations Research*, 44, pp. 543-569.
24. Aprahamian, H., Bish, D.R. and Bish, E.K. (2019). Optimal risk-based group testing. *Management Science*, 65, pp. 4365–4384.
25. Bilder, C. (2019). Group Testing for Estimation, *Wiley StatsRef: Statistics Reference Online* 10.1002/9781118445112.stat08231, (1-11).
26. Abdalhamid, B., Bilder, C. R., McCutchen, E. L., Hinrichs, S. H., Koepsell, S. A., and Iwen, P. C. (2020). Assessment of specimen pooling to conserve SARS CoV-2 testing resources. *American Journal of Clinical Pathology*, 153, pp. 715–718.
27. Lohse, S., Pfuhl, T., Berkó-Göttel, B., Rissland, J., Geißler, T., Gärtner, B., Becker, S.L., Schneitler, S. and Smola, S., (2020). Pooling of samples for testing for Sars-CoV-2 in asymptomatic people. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30362-5](https://doi.org/10.1016/S1473-3099(20)30362-5)
28. Yelin, I., Aharony, N., Shaer-Tamar, E., Argoetti, A., Messer, E., Berenbaum, D., Shafran, E., Kuzli, A., Gandali, N., Hashimshony, T. and Mandel-Gutfreund, Y., (2020). Evaluation of COVID-19 RT-qPCR test in multi-sample pools. *medRxiv*.

29. Ferran, M. (2020, May 6) Coronavirus tests are pretty accurate, but far from perfect. The Conversation. Retrieved from <https://theconversation.com/coronavirus-tests-are-pretty-accurate-but-far-from-perfect-136671>
30. Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D., and Lessler, J. (2020). Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Annals of Internal Medicine*.
31. Cohen, A. N., and Kessel, B. (2020). False positives in reverse transcription PCR testing for SARS-CoV-2. medRxiv.
32. Ioannidis, J. P. (2020, March 17). A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data. *Stat*, 17. Retrieved from <https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>
33. Thacker, T.,(2020, April 21).No symptoms in 80% of Covid cases raise concerns. *The Economic Times*. Retrieved from <https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/no-symptoms-in-80-of-covid-cases-raise-concerns/articleshow/75260387.cms>
34. John, T. (2020, April 3). Iceland lab’s testing suggests 50% of coronavirus cases have no symptoms. CNN. Retrieved from <https://edition.cnn.com/2020/04/01/europe/iceland-testing-coronavirus-intl/index.html>
35. Arnold, A. (2020, April 21). How Many People With the Coronavirus Are Asymptomatic? *The Cut*. Retrieved from <https://www.thecut.com/2020/04/how-many-people-with-the-coronavirus-are-asymptomatic.html>
36. Regalado, A.(2020, April 17). Up to 4% of Silicon Valley is already infected with coronavirus. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/2020/04/17/1000113/up-to-4-of-silicon-valley-already-infected-with-coronavirus/>
37. Bajak, A. and Howe, J. (2020, May 14) A Study Said Covid Wasn’t That Deadly. The Right Seized It. How coronavirus research is being weaponized. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/05/14/opinion/coronavirus-research-misinformation.html>
38. McCormick (2020, April 23). Why experts are questioning two hyped antibody studies in coronavirus hotspots. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2020/apr/23/coronavirus-antibody-studies-california-stanford>

Table 1a: Optimal group size (k^*), efficiency in terms of reduction in test numbers as percentage of number of individuals tested (Eff), pooled testing sensitivity (S_e^D) and specificity (S_p^D) false positive predictive value ($FPPV$) and false negative predictive value ($FNPV$) corresponding to different prevalence rates (p), sensitivities (S_e) and specificities (S_p) for Dorfman's algorithm. False positive predictive value ($FPPV_I$) and false negative predictive value ($FNPV_I$) for individual tests are given within parentheses to facilitate comparison.

