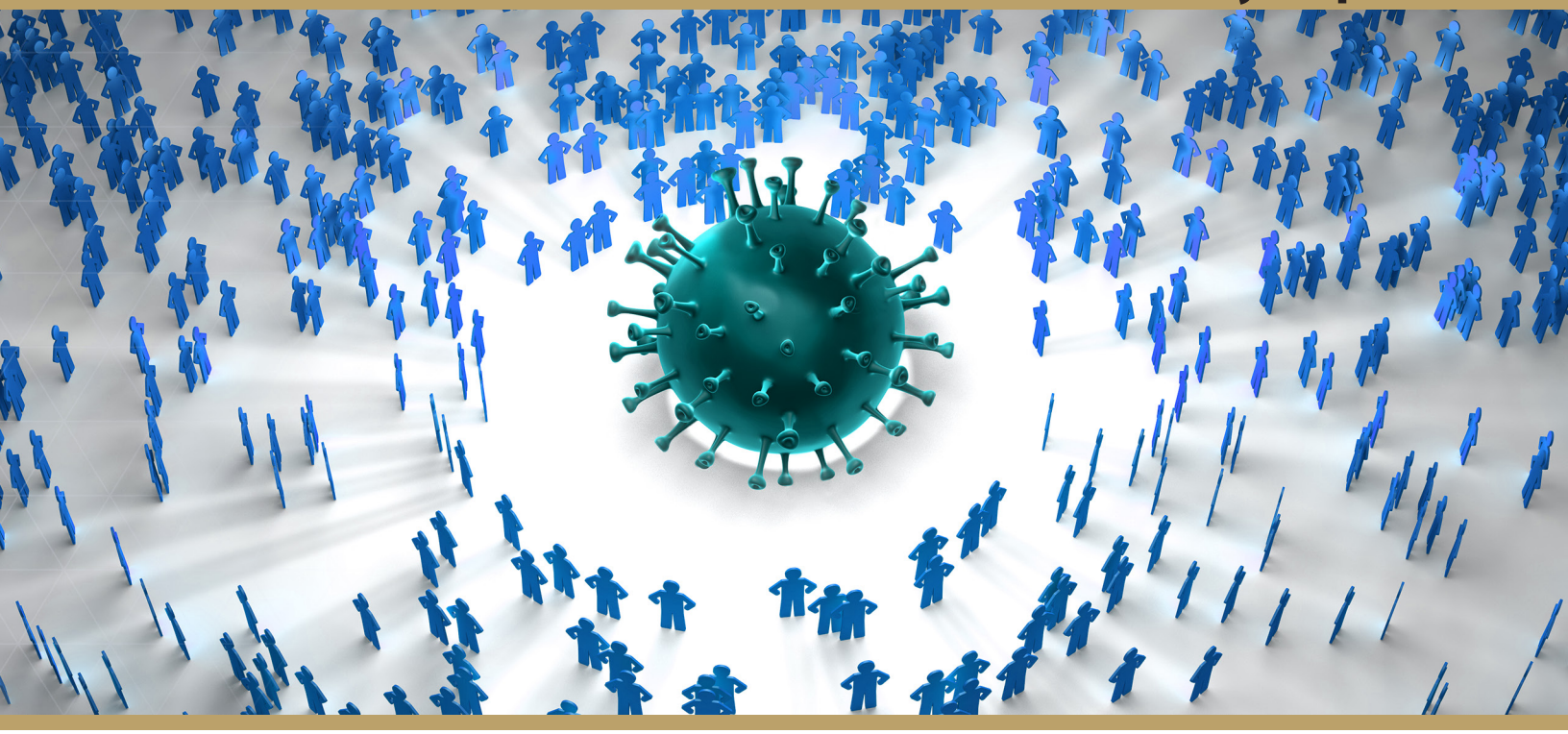
OPERATIONAL ANALYSIS FOR CORONAVIRUS TESTING

National Security Report



Marc Mangel | Alan Brown

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

OPERATIONAL ANALYSIS FOR CORONAVIRUS TESTING

Marc Mangel

Alan Brown

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

NSAD-R-21-041

Contents

Figures

Summary

Testing will remain a key tool for those managing health care and making health policy for the current coronavirus pandemic, and testing will probably be an important tool in future pandemics. Because of test errors (false negative tests in which an infected individual tests as uninfected and false positive tests in which an uninfected individual tests as infected), the observed fraction of positive tests out of a total of T tests, the surface positivity, is generally different from the underlying incidence rate of the disease.

In a companion report,[1] we describe a method for translating from the surface positivity to a point estimate for the incidence rate, then to an appropriate range of values for the incidence rate (the test range), and finally to the risk (the probability of including one infected individual) associated with groups of different sizes. Three key messages of that report are (1) surface positivity is not an accurate indicator of the incidence of coronavirus; (2) false negative tests lead to overestimation of the incidence rate, and false positive tests lead to underestimation of the incidence rate; and (3) the risk of groups of different sizes is not an either-or situation but can be graded according to the incidence rate, the size of the group, and the specified level of tolerance for risk.

The main purpose of this report is to provide supporting analysis for the recommendations for practice given in our companion report,[2] summarized in Equations 1–4 in this report, and to provide an additional recommendation for practice. To do so, we model the process generating test data in which the true state of the world (incidence rate, probability of a false negative test, and probability of a false positive test) is assumed to be known. This allows us to compare analytical predictions with a known situation. We begin by showing how to compute the risk associated with groups of different sizes (defined to be the probability of including at least one infected individual) when one has an estimate for incidence rate.

When test errors are known, we show that surface positivity can be a very poor proxy for the underlying incidence rate and that the estimate for incidence rate in our companion report is the maximum likelihood estimate (MLE). We show how to calculate test range directly from the simulated data and that the approximation given in our companion report is very accurate if the number of tests is sufficiently large. We evaluate the MLE for test numbers ranging from 50 to 5,000 and show that its mean is an accurate estimate (close to the true value of the incidence rate) for even a modest number of tests, but that its variance is so large for a modest number of tests that the MLE is very imprecise. However, the standard deviation of the MLEs and the test range decline as $1/\sqrt{T}$ so that with a sufficient number of tests our method[3] gives both an accurate and precise prediction. Because the

[1] Brown and Mangel, *Recommendations for Practice*.

[2] Brown and Mangel, *Recommendations for Practice*.

[3] Brown and Mangel, *Recommendations for Practice*.

test range declines with the number of tests, it is possible to oversample spatial regions by allocating too many tests. When test errors are not known, we generate data using the true test errors and incidence rate but compute the estimate of incidence rate with test errors that may differ from the true ones. Using the simulation, we show that when the choice of probability of a false negative test exceeds the true value, one overestimates the true incidence rate. Similarly, when the choice of the probability of a false positive test used in computing the estimate of incidence rate exceeds the true value, one underestimates the true incidence rate. Using a certainty equivalent (CE) approximation, in which we replace stochastic surface positivity by its mean, we analytically confirm these observations from the simulation. In addition, using the certainty equivalent approximation, we show how to include distributions of test errors in the construction of the estimate of incidence rate.[4] We also explicitly consider the case in which surface positivity is 0, which is likely to happen during testing as the pandemic wanes, and derive a formula for the maximum incidence rate consistent with no positive tests. This estimate of the incidence rate can be used in a risk calculation. In the supplementary figures,[5] we use an additional 27 sets of parameters characterizing the true state of the world and confirm that all the qualitative patterns described in the main text remain, although some minor quantitative details change.

[4] Brown and Mangel, *Recommendations for Practice*, Appendix B.

[5] Available at https://www.jhuapl.edu/Content/figures/Mangel_Brown_SuppFigs.pdf.

Introduction, Summary of Our Companion Report, and Overview of This Report

> . . . if you don't do the best you can with what you happen to have got, you'll never do the best you might have done with what you should have had. . .
>
> —*Aris,* Discrete Dynamic Programming*, 27*

Introduction

Even though vaccines for coronavirus are increasingly available, testing remains a key tool for those managing health care and making policy decisions, and large-scale testing is probably essential for resilience to the pandemic and reopening society.[1]

In the companion to this report,[2] we provide a method for interpreting coronavirus tests, addressing the following questions:

- How does one go from surface positivity (the observed fraction of positive tests) to the incidence rate (the unobserved fraction of individuals infected with coronavirus) knowing that there are test errors (false negatives in which infected individuals give a negative test and false positives in which noninfected individuals give a positive test)?
- How does one go from a point estimate for the incidence rate to a range of reasonably likely incidence rates?
- How does one compute the risk (the probability of including one infected individual) of coronavirus transmission in groups of different sizes, given the point estimate and range of values for the incidence rate?