Test Quality	p	k^*	Eff (%)	S_e^D	S_p^D	$FPPV$ ($FPPV_I$)	$FNPV$ ($FNPV_I$)
$S_e = 1$ $S_p = 1$	0.01	11	80.44	1	1	0 (0)	0 (0)
	0.02	8	72.58	1	1	0 (0)	0 (0)
	0.05	5	57.38	1	1	0 (0)	0 (0)
	0.10	4	40.61	1	1	0 (0)	0 (0)
$S_e = 0.99$ $S_p = 0.99$	0.01	11	79.65	0.98	0.999	0.09 (0.50)	0.0002 (0.0001)
	0.02	8	71.87	0.98	0.999	0.07 (0.33)	0.0004 (0.0002)
	0.05	5	56.83	0.98	0.998	0.04 (0.16)	0.0010 (0.0005)
	0.10	4	40.30	0.98	0.997	0.02 (0.08)	0.0022 (0.0011)
$S_e = 0.99$ $S_p = 0.95$	0.01	11	76.07	0.98	0.993	0.41 (0.83)	0.0002 (0.0001)
	0.02	8	68.47	0.98	0.991	0.30 (0.71)	0.0004 (0.0002)
	0.05	5	53.74	0.98	0.989	0.18 (0.49)	0.0011 (0.0006)
	0.10	4	37.67	0.98	0.985	0.12 (0.31)	0.0022 (0.0012)
$S_e = 0.99$ $S_p = 0.90$	0.01	11	71.59	0.98	0.981	0.65 (0.91)	0.0002 (0.0001)
	0.02	8	64.22	0.98	0.978	0.52 (0.83)	0.0004 (0.0002)
	0.05	5	49.87	0.98	0.973	0.34 (0.66)	0.0011 (0.0006)
	0.10	4	34.39	0.98	0.966	0.24 (0.48)	0.0023 (0.0012)
$S_e = 0.99$ $S_p = 0.80$	0.01	12	62.69	0.98	0.943	0.85 (0.95)	0.0002 (0.0001)
	0.02	9	55.75	0.98	0.936	0.76 (0.91)	0.0004 (0.0003)
	0.05	6	42.41	0.98	0.924	0.59 (0.79)	0.0011 (0.0007)
	0.10	4	27.83	0.98	0.917	0.43 (0.65)	0.0024 (0.0014)
$S_e = 0.99$ $S_p = 0.70$	0.01	13	53.86	0.98	0.886	0.92 (0.97)	0.0002 (0.0001)
	0.02	9	47.42	0.98	0.879	0.86 (0.94)	0.0005 (0.0003)
	0.05	6	35.05	0.98	0.863	0.73 (0.85)	0.0012 (0.0008)
	0.10	5	21.74	0.98	0.839	0.60 (0.73)	0.0026 (0.0016)
$S_e = 0.95$ $S_p = 0.99$	0.01	11	80.07	0.90	0.999	0.10 (0.51)	0.0010 (0.0005)
	0.02	8	72.47	0.90	0.999	0.07 (0.34)	0.0020 (0.0010)
	0.05	5	57.74	0.90	0.998	0.04 (0.17)	0.0051 (0.0027)
	0.10	4	41.67	0.90	0.997	0.03 (0.09)	0.0107 (0.0056)
$S_e = 0.95$ $S_p = 0.95$	0.01	11	76.49	0.90	0.993	0.43 (0.84)	0.0010 (0.0005)
	0.02	8	69.07	0.90	0.992	0.31 (0.72)	0.0020 (0.0011)
	0.05	5	54.64	0.90	0.989	0.19 (0.50)	0.0052 (0.0028)
	0.10	4	39.05	0.90	0.985	0.13 (0.32)	0.0109 (0.0058)
$S_e = 0.95$ $S_p = 0.90$	0.01	11	72.01	0.90	0.982	0.67 (0.91)	0.0010 (0.0006)
	0.02	8	64.81	0.90	0.979	0.54 (0.84)	0.0020 (0.0011)
	0.05	6	50.82	0.90	0.971	0.38 (0.67)	0.0053 (0.0029)
	0.10	4	35.77	0.90	0.967	0.25 (0.49)	0.0111 (0.0061)
$S_e = 0.95$ $S_p = 0.80$	0.01	12	63.15	0.90	0.944	0.86 (0.95)	0.0010 (0.0006)
	0.02	9	56.42	0.90	0.938	0.77 (0.91)	0.0021 (0.0013)
	0.05	6	43.47	0.90	0.926	0.61 (0.80)	0.0055 (0.0033)
	0.10	5	29.29	0.90	0.908	0.48 (0.65)	0.0118 (0.0069)
$S_e = 0.95$ $S_p = 0.70$	0.01	13	54.35	0.90	0.888	0.92 (0.97)	0.0011 (0.0007)
	0.02	10	48.11	0.90	0.878	0.87 (0.94)	0.0023 (0.0015)
	0.05	6	36.11	0.90	0.866	0.74 (0.86)	0.0059 (0.0037)
	0.10	5	23.38	0.90	0.843	0.61 (0.74)	0.0127 (0.0079)

Table 1b: Optimal group size (k^*), efficiency in terms of reduction in test numbers as percentage of number of individuals tested (Eff), pooled testing sensitivity (S_e^D) and specificity (S_p^D) false positive predictive value ($FPPV$) and false negative predictive value ($FNPV$) corresponding to different prevalence rates (p), sensitivities (S_e) and specificities (S_p) for Dorfman's algorithm (Continued).

Test Quality	p	k^*	Eff (%)	S_e^D	S_p^D	$FPPV$ ($FPPV_I$)	$FNPV$ ($FNPV_I$)
$S_e = 0.90$ $S_p = 0.99$	0.01	11	80.59	0.81	0.999	0.10 (0.52)	0.0019 (0.0010)
	0.02	8	73.22	0.81	0.999	0.07 (0.35)	0.0039 (0.0021)
	0.05	5	58.87	0.81	0.998	0.04 (0.17)	0.0099 (0.0053)
	0.10	4	43.39	0.81	0.997	0.03 (0.09)	0.0207 (0.0111)
$S_e = 0.90$ $S_p = 0.95$	0.01	11	77.01	0.81	0.993	0.45 (0.85)	0.0019 (0.0011)
	0.02	8	69.81	0.81	0.992	0.33 (0.73)	0.0039 (0.0021)
	0.05	6	55.82	0.81	0.988	0.22 (0.51)	0.0100 (0.0055)
	0.10	4	40.77	0.81	0.986	0.13 (0.33)	0.021 (0.0116)
$S_e = 0.90$ $S_p = 0.90$	0.01	12	72.58	0.81	0.982	0.69 (0.92)	0.0020 (0.0011)
	0.02	9	65.59	0.81	0.978	0.57 (0.84)	0.0039 (0.0023)
	0.05	6	52.14	0.81	0.972	0.40 (0.68)	0.0102 (0.0058)
	0.10	4	37.49	0.81	0.968	0.26 (0.5)	0.0213 (0.0122)
$S_e = 0.90$ $S_p = 0.80$	0.01	13	63.73	0.81	0.944	0.87 (0.96)	0.0020 (0.0013)