As Aris notes in the epigraph, one will almost never have an ideal situation for analysis of a problem but should strive to do the best possible analysis in an imperfect situation. In addition to providing insight into the current problem, effective analysis gives guidance on how to prepare for the next analogous problem.

We next briefly review relevant aspects of coronavirus infection, transmission, and testing, after which we give an overview of the entire report.

[1] Allen et al., *Roadmap to Pandemic Resilience*; Auger et al., "School Closure and COVID-19 Incidence and Mortality"; Booeshaghi et al., "Heterogeneous COVID-19 Testing Plans"; National Academies, *Reopening K–12 Schools*; and Wilson et al., "Multiple COVID-19 Clusters on a University Campus."

[2] Brown and Mangel, *Recommendations for Practice*.

Relevant Etiology of Coronavirus Infection

Coronavirus is highly infectious,[3] with a large reproductive number, R_0, and a long asymptomatic phase during which infection can be transmitted.[4] For example, He et al.[5] estimated that 44% of secondary cases occurred before the onset of symptoms and that infectiousness peaked on or before the onset of symptoms. Oran and Topol[6] estimated asymptomatic individuals on average accounted for 40–45% of coronavirus infections, with a range from 6.3% (King County, Washington, nursing facility residents) to almost 100% (inmates in Arkansas, North Carolina, Ohio, and Virginia). Shental et al.[7] estimated that the percentage of asymptomatic infected individuals was 10–30%, but Li et al.[8] estimated that asymptomatic individuals could have been responsible for 79% of documented cases.

The high R_0 and asymptomatic infections are complicated by coronavirus tests having false positive results (a noninfected individual tests positive) and false negative results (an infected individual tests negative). Kucirka et al.[9] estimated that on the day of infection the probability[10] of a false negative test is 100%, and then it falls to 67% on day 4 of infection, 38% on day 5 (assumed onset of symptoms), 20% on day 8, 21% on day 9, and 66% on day 21. (See Figure 2 of their paper for the full temporal dynamics.)

Summary of Our Companion Report

In Table 1, we list the major symbols used in the remainder of the report, their interpretation, and their first appearance in this text.

The operational situation is that T tests are administered to a population, and each individual tested has either a positive or negative test result for coronavirus. An unknown fraction f_t of these individuals are infected with coronavirus and are antigen positive, with the subscript t denoting the true but unknown value. However, such individuals have a probability p_{FN_t} of a false negative result in which the test reports no infection. The remaining individuals, a fraction $1 - f_t$ of the sample, are not infected (i.e., are antigen negative) but have a probability p_{FP_t} of testing positive.

[3] Bi et al., "Epidemiology and Transmission of COVID-19."

[4] Oran and Topol, "Prevalence of Asymptomatic SARS-CoV-2 Infection."

[5] He et al., "Viral Shedding and Transmissibility."

[6] Oran and Topol, "Prevalence of Asymptomatic SARS-CoV-2 Infection."

[7] Shental et al., "Asymptomatic Carriers."

[8] Li et al., "Substantial Undocumented Infection."

[9] Kucirka et al., "Variation in False-Negative Rate."

[10] Green et al., "Molecular Tests"; He et al., "Diagnostic Performance"; Kucirka et al., "Variation in False-Negative Rate"; and Watson, Whiting, and Brush, "Interpreting a Covid-19 Test Result."

Table 1. Variables, Symbols, Their Interpretation, and First Appearance in Text

Variable/Symbol	Interpretation	First appearance in text
Incidence rate		
f	Incidence of coronavirus infection	Summary of Our Companion Report
f_t	True but unknown incidence of coronavirus infection generating the observations	Summary of Our Companion Report
$\hat{f}$	Recommendation for practice of the estimate of incidence of coronavirus infection	Eq. 2
$\hat{f}_n$	Estimate of coronavirus infection incidence on the *n*th simulation of the testing process	Eq. 8
Test errors		
p_{FN_t}	True probability of a false negative test	Summary of Our Companion Report
p_{FP_t}	True probability of a false positive test	Summary of Our Companion Report
p_{FN}	Assumed probability of a false negative test	Summary of Our Companion Report
p_{FP}	Assumed probability of a false positive test	Summary of Our Companion Report
Test data		
T	Number of tests administered	Summary of Our Companion Report
$P, \tilde{P}$	Number of positive test results obtained	Summary of Our Companion Report
$p = P/T$	Positivity rate	Summary of Our Companion Report
$p_+(f)$	Probability that a positive test result is obtained when the incidence rate is f	Eq. 1
$Range(\hat{f})$	Range of incidence rates that is compatible with the data and model	Eq. 3
f_{lower}, f_{upper}	Smallest and largest values, respectively, for the estimate of the incidence rate compatible with the data and model	Below Eq. 3
Risk		
$\mathcal{R}(g, \hat{f})$	Probability that a group of size g contains at least one infected individual when the estimate of incidence rate is $\hat{f}$	Eq. 4
$\mathcal{R}_{acc}$	Specified level of risk that can be tolerated	Below Eq. 7
$g(\hat{f}, \mathcal{R}_{acc})$	Group size giving risk no larger than $\mathcal{R}_{acc}$	Eq. 8

(***continued***)

Table 1 (***continued***)

Variable/Symbol	Interpretation	First appearance in text
Simulation method		
N	Number of simulations	Simulating the Test Data
n	Index for the simulation, running from 1 to N	Simulating the Test Data
P_n	Number of positive tests on the *n*th iteration of the simulation	Simulating the Test Data
$\mathcal{B}(\cdot, T, p_+(f_t)$	Binomial distribution for the number of positive tests when T tests are given and the probability of a positive test is $p_+(f_t)$	Eq. 9
$\mathcal{L}(p_+(f)\|P_n, T)$	Likelihood that the probability of a positive test is $p(f)$ given P_n positive tests out of T total tests on the *n*th simulation	The Likelihood of an Incidence Rate Given the Number of Positive Tests
$\phi(f\|P_n, T)$	Probability density for the incidence rate in the nth iteration of the simulation given that P_n positive tests are obtained	Eq. 12
$\mathcal{E}(\hat{f}_n)$	Expectation of the estimate of the incidence rate, $\hat{f}_n$, on the nth replicate of the simulation	Eq. 14
$Var(\hat{f}_n)$	Variance of the estimate of the incidence rate, $\hat{f}_n$, on the nth replicate of the simulation	Eq. 16
Uncertainty in test errors		
$RE(\hat{f}(p_{FN}, p_{FP}\|p_{FN_t}, p_{FP_t})$	Relative error in the estimate of incidence rate when the true test errors are p_{FN_t} and p_{FP_t} and the assumed test errors are p_{FN} and p_{FP}	Eq. 18
Certainty equivalent (CE) approximation		
$p_{CE}(f_t, p_{FN_t}, p_{FP_t})$	Probability of a positive test uncertainty using the CE approximation that the surface positivity can be replaced by its mean	Eq. 19
$\hat{f}_{CE}(p_{FN}, p_{FP})$	CE estimate for the incidence rate, in which the surface positivity is replaced by $p_{CE}(f_t, p_{FN_t}, p_{FP_t})$	Eq. 20
Heterogeneity in test errors		
$\overline{p}_{FN}, \overline{p}_{FP}$	Means of the probabilities of a false negative or positive test	Accounting for Heterogeneity in Test Errors
$V_{pFN}, V_{pFP}, Cov(p_{FN}, p_{FP})$	Variances of the probabilities false negative or positive test and their covariance	Accounting for Heterogeneity in Test Errors

The goal is to begin with the test data, P positive results out of T tests administered, giving surface positivity P/T, and obtain an estimate $\hat{f}$ for the unobserved incidence rate of coronavirus, f_t, recognizing that f_t will never be known exactly.

The expected positivity rate $p_+(f_t)$ (the probability of a positive test) is composed of two terms: (1) the fraction of antigen-positive individuals tested who test positive (a true positive result) and (2) the fraction of antigen-negative individuals tested who test negative (a false positive result), so that

$$p_+(f_t) = f_t(1 - p_{FN}) + (1 - f_t)p_{FP}. \tag{1}$$

It is clear from this equation that the surface positivity rate is not equal to the incidence rate unless there are no test errors. The analytical challenge is that we observe the surface positivity but want to know the incidence rate. False negative tests reduce the contribution of infected individuals to the positive tests obtained in the sample, and false positive tests increase the contribution of uninfected individuals to the positive tests obtained in the sample.