	0.02	9	57.25	0.81	0.939	0.79 (0.92)	0.0041 (0.0025)
	0.05	6	44.79	0.81	0.928	0.63 (0.81)	0.0107 (0.0065)
	0.10	5	31.33	0.81	0.912	0.49 (0.67)	0.0226 (0.0137)
$S_e = 0.90$ $S_p = 0.70$	0.01	14	54.98	0.81	0.888	0.93 (0.97)	0.0022 (0.0014)
	0.02	10	49.02	0.81	0.880	0.88 (0.94)	0.0044 (0.0029)
	0.05	7	37.61	0.81	0.862	0.76 (0.86)	0.0115 (0.0075)
	0.10	5	25.43	0.81	0.848	0.63 (0.75)	0.0243 (0.0156)
$S_e = 0.80$ $S_p = 0.99$	0.01	12	81.69	0.64	0.999	0.13 (0.55)	0.0036 (0.0020)
	0.02	9	74.75	0.64	0.999	0.09 (0.38)	0.0073 (0.0041)
	0.05	6	61.41	0.64	0.998	0.05 (0.19)	0.0186 (0.0105)
	0.10	4	46.83	0.64	0.998	0.03 (0.10)	0.0385 (0.0220)
$S_e = 0.80$ $S_p = 0.95$	0.01	12	78.15	0.64	0.994	0.50 (0.86)	0.0036 (0.0021)
	0.02	9	71.42	0.64	0.992	0.38 (0.75)	0.0074 (0.0043)
	0.05	6	58.47	0.64	0.989	0.25 (0.54)	0.0188 (0.0110)
	0.10	5	44.29	0.64	0.985	0.18 (0.36)	0.0390 (0.0229)
$S_e = 0.80$ $S_p = 0.90$	0.01	13	73.73	0.64	0.982	0.74 (0.93)	0.0037 (0.0022)
	0.02	9	67.25	0.64	0.980	0.61 (0.86)	0.0074 (0.0045)
	0.05	6	54.79	0.64	0.974	0.43 (0.70)	0.0191 (0.0116)
	0.10	5	41.33	0.64	0.966	0.32 (0.53)	0.0398 (0.0241)
$S_e = 0.80$ $S_p = 0.80$	0.01	14	64.98	0.64	0.945	0.89 (0.96)	0.0038 (0.0025)
	0.02	10	59.02	0.64	0.940	0.82 (0.92)	0.0078 (0.0051)
	0.05	7	47.61	0.64	0.928	0.68 (0.83)	0.0200 (0.0130)
	0.10	5	35.43	0.64	0.919	0.53 (0.69)	0.0417 (0.0270)
$S_e = 0.80$ $S_p = 0.70$	0.01	15	56.34	0.64	0.890	0.94 (0.97)	0.0041 (0.0029)
	0.02	11	50.95	0.64	0.883	0.90 (0.95)	0.0083 (0.0058)
	0.05	8	40.67	0.64	0.865	0.80 (0.88)	0.0214 (0.0148)
	0.10	6	29.91	0.64	0.849	0.68 (0.77)	0.0450 (0.0308)

Table 2: Expected values of the proportion of sample testing positive (π) corresponding to different prevalence rates, sensitivities (S_e) and specificities (S_p) for individual testing (π_{INDIV}) Dorfman's algorithm (π_{DORF}). The π_{DORF} values are based on $k = 4$, the most conservative pooling choice from Tables 1a and 1b.

Test Quality	p	π_{INDIV}	π_{DORF}	Test Quality	p	π_{INDIV}	π_{DORF}
$S_e = 1$ $S_p = 1$	0.01	0.01	0.01	$S_e = 0.80$ $S_p = 1$	0.01	0.01	0.01
	0.02	0.02	0.02		0.02	0.02	0.01
	0.05	0.05	0.05		0.05	0.04	0.03
	0.10	0.10	0.10		0.10	0.08	0.06
$S_e = 1$ $S_p = 0.90$	0.01	0.11	0.02	$S_e = 0.80$ $S_p = 0.90$	0.01	0.11	0.02
	0.02	0.12	0.03		0.02	0.11	0.03
	0.05	0.14	0.07		0.05	0.13	0.05
	0.10	0.19	0.13		0.10	0.17	0.09
$S_e = 1$ $S_p = 0.80$	0.01	0.21	0.05	$S_e = 0.80$ $S_p = 0.80$	0.01	0.21	0.05
	0.02	0.22	0.07		0.02	0.21	0.06
	0.05	0.24	0.11		0.05	0.23	0.09
	0.10	0.28	0.18		0.10	0.26	0.13
$S_e = 1$ $S_p = 0.70$	0.01	0.31	0.11	$S_e = 0.80$ $S_p = 0.70$	0.01	0.31	0.10
	0.02	0.31	0.12		0.02	0.31	0.11
	0.05	0.34	0.16		0.05	0.33	0.14
	0.10	0.37	0.23		0.10	0.35	0.18
$S_e = 0.90$ $S_p = 1$	0.01	0.01	0.01	$S_e = 0.70$ $S_p = 1$	0.01	0.01	0
	0.02	0.02	0.02		0.02	0.01	0.01
	0.05	0.05	0.04		0.05	0.03	0.02
	0.10	0.09	0.08		0.10	0.07	0.05
$S_e = 0.90$ $S_p = 0.90$	0.01	0.11	0.02	$S_e = 0.70$ $S_p = 0.90$	0.01	0.11	0.02
	0.02	0.12	0.03		0.02	0.11	0.02
	0.05	0.14	0.06		0.05	0.13	0.04
	0.10	0.18	0.11		0.10	0.16	0.07
$S_e = 0.90$ $S_p = 0.80$	0.01	0.21	0.05	$S_e = 0.70$ $S_p = 0.80$	0.01	0.20	0.05
	0.02	0.21	0.06		0.02	0.21	0.05
	0.05	0.23	0.10		0.05	0.22	0.08
	0.10	0.27	0.15		0.10	0.25	0.11
$S_e = 0.90$ $S_p = 0.70$	0.01	0.31	0.10	$S_e = 0.70$ $S_p = 0.70$	0.01	0.30	0.10
	0.02	0.31	0.11		0.02	0.31	0.10
	0.05	0.33	0.15		0.05	0.32	0.13
	0.10	0.36	0.21		0.10	0.34	0.16

Table 3a: Estimation results when perceived sensitivity (S_e^P) and specificity (S_p^P) are higher than the true situations. Bias $bias_I$ corresponds to individual tests, while $bias_D$ corresponds to Dorfman tests. The Dorfman values are based on $k = 4$, the most conservative pooling choice from Tables 1a and 1b.

Perceived test quality	p	$S_e^T = S_e^P$ $S_p^T = S_p^P$ 10% less		S_e^T 10% less $S_p^T = S_p^P$		Both S_e^T and S_p^T 10% less		Both S_e^T and S_p^T 20% less	
		$bias_I$	$bias_D$	$bias_I$	$bias_D$	$bias_I$	$bias_D$	$bias_I$	$bias_D$
$S_e^P = 1$ $S_p^P = 1$	0.01	0.10	0.01	0	0	0.10	0.01	0.20	0.04
	0.02	0.10	0.01	0	0	0.10	0.01	0.19	0.04
	0.05	0.10	0.02	-0.01	-0.01	0.09	0.01	0.18	0.04
	0.10	0.09	0.03	-0.01	-0.02	0.08	0.01	0.16	0.03