From Surface Positivity to Incidence Rate

The estimate of the incidence rate $\hat{f}$ from the test results P positive tests out of a total of T tests when test errors are assumed to be p_{FN} and p_{FP} is

$$\hat{f} = \frac{P/T - p_{FP}}{1 - p_{FN} - p_{FP}}. \tag{2}$$

We show below that $\hat{f}$ given by Eq. 2 is the maximum likelihood estimate (MLE) for the incidence rate, and is to be interpreted as 0 whenever $P/T \leq p_{FP}$.

From a Point Estimate of the Incidence Rate to a Range for the Incidence Rate

Surface positivity P/T is a random variable; if one were to repeatedly sample the same population, the values of P/T would generally be different but around the expected positivity rate.

Variation in P/T generates variation in $\hat{f}$. A range of values of f that is compatible with the data and model is[11]

$$Range(\hat{f}) = 3.92\sqrt{\frac{p_+(\hat{f})(1 - p_+(\hat{f}))}{T(1 - p_{FN} - p_{FP})^2}}. \tag{3}$$

[11] In the section on the likelihood of an incidence rate, we explain our determination of range, which we interpret as a compatibility interval (McElreath, *Statistical Rethinking*, 54) and thus avoid the undesired implications of words such as *confidence* or *credible* (Morey et al., "Fallacy of Placing Confidence in Confidence Intervals").

We refer to Eq. 3 as the test range and will show that (1) $Range(\hat{f})$ is symmetrically distributed around the true range, $Range(f_t)$, which is obtained by replacing $\hat{f}$ by f_t in Eq. 3; and (2) the mean error between to the two is a fraction of a percent, so that Eq. 3 is, on average, a very accurate characterization of the range. Thus, reasonable lower and upper limits, respectively, for the estimated infection rate are $\hat{f}_{lower} = \hat{f} - 0.5 \cdot Range(\hat{f})$ and $\hat{f}_{upper} = \hat{f} + 0.5 \cdot Range(\hat{f})$.

From Incidence Rate to Risk Associated with Groups of Different Sizes

In our companion report,[12] we define the risk associated with groups of different sizes as the probability that a group of g individuals contains at least one infected individual when the estimate for incidence rate is $\hat{f}$ and denoted by $\mathcal{R}(g, \hat{f})$. This is,

$$\mathcal{R}(g, \hat{f}) = 1 - (1 - \hat{f})^g . \tag{4}$$

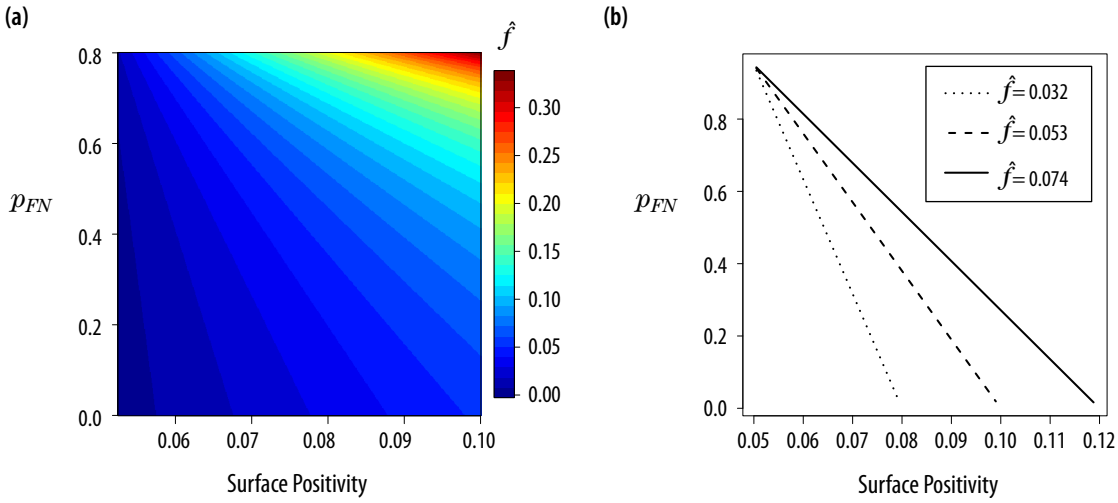
We will derive both this equation and an equation for the maximum group size consistent with a given level of risk.

An Additional Recommendation for Practice: A Caution about Relying on Positivity Alone

If p_{FP} is held constant, Eq. 2 is a relationship between the surface positivity P/T, the probability of a false negative test p_{FN}, and the estimate $\hat{f}$ of the incidence rate. Specifying any two of these will determine the third. It is natural and intuitive to specify the terms on the right-hand side of Eq. 2 and from that compute the estimate for the incidence rate. When p_{FP} is held constant, it is clear from Eq. 2 that as p_{FN} increases, the estimate of incidence rate for a given surface positivity increases, since the denominator on the right-hand side of Eq. 2 becomes smaller. Similarly, increasing surface positivity for given probability of a false negative test will increase the estimate of incidence rate because the numerator in Eq. 2 will increase. Thus, by holding p_{FP} constant, we are able to create a heat map (Figure 1a) in which the x axis is surface positivity, the y axis is p_{FN}, and the entries are $\hat{f}$ from Eq. 2.

We intentionally drew Figure 1a using only 32 colors for the heat map, because doing so suggests that the estimate of incidence rate is constant along negatively sloping lines in the plane determined by surface positivity and the probability of a false negative test. This observation begs the question of how to compare surface positivity rates. That is, is a surface positivity of 8%, say, better than a surface positivity of 12% when the test errors are not known?

[12] Brown and Mangel, *Recommendations for Practice.*



(a) When p_{FP} is held constant (in this case, $p_{FP} = 0.05$), Eq. 2 allows us to construct a heat map for the incidence rate as a function of the surface positivity (x axis) and probability of a false negative test p_{FN} (y axis). As explained in the text, the estimate of incidence rate is highest when surface positivity and the probability of a false negative test are both high. (b) Lines of constant $\hat{f}$ when $p_{FP} = 0.05$ and surface positivity is 12%, 10%, or 8% when there are no false negative tests. For any surface positivity smaller than the value corresponding to $p_{FN} = 0$, by drawing a vertical line from the value of positivity to the line of constant $\hat{f}$ and then going across to the y axis, we determine the value of p_{FN} giving the same estimate of incidence rate as when there is no chance of a false negative test.

Figure 1. Heat Map of Incidence Rate

We can also address this question using Eq. 2. To do so, we hold p_{FP} constant and imagine a surface positivity p_0 corresponding to the situation of no false negatives, so that we set $p_{FN} = 0$ in Eq. 2 to obtain the estimate for incidence rate $\hat{f}_0 = \frac{p_0 - p_{FP}}{1 - p_{FP}}$. We can obtain the value of p_{FN} needed to give an incidence rate $\hat{f}_0$ for surface positivity $p < p_0$, by using $\hat{f}_0$ on the left-hand side of Eq. 2 and p_0 for P/T on the right-hand side of Eq. 2; that is,

$$\hat{f}_0 = \frac{p - p_{FP}}{1 - p_{FN} - p_{FP}}. \tag{5}$$

Solving this equation for p_{FN}, we obtain

$$p_{FN} = 1 - p_{FP} - \left[\frac{p - p_{FP}}{\hat{f}_0}\right]. \tag{6}$$

Eq. 6 allows us to draw lines of constant $\hat{f}$ in a plane in which surface positivity is on the x axis and p_{FN} on the y axis (Figure 1b). For any surface positivity smaller than the value corresponding to $p_{FN} = 0$, by drawing a vertical line from the value of positivity to the line of constant $\hat{f}$ and then going across to the y axis, we determine the value of p_{FN} giving the same estimate of incidence rate as when there is no chance of a false negative test.

Thus, to answer the question raised above (Is surface positivity of 8% better than 12%?), we draw a vertical line from 0.08 on the x axis to where it intersects the line corresponding to $\hat{f} = 0.074$ and then draw a horizontal line to the y axis. In this case, the intersection is at p_{FN} about 0.6. If p_{FN} is this value, surface positivity of 8% is no better than 12%; if p_{FN} is less than this value, surface positivity of 8% is better than 12%; if p_{FN} is greater than this value, surface positivity of 8% is worse than 12%.

This example illustrates the danger of reporting surface positivity without reference to the test errors; since Eq. 6 is conditioned on knowing p_{FP}, both test errors are relevant.

Overall Approach

We intend this report to be as accessible as possible, but there is no way to avoid equations and mathematical analysis. The great biologist and operations analyst Baron Solly Zuckerman is reputed to have said that when he came to an equation in a paper he "hummed through it," which people often take as license to skip equations. But Zuckerman recognized the power of mathematical analysis and was no slouch in quantitative methods.[13]

We have structured the report to help a dedicated reader understand all the ideas, with the approach based on a dictum of Oliver Wendell Holmes Jr.: "I would not give a fig for the simplicity this side of complexity, but I would give my life for the simplicity on the other side of complexity."

The Operational Situation and Analytical Approach

We assume that T tests are administered to a population in which a fraction f of individuals are antigen positive, that each individual tested provides either a positive or negative test result for coronavirus (Figure 2), and that a total of P positive tests are obtained. Infected individuals have a probability p_{FN} of a false negative result in which the test reports no infection. We assume that the probability of a false negative test is known.[14] Individuals who are not infected (i.e., are antigen negative) have a probability p_{FP} of a positive test result. We also assume that the probability of a false positive test is known.[15] The values of p_{FP} and p_{FN} will generally be both medically and operationally heterogeneous; we discuss how to approach heterogeneity in the section on accounting for heterogeneity in test errors.

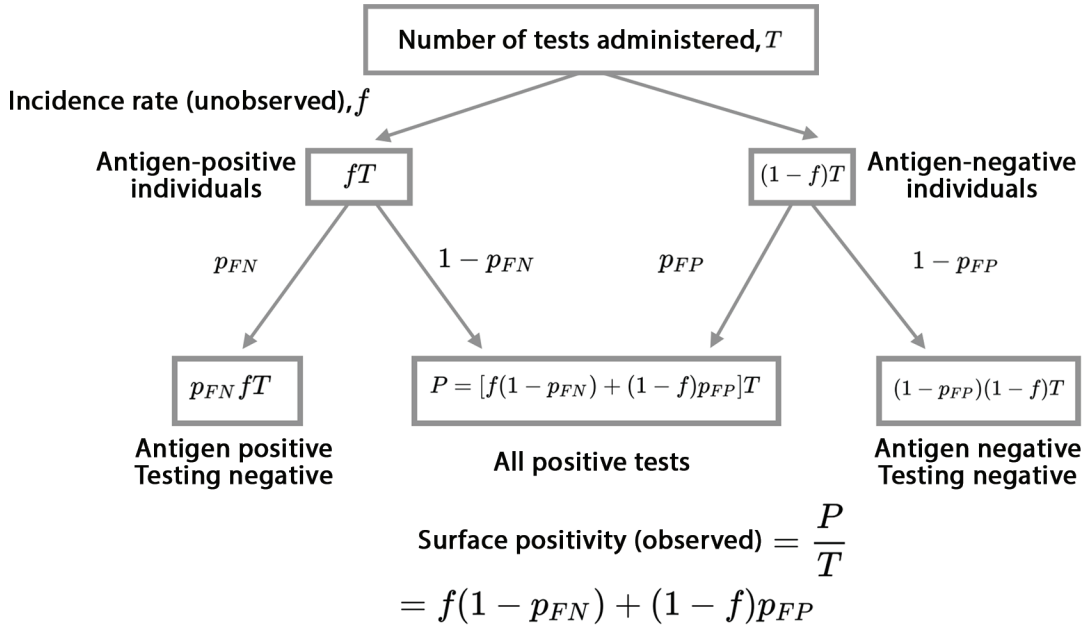
The goal is to begin with the surface positivity P/T and obtain an estimate $\hat{f}$ for the fraction of coronavirus infections. We characterize $\hat{f}$ by *accuracy*—how close is $\hat{f}$ to the true but

[13] For example, Zuckerman, *Scientists and War*; and *Beyond the Ivory Tower*.

[14] For example, Kucirka et al., "Variation in False-Negative Rate"; and Watson, Whiting, and Brush, "Interpreting a Covid-19 Test Result."

[15] For example, He et al., "Diagnostic Performance," Table 2.

unknown value *f*?—and *precision*—what is a plausible range of estimates for the fraction of coronavirus infections?



The surface positivity (fraction of positive tests) is observed, and from that we want to infer the unobserved incidence rate *f*. See text for further details.

Figure 2. Visual Representation of the Operational Situation

The Analytical Approach

To have confidence in methods that allow one to infer incidence rate from surface positivity and test errors, we begin by modeling the process of testing in which one knows the true state of the world, rather than the data from testing in which such knowledge is lacking.[16] We do so using simulation and analytical methods in which we specify the true state of the environment, which is characterized by the (unknown) fraction f_t of coronavirus infections in the population and the values of test errors p_{FN_t} and p_{FP_t} (where the subscript t denotes the true value). These are sufficient to simulate the number of positive tests P in a total of T tests.

Our approach is to first specify the true state of nature and use a simulation method to create the kind of data obtained in testing operations. We then develop methods to analyze those test data, when we know the true situation, and compare the predictions based on our methods with the true incidence rate. We do this because it is only when we know that the results of analyses are accurate that one can have confidence that the methods can be used to analyze test data for which we do not know the true state of nature.

[16] Shelton and Mangel, "Variability in Fish Populations."

We use three analytical tools: the binomial distribution, the binomial likelihood, and the Gaussian or normal approximation to the binomial distribution.[17] Although these are "elementary" tools of probability theory, such elementary tools when applied in mature ways can lead to novel and important insights.

We review these tools in the appendix, in which we delineate the distinction between the binomial probability and the binomial likelihood. This is an important conceptual distinction for our work.

The Risk Associated with Groups of Different Sizes and with Different Levels of Risk

To begin, assume that one has already constructed an estimate $\hat{f}$ for the incidence rate. Thus, the probability that a coronavirus-infected individual is joining a group is $\hat{f}$ and the probability that a noninfected individual is joining is $1 - \hat{f}$ (with obvious extension to the lower and upper estimates of the incidence rate).

We define the estimated risk $\mathcal{R}(g, \hat{f})$ associated with a group of size g when the estimated incidence rate is $\hat{f}$ to be the probability that at least one individual in the group is infected with coronavirus. Since the probability that none of the individuals is infected is $(1 - \hat{f})^g$, the probability that at least one of them is infected is

$$\mathcal{R}(g, \hat{f}) = 1 - (1 - \hat{f})^g, \tag{7}$$

which is Eq. 4.

The inverse question is perhaps more relevant to health policy: given a specified level of tolerable risk, $\mathcal{R}_{acc}$, what is the maximum group size concordant with this level of risk? To find this group size, we set $\mathcal{R}_{acc} = 1 - (1 - \hat{f})^g$ and solve for the group size, obtaining

$$g(\hat{f}, \mathcal{R}_{acc}) = \frac{log(1 - \mathcal{R}_{acc})}{log(1 - \hat{f})}, \tag{8}$$

where *log* denotes the logarithm (natural, base 2, or base 10 are all fine as long as they are used consistently). Choosing $\mathcal{R}_{acc}$ is the key policy decision, and once it is chosen the maximum group size for a given estimate of incidence rate emerges from Eq. 8.

Simulating and Analyzing Test Data with Known Test Errors

For computations, the base-case parameters are the fraction $f_t = 0.03$ of coronavirus infections, and the probability of a false negative test is $p_{FN_t} = 0.3$ and of a false positive test, $p_{FP_t} = 0.05$. In the the supplementary figures,[18] we explore other parameter choices.

[17] Feller, *Probability Theory*; and Mangel, *Theoretical Biologist's Toolbox*.

[18] Available at https://www.jhuapl.edu/Content/figures/Mangel_Brown_SuppFigs.pdf.

Simulating the Test Data

We simulate the testing process N times for which the number of tests T is fixed. Each simulation generates a number of positive tests, denoted by P_1, P_2, ... P_N where P_n is the number of positive results from T tests for the nth simulated test. Because the probability of a positive test is $p_+(f_t)$ (Eq. 1), P_n is binomially distributed with parameters T and $p_+(f_t)$, which we write as

$$P_n = \mathcal{B}(\cdot, T, p_+(f_t)) . \tag{9}$$

The Likelihood of an Incidence Rate Given the Number of Positive Tests

Given the number of positive tests P_n on the nth simulation, the likelihood of the data for any value of f is

$$\mathcal{L}(p_+(f)|P_n, T) \propto \mathcal{B}(P_n, T, p_+(f)).$$

The maximum likelihood estimate (MLE) for $p_+(f)$[19] is

$$\hat{p}_+(f) = \frac{P_n}{T} .$$

Letting $\hat{f}_n$ denote the MLE for incidence rate in the nth simulation, we write $\hat{p}_+(\hat{f}_n) = \hat{f}_n(1 - p_{FN}) + (1 - \hat{f}_n)p_{FP}$ so that

$$\hat{f}_n(1 - p_{FN}) + (1 - \hat{f}_n)p_{FP} = \frac{P_n}{T} . \tag{10}$$

Solving Eq. 10 for $\hat{f}_n$ gives

$$\hat{f}_n = \frac{P_n/T - p_{FP}}{1 - p_{FN} - p_{FP}} , \tag{11}$$

which is Eq. 2; we understand that $\hat{f}_n = 0$ if $P_n/\mathrm{T} < p_{FP}$. Thus, the estimate of incidence rate in our companion report[20] is the MLE. We consider a method to be accurate if, on average, the MLE is close to the true value that generated the data.

We normalized the likelihood to obtain a probability density $\phi_n(f \,|\, P_n, T)$ for f given the test data from the nth simulation:[21]

$$\phi_n(f|P_n, T) = \frac{\mathcal{L}(f|P_n, T)}{\sum_{f'=0}^{1} \mathcal{L}(f'|P_n, T)} , \tag{12}$$

[19] Mangel, *Theoretical Biologist's Toolbox.*

[20] Brown and Mangel, *Recommendations for Practice.*

[21] The incidence rate is a continuous variable, ranging between 0 and 1. To avoid the more complicated mathematics associated with integrating the likelihood, we work from the outset with a set of discrete values for f with increments of 0.001.

The $\phi_n(f \mid P_n, T)$[22] allow us to compute a range of values of f that are compatible with the model and data. We follow McElreath[23] and call this the compatibility interval (CI), thus avoiding the undesired implications of words such as *confidence* or *credible*.[24] We used the method of McElreath[25] to compute 95% CIs that are symmetrical around the MLE. In brief, for each simulation we used the $\phi_n(f \mid P_n, T)$ to generate 100,000 samples of f and then the quantile tool in R to construct the symmetric 95% quantile. The test range is then the upper limit of the quantile (i.e., the 0.975 point of the sample) minus the lower limit (i.e., the 0.025 point of the sample) of the quantile. We consider the 95% CI to be a measure of the precision of a method.

A method might be accurate (on average $\hat{f}$ is close to f_t) but the 95% CI so wide that it is useless. Similarly, a method might be precise, with a very small 95% CI, but inaccurate in that $\hat{f}$ and f_t are very far apart.

Gaussian Approximation for the Test Range

In this section, we suppress the subscript n on the number of positive tests, which we now denote by $\tilde{P}$ to emphasize that it is a random variable. We use these results about the mean and variance of a random variable $\tilde{X}$ and constant c: $\mathcal{E}(c) = c$, $\mathcal{E}(c\tilde{X}) = c\mathcal{E}(\tilde{X})$, $Var(c) = 0$, and $Var(c\tilde{X}) = c^2 Var(\tilde{X})$.

We multiply the numerator and denominator in Eq. 11 by the number of tests T to obtain

$$\hat{f} = \frac{\tilde{P} - Tp_{FP}}{T(1 - p_{FN} - p_{FP})}. \tag{13}$$

We take the expectation of both sides of Eq. 13 and then use Eq. 1 to obtain

$$\begin{aligned}
\mathcal{E}(\hat{f}) &= \frac{\mathcal{E}(\tilde{P}) - Tp_{FP}}{T(1 - p_{FN} - p_{FP})} \\
&= \frac{f_t T(1 - p_{FN}) + (1 - f_t)Tp_{FP} - Tp_{FP}}{T(1 - p_{FN} - p_{FP})} \\
&= \frac{f_t T(1 - p_{FN} - p_{FP})}{T(1 - p_{FN} - p_{FP})} \\
&= f_t \qquad .
\end{aligned} \tag{14}$$

[22] Readers familiar with Bayesian analysis will recognize Eq. 12 as a posterior given a uniform prior on f. Were we working with integrals rather than sums, this posterior would be a beta density.

[23] McElreath, *Statistical Rethinking*, 54.

[24] Morey et al., "Fallacy of Placing Confidence in Confidence Intervals."

[25] McElreath, *Statistical Rethinking*, 53ff.

This demonstrates that, on average, Eq. 2 is the true value of the (unobserved) incidence rate.

It is immediately clear that Eqs. 2 and 11 are not always sensible: whenever $P_n/T \leq p_{FP}$, the right-hand side of these equations is ≤ 0. The appropriate interpretation is that the MLE for incidence rate is 0, with implication that a different set of tools is needed. We discuss this in the section The Coronavirus End Game: The Case of No Positive Tests on page 27.

Using Eq. 13 and the properties of variance given above,

$$Var(\hat{f}) = \frac{Var(\tilde{P})}{T^2(1-p_{FN}-p_{FP})^2}. \tag{15}$$

The number of positive tests $\tilde{P}$ is a binomial random variable with parameters T and $p_+(f_t)$, so that $Var(\tilde{P}) = Tp_+(f_t)(1 - p_+(f_t))$. Thus, Eq. 15 becomes

$$\begin{aligned} Var(\hat{f}) &= \frac{Tp_+(f_t)(1-p_+(f_t))}{T^2(1-p_{FN}-p_{FP})^2} \\ &= \frac{p_+(f_t)(1-p_+(f_t))}{T(1-p_{FN}-p_{FP})^2} \ . \end{aligned} \tag{16}$$

In the Gaussian approximation for the binomial distribution, the distribution of $\hat{f}$ is normal with mean f_t and variance given by Eq. 16. For a Gaussian distribution, 95% of the probability is contained in a region centered on the mean and of length about 3.92 times the standard deviation so that the test range in the Gaussian approximation to the binomial distribution is

$$Range(f) = 3.92\sqrt{\frac{p_+(f_t)(1-p_+(f_t))}{T(1-p_{FN}-p_{FP})^2}} \ . \tag{17}$$

To obtain Eq. 3, we replace f_t by $\hat{f}$. We discuss this replacement in detail in the section on how the test range depends on the number of tests.

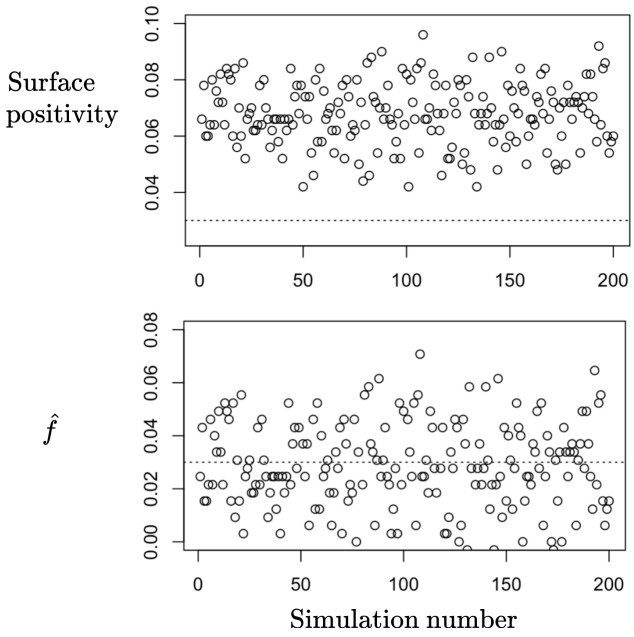
Results with Known Test Errors

We begin by illustrating results for 5,000 simulations generating data with $T = 500$ tests, and we subsequently explore the consequences of varying test numbers.

Surface Positivity and Estimates of Incidence Rate

In Figure 3, we show the first 200 replicates of testing $T = 500$ individuals for the base-case parameters $f_t = 0.03$, $p_{FN_t} = 0.3$, and $p_{FP_t} = 0.05$. In both panels, the horizontal line is f_t (i.e.,

what we expect if there were no test errors). In the upper panel we show the surface positivity, which is strongly positively biased in this case (the bias depends on the true incidence rate and true test errors). In the lower panel we show the individual MLEs for incidence rate (Eq. 13). Although there is noticeable variation, it is symmetric around f_t.



The results of testing $T = 500$ individuals for the base-case parameters $f_t = 0.03$, $p_{FN_t} = 0.3$, and $p_{FP_t} = 0.05$. In both panels, the horizontal line is f_t (i.e., what we expect if there were no test errors). In the upper panel we show the surface positivity, which is strongly positively biased (see text for an explanation). In the lower panel we show the individual MLEs for incidence rate (Eq. 13). Although there is noticeable variation, it is symmetric around f_t.

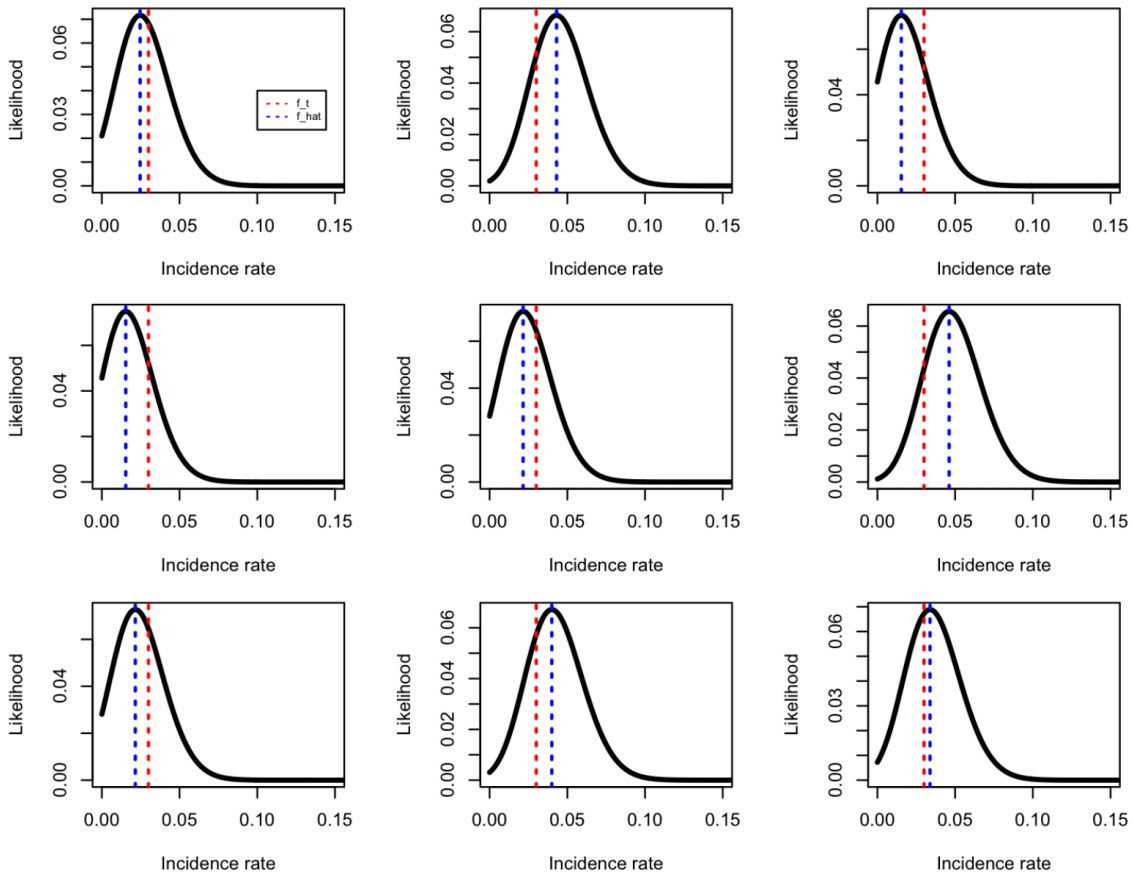
Figure 3. Results of the First 200 of 5,000 Replicates of 500 Tests for the Base-Case Parameters

The bias in surface positivity can be understood in terms of the test errors and the expected number of coronavirus infections in the test sample. That is, for $f_t = 0.03$ and 500 tests, we expect about 15 infected individuals in the sample; about 30% of the individuals will test negative, so that infected individuals provide about 10 positive tests. Of the remaining 485 individuals, about 5% will test positive, providing on average about 25 positive tests. Thus, on average we expect surface positivity to be about $(10 + 25)/500 = 0.07$; note that the points in the upper panel are symmetrically distributed around 0.07.

It is clear from this figure that surface positivity is a greatly biased estimate of incidence rate. For the 5,000 simulations in this case, the mean difference between the surface positivity and f_t is about 0.04—more than 100% of f_t. On the other hand, the mean difference between $\hat{f}$ and f_t is 0.0002.

Likelihoods

Each $\hat{f}_n$ has its own likelihood. In Figure 4, we show the likelihoods $\mathcal{L}(f \mid P_n, T)$ for replicates $n = 1, \ldots 9$ for $T = 500$ tests. The blue line is $\hat{f}_n$, which must always fall at the peak of the likelihood; hence, it moves around according to the surface positivity. The red line is f_t, which is stationary because it is the base-case true incidence rate. Sometimes $\hat{f}_n$ is very close to f_t and other times not—this is why computing the rest of the range is important.



The blue line is $\hat{f}_n$, which must always fall at the peak of the likelihood; hence, it moves around according to the surface positivity. The red line is f_t and is stationary because it is the true incidence rate.

Figure 4. The First Nine Likelihoods from 5,000 Simulations with 500 Tests

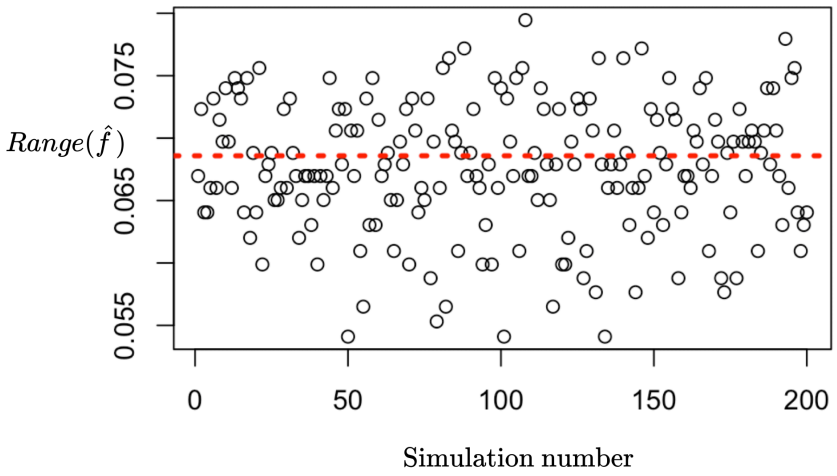
Test Ranges

There are three possible versions of the test range: (1) directly from the simulated data, so that each replicate produces a different test range based on the underlying but never observed incidence rate; (2) the Gaussian approximation for the test range using the true but never observed value of f_t as in Eq. 17, and (3) the Gaussian approximation to the test

range using the $\hat{f}_n$ as in Eq. 3 (this is the only method for which the unknown state of nature is not involved).

We first compare (Figure 5) the Gaussian approximation for the test range using the true but never observed value of incidence rate (shown as the red dotted line in Figure 5) and the test ranges obtained using the MLE for incidence rate in each replicate of the simulation (the points in Figure 5).

The simulated test ranges are symmetrically distributed around the test range using f_t. The mean difference between the $Range(\hat{f})$ and $Range(f_t)$ is 0.006. We next explore how these approximations depend on the number of tests.



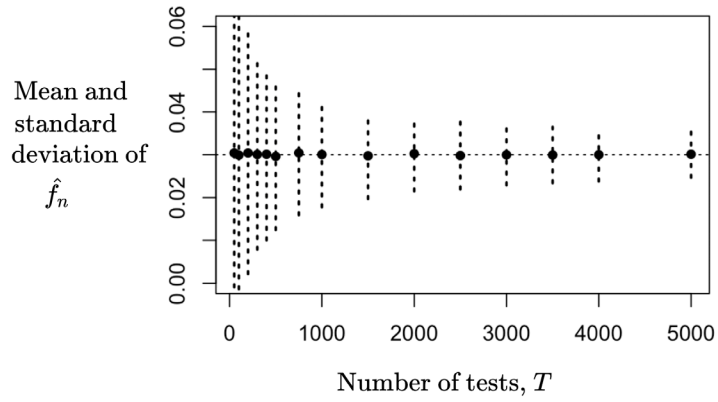
Each point is a replicate using Eq. 3 for the range (i.e., $\hat{f}_n$) The dotted red line shows the Gaussian approximation for the test range using Eq. 17 (i.e., f_t).

Figure 5. The First 200 of 5,000 Replicates of Test Range with the Base-Case Parameters

Accuracy and Precision Depending on the Number of Tests

Thus far we fixed $T = 500$; we now explore the consequences of varying the number of tests in the simulations. We report results for 50, 100, 200, 300, 400, 500, 750, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, and 5,000 tests. Including 50 and 100 tests allows us to explore how and why the methods fail.

In Figure 6, we show the mean and one standard deviation of the MLEs for incidence rate as a function of the number of tests. We conclude that even with 50 or 100 tests, the mean MLE is quite accurate—note that the points in Figure 6 sit on the horizontal dotted line almost from the outset. However, for small numbers of tests, the standard deviation of the MLEs is so large that we conclude that the accurate estimate is woefully imprecise.



Mean (points) and standard deviation (f_t, dotted line) of the MLEs for incidence rate.

Figure 6. Mean and +/– One Standard Deviation of the MLEs $\hat{f}_n$ for the Incidence Rate as They Depend on the Number of Tests

It is also clear from Figure 6 that although the standard deviation of the $\hat{f}_n$ decreases as the number of tests increases, the rate at which it decreases declines as the number of tests increases. We emphasize this point by showing just the standard deviation of $\hat{f}_n$ as a function of the number of tests in Figure 7. This figure shows that the gain in precision as one goes from 2,000 to 5,000 tests is minimal and that after about 1,500 tests, there is little precision to gain from additional testing.

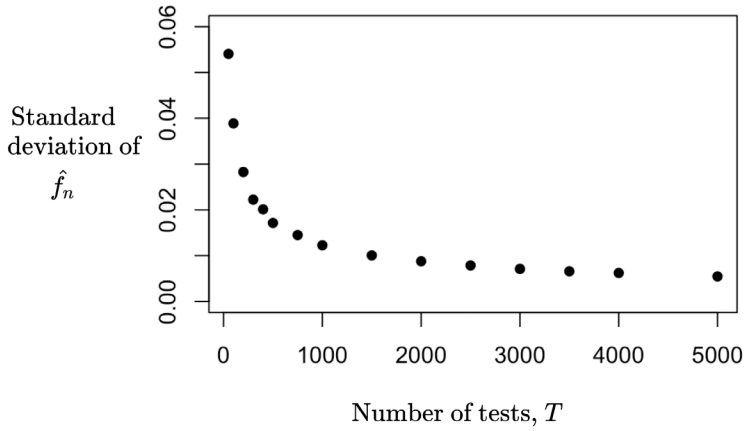
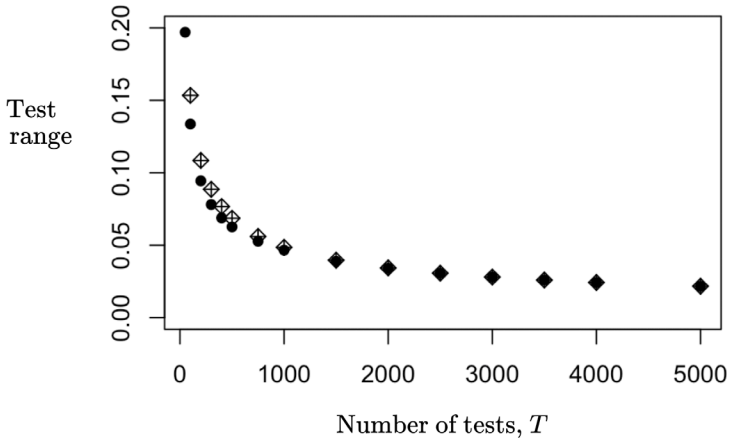


Figure 7. The Standard Deviation of the $\hat{f}_n$ Depending on the Number of Tests

How the Test Range Depends on the Number of Tests

We now more fully explore the test range, the Gaussian approximation to it, and how the test range depends on the number of tests.

To begin, in Figure 8, we compare the test ranges from the simulation (filled points, based on the unknown incidence rate f_t) with the test range from the Gaussian approximation (hatched diamonds). We see that after about 1,500 tests, the diamonds and points coincide. That is, the Gaussian approximation for test range using f_t is excellent with more than about 1,500 tests.



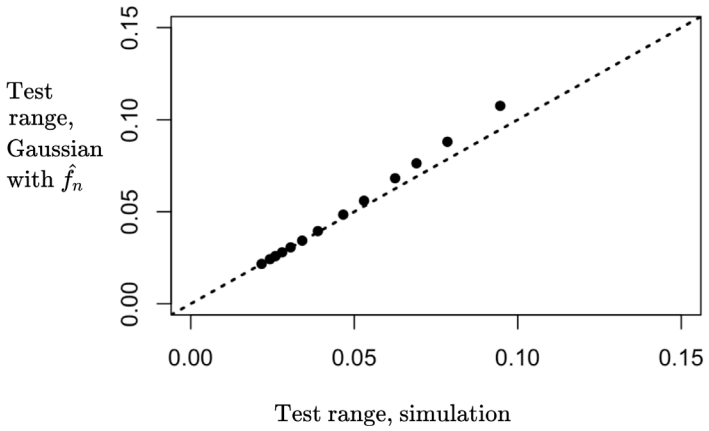
We show the mean test range from the simulation as solid points and test range from the Gaussian approximation using f_t to the binomial as hatched diamonds.

Figure 8. Test Range as It Depends on the Number of Tests

Figures 7 and 8 have a profound implication for the operational allocation of tests across space. Clearly, one wants to do many tests to identify who is and is not carrying antibodies so that contact tracing is possible. However, from the perspective of estimation and cost, it appears unnecessary to do more tests than the figures imply for estimating incidence. In particular, for the base-case parameters, doing more than 1,500 to 2,000 tests for estimating the incidence rate will have very little marginal effect on precision of the test results.

The remaining question about test range is whether the error introduced by replacing f_t by the estimate $\hat{f}$ is acceptable. To address this question, in Figure 9, we show mean test range from the simulation and mean test range from the Gaussian approximation (Eq. 3) using the MLEs rather than f_t and the 1:1 line. The far right-hand point corresponds to 50 tests and the far left-hand point to 5,000 tests. The points start falling onto the line after 1,000 tests, showing that for more than about 1,000 tests the Gaussian approximation to test range using the MLEs is highly accurate. The particular value of 1,000 tests is conditioned on the parameters; in the supplementary figures,[26] we show that the qualitative patterns in Figures 6–9 do not change as parameters change but the quantitative details may.

[26] Available at https://www.jhuapl.edu/Content/figures/Mangel_Brown_SuppFigs.pdf.



The mean test range from the simulation is shown on the *x* axis. The mean test range from the Gaussian approximation (Eq. 3) is shown on the *y* axis. The number of tests decreases as one goes from left to right on the *x* axis.

Figure 9. Mean Test Range from the Simulation and the Gaussian Approximation Using the MLEs Rather Than f_t and the 1:1 Line

The Consequences of Uncertainty Test Errors

To explore the consequences when p_{FN_t} and p_{FP_t} that generate the observed number of positive tests differ from the test errors that are assumed when constructing $\hat{f}$, we let $\hat{f}_n(p_{FN}, p_{FP} | P_n, T, p_{FN_t}, p_{FP_t})$ denote the estimate of incidence rate on the nth simulation when the assumed test errors are p_{FN} and p_{FP} but the true test errors are p_{FN_t} and p_{FP_t}.

We evaluate the consequence of not knowing the test error probabilities by the relative error (*RE*) of the average of the MLEs and true value of the incidence rate used to generate the data f_t. Suppressing the dependence of the *RE* on the P_n, it is

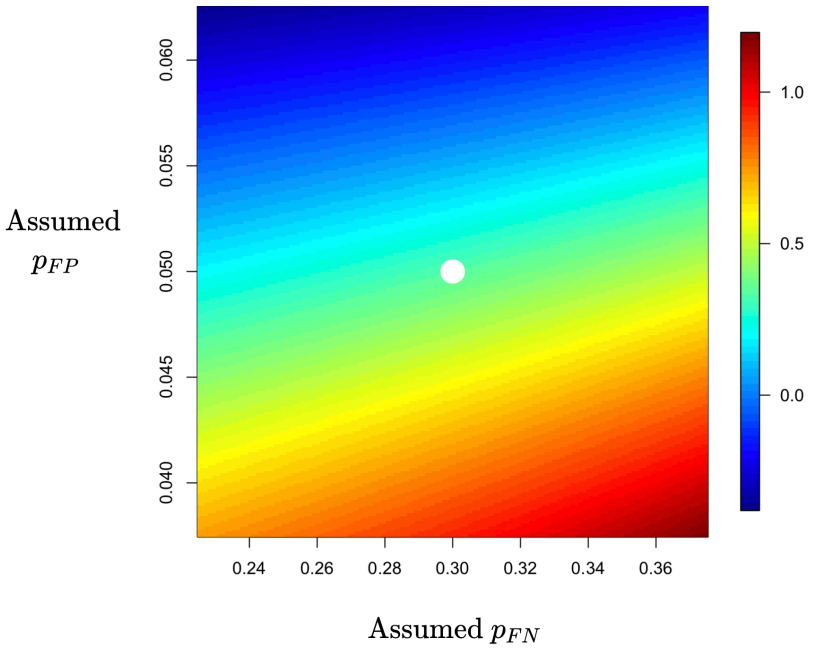
$$RE(\hat{f}(p_{FN}, p_{FP}|p_{FN_t}, p_{FP_t})) = \frac{\frac{1}{N}\sum_{n=1}^{N} \hat{f}_n(p_{FN}, p_{FP}|P_n, T, p_{FN_t}, p_{FP_t}) - f_t}{f_t}. \quad (18)$$

Simulation Results

In Figure 10, we show the *RE* for $T = 3{,}000$ tests; the qualitative pattern is insensitive to the number of tests, although the numerical scale of the axis varies with the number of tests.

Note these properties:

(1) $RE(\hat{f}(p_{FN}, p_{FP} | p_{FN_t}, p_{FP_t}))$ has a minimum when the assumed test errors are the true values of the test errors. However, there is a diagonal swath in the test error plane in which this *RE* is small.



True test errors are denoted by the white circle.

Figure 10. The Relative Error, Eq. 17, of Assuming Test Errors That Differ from the True Test Errors

(2) As the assumed value of p_{FN} increases, the *RE* increases, as seen by following a horizontal line from the y axis. The reason for this is that given a number of positive tests, as p_{FN} increases but p_{FP} is held constant, a higher value of incidence rate is required to generate the observations. A simple way of seeing this is to set $p_{FP} = 0$ in Eq. 11, so that $\hat{f}_n = \frac{P_n/T}{1-p_{FN}}$. It is clear that increasing p_{FN} increases the estimate of the incidence rate.

(3) Similarly, as p_{FP} increases, *RE* declines and becomes negative, as seen by following a vertical line from the x axis. The reason for this is that given a number of positive tests, as the probability of false positive test p_{FP} increases but the p_{FN} is held constant, a lower value of the true incidence rate is required to generate the data. Setting $p_{FN} = 0$ in Eq. 11 does not allow a simplification that has obvious implication because p_{FP} appears in both the numerator and denominator of the estimate for incidence rate. However, in the next section we analytically demonstrate that increasing p_{FP} decreases the estimate of incidence rate.

(4) The diagonal swath of small values of *RE* in Figure 10 is a result of the counterbalancing factors in points 2 and 3.

The Certainty Equivalent (CE) Approximation and the Analytical Verification of the Patterns of Error

In the CE approximation, random variables are replaced by their means.[27] When the true values of the incidence rate and test errors are f_t, p_{FN_t}, and p_{FN_t}, the mean of P_n/T is

$$p_{CE}(f_t, p_{FN_t}, p_{FP_t}) = f_t(1 - p_{FN_t}) + (1 - f_t)p_{FP_t}. \tag{19}$$

Substituting into Eq. 11, we obtain a fully deterministic function of the assumed test errors, conditioned on the true test errors. We replace the subscript n by the subscript CE and write

$$\hat{f}_{CE} = \frac{p_{CE}(f_t, p_{FN_t}, p_{FP_t}) - p_{FP}}{1 - p_{FN} - p_{FP}}. \tag{20}$$

Patterns with Test Errors Using the CE Approximation

In Figure 11 we show a sensitivity analysis of the RE as the true test errors vary using the CE approximation.

We use Eq. 20 to explore the patterns in Figures 10 and 11 and demonstrate the following:[28]

- The rate of change of $\hat{f}_{CE}$ with p_{FN} is positive. That is, for the same value of fraction of positive tests, increasing p_{FP} leads to an overestimation of the fraction of coronavirus infections.
- The rate of change of $\hat{f}_{CE}$ with p_{FP} is negative. That is, for the same value of fraction of positive tests, increasing p_{FP} leads to an underestimation of the fraction of coronavirus infections.
- The contours of $\hat{f}_{CE}$ are straight lines. That is, $\hat{f}_{CE} = c$ implies

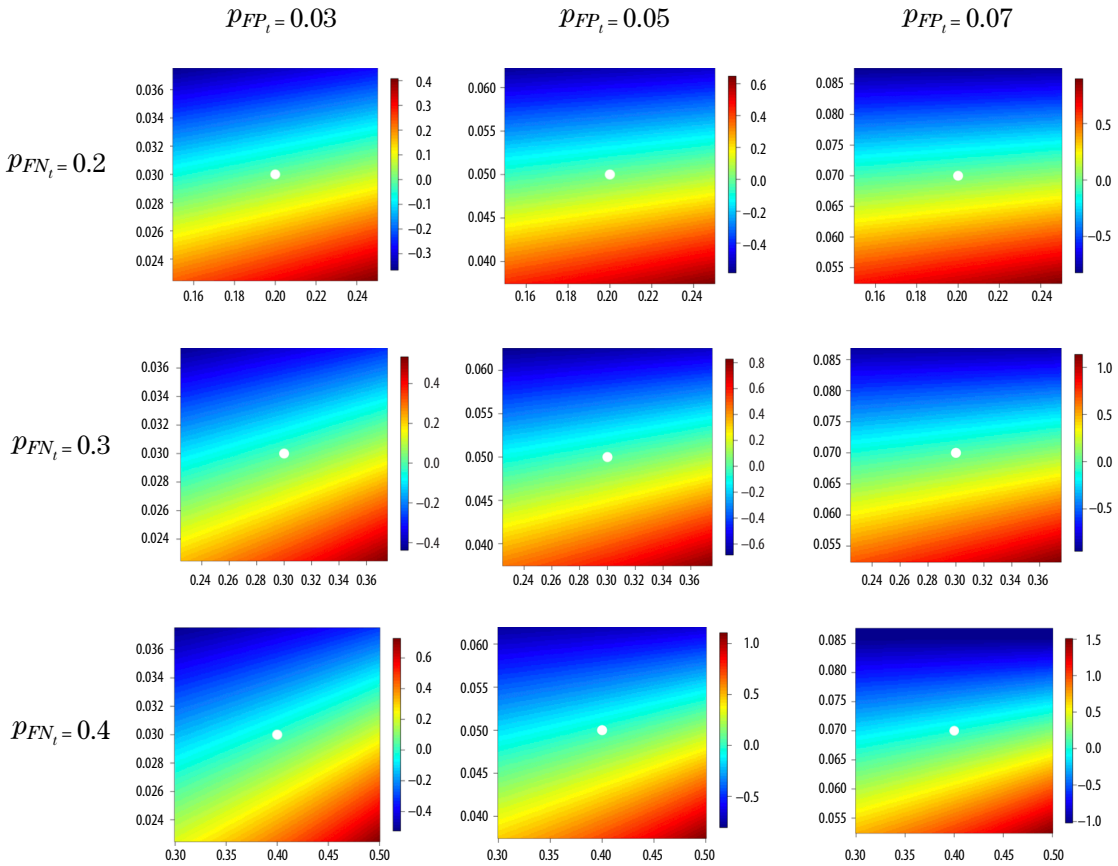
$$p_{FP} = \frac{c}{1 - c}p_{FN} + \left[\frac{p_{CE}(f_t, p_{FN_t}, p_{FP_t}) - c}{1 - c}\right]. \tag{21}$$

The rate of change of $\hat{f}_{CE}$ with p_{FN} is positive. We assume (perfectly reasonably because otherwise there is no point in conducting testing) that $p_{FN} + p_{FP} < 1$. To simplify the symbology, we set $x = p_{FN}$, $y = p_{FP}$, suppress the dependence of the surface positivity on the true test errors, and rewrite Eq. 20 as

$$\hat{f}_{CE}(x, y) = \frac{p_{CE} - y}{1 - x - y} = [p_{CE} - y][1 - x - y]^{-1}. \tag{22}$$

[27] Mangel, *Decision and Control in Uncertain Resource Systems*.

[28] The analysis that follows is identical if one works with the stochastic function given by Eq. 11 but considers that the results are conditioned on the value of P_n/T.



We investigate three values of the true probability of a false negative test (rows) and three values of the true probability of a false positive test (columns). In each heat plot, the x axis is p_{FN}, and the y axis is p_{FP}. We let these range between 50% and 150% of the true values, so that the numerical values on the axes differ across plots. The entry in each heat plot is the corresponding RE in the estimate of fraction of coronavirus infections in the population, which varies. In the first row, the hottest values are approximately 0.4, 0.6, and 0.7 (left, middle, and right columns); in the second row they are approximately 0.4, 0.8, and 1.0 (left, middle, and right columns); in the third row they are approximately 0.6, 1.0, and 1.5 (left, middle, and right columns).

Figure 11. Sensitivity Analysis of the Relative Error for a Wider Range of True Test Errors Using the CE Approximation

The rate of change of $\hat{f}_{CE}$ with respect to p_{FN} is the same as the rate