$S_e^P = 0.99$ $S_p^P = 0.99$	0.01	0.10	0.01	0	0	0.10	0.01	0.20	0.04
	0.02	0.10	0.02	0	0	0.10	0.01	0.20	0.04
	0.05	0.10	0.02	-0.01	-0.01	0.09	0.01	0.18	0.04
	0.10	0.09	0.03	-0.01	-0.02	0.08	0.01	0.16	0.03
$S_e^P = 0.99$ $S_p^P = 0.95$	0.01	0.11	0.02	0	0	0.10	0.02	0.21	0.05
	0.02	0.10	0.02	0	0	0.10	0.02	0.20	0.05
	0.05	0.10	0.03	-0.01	-0.01	0.10	0.02	0.19	0.05
	0.10	0.10	0.04	-0.01	-0.02	0.09	0.01	0.17	0.04
$S_e^P = 0.99$ $S_p^P = 0.90$	0.01	0.11	0.03	0	0	0.11	0.02	0.22	0.06
	0.02	0.11	0.03	0	0	0.11	0.02	0.22	0.06
	0.05	0.11	0.03	-0.01	-0.01	0.10	0.02	0.20	0.06
	0.10	0.10	0.04	-0.01	-0.02	0.09	0.02	0.18	0.05
$S_e^P = 0.99$ $S_p^P = 0.80$	0.01	0.13	0.04	0	0	0.12	0.04	0.25	0.09
	0.02	0.12	0.04	0	0	0.12	0.03	0.24	0.08
	0.05	0.12	0.04	-0.01	-0.01	0.11	0.03	0.23	0.08
	0.10	0.11	0.05	-0.01	-0.02	0.10	0.03	0.20	0.06
$S_e^P = 0.99$ $S_p^P = 0.70$	0.01	0.14	0.05	0	0	0.14	0.05	0.28	0.11
	0.02	0.14	0.05	0	0	0.14	0.05	0.28	0.11
	0.05	0.14	0.05	-0.01	-0.01	0.13	0.04	0.26	0.10
	0.10	0.13	0.06	-0.01	-0.02	0.12	0.03	0.23	0.08
$S_e^P = 0.95$ $S_p^P = 0.99$	0.01	0.11	0.02	0	0	0.10	0.01	0.21	0.05
	0.02	0.10	0.02	0	0	0.10	0.01	0.20	0.05
	0.05	0.10	0.02	-0.01	-0.01	0.10	0.01	0.19	0.04
	0.10	0.10	0.03	-0.01	-0.02	0.09	0.01	0.17	0.03
$S_e^P = 0.95$ $S_p^P = 0.95$	0.01	0.11	0.02	0	0	0.11	0.02	0.22	0.06
	0.02	0.11	0.02	0	0	0.11	0.02	0.21	0.06
	0.05	0.11	0.03	-0.01	-0.01	0.10	0.02	0.20	0.05
	0.10	0.10	0.04	-0.01	-0.02	0.09	0.01	0.18	0.04
$S_e^P = 0.95$ $S_p^P = 0.90$	0.01	0.12	0.03	0	0	0.12	0.03	0.23	0.07
	0.02	0.12	0.03	0	0	0.11	0.03	0.23	0.07
	0.05	0.11	0.03	-0.01	-0.01	0.11	0.02	0.21	0.06
	0.10	0.11	0.04	-0.01	-0.02	0.09	0.02	0.19	0.05
$S_e^P = 0.95$ $S_p^P = 0.80$	0.01	0.13	0.04	0	0	0.13	0.04	0.26	0.09
	0.02	0.13	0.04	0	0	0.13	0.04	0.26	0.09
	0.05	0.13	0.04	-0.01	-0.01	0.12	0.03	0.24	0.08
	0.10	0.12	0.05	-0.01	-0.02	0.11	0.03	0.21	0.06
$S_e^P = 0.95$ $S_p^P = 0.70$	0.01	0.15	0.05	0	0	0.15	0.05	0.30	0.12
	0.02	0.15	0.05	0	0	0.15	0.05	0.30	0.12
	0.05	0.15	0.06	-0.01	-0.01	0.14	0.04	0.28	0.11
	0.10	0.14	0.06	-0.02	-0.02	0.12	0.04	0.25	0.08

Table 3b: Estimation results when perceived sensitivity (S_e^P) and specificity (S_p^P) are higher than the true situations. Bias $bias_I$ corresponds to individual tests, while $bias_D$ corresponds to Dorfman tests (Continued).

Perceived test quality	p	$S_e^T = S_e^P$ S_p^T 10% less		S_e^T 10% less $S_p^T = S_p^P$		Both S_e^T and S_p^T 10% less		Both S_e^T and S_p^T 20% less	
		$bias_I$	$bias_D$	$bias_I$	$bias_D$	$bias_I$	$bias_D$	$bias_I$	$bias_D$
		$S_e^P = 0.90$ $S_p^P = 0.99$	0.01 0.02 0.05 0.10	0.11 0.11 0.11 0.10	0.02 0.02 0.03 0.04	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.11 0.11 0.10 0.09	0.01 0.01 0.01 0.01
$S_e^P = 0.90$ $S_p^P = 0.95$	0.01 0.02 0.05 0.10	0.12 0.12 0.11 0.11	0.02 0.03 0.03 0.04	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.12 0.11 0.11 0.09	0.02 0.02 0.02 0.02	0.23 0.23 0.21 0.19	0.06 0.06 0.06 0.04
$S_e^P = 0.90$ $S_p^P = 0.90$	0.01 0.02 0.05 0.10	0.12 0.12 0.12 0.11	0.03 0.03 0.04 0.04	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.12 0.12 0.11 0.10	0.03 0.03 0.03 0.02	0.25 0.24 0.22 0.20	0.08 0.08 0.07 0.05
$S_e^P = 0.90$ $S_p^P = 0.80$	0.01 0.02 0.05 0.10	0.14 0.14 0.14 0.13	0.04 0.05 0.05 0.05	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.14 0.14 0.13 0.11	0.04 0.04 0.04 0.03	0.28 0.27 0.26 0.23	0.11 0.10 0.09 0.07
$S_e^P = 0.90$ $S_p^P = 0.70$	0.01 0.02 0.05 0.10	0.16 0.16 0.16 0.15	0.06 0.06 0.06 0.07	0 0 -0.01 -0.02	0 0 -0.01 -0.02	0.16 0.16 0.15 0.13	0.06 0.06 0.05 0.04	0.33 0.32 0.30 0.27	0.14 0.13 0.12 0.09
$S_e^P = 0.80$ $S_p^P = 0.99$	0.01 0.02 0.05 0.10	0.13 0.12 0.12 0.11	0.02 0.02 0.03 0.04	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.12 0.12 0.11 0.10	0.02 0.02 0.02 0.01	0.25 0.24 0.23 0.20	0.07 0.06 0.06 0.05
$S_e^P = 0.80$ $S_p^P = 0.95$	0.01 0.02 0.05 0.10	0.13 0.13 0.13 0.12	0.03 0.03 0.04 0.05	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.13 0.13 0.12 0.11	0.03 0.03 0.02 0.02	0.26 0.26 0.24 0.21	0.08 0.08 0.07 0.05
$S_e^P = 0.80$ $S_p^P = 0.90$	0.01 0.02 0.05 0.10	0.14 0.14 0.14 0.13	0.04 0.04 0.05 0.05	0 0 -0.01 -0.01	0 0 -0.01 -0.02	0.14 0.14 0.13 0.11	0.04 0.03 0.03 0.03	0.28 0.27 0.26 0.23	0.10 0.09 0.08 0.07
$S_e^P = 0.80$ $S_p^P = 0.80$	0.01 0.02 0.05 0.10	0.16 0.16 0.16 0.15	0.06 0.06 0.06 0.07	0 0 -0.01 -0.02	0 0 -0.01 -0.02	0.16 0.16 0.15 0.13	0.05 0.05 0.05 0.04	0.33 0.32 0.30 0.27	0.13 0.13 0.11 0.09
$S_e^P = 0.80$ $S_p^P = 0.70$	0.01 0.02 0.05 0.10	0.20 0.20 0.19 0.18	0.08 0.08 0.08 0.08	0 0 -0.01 -0.02	0 0 -0.01 -0.03	0.20 0.19 0.18 0.16	0.07 0.07 0.06 0.05	0.39 0.38 0.36 0.32	0.18 0.18 0.16 0.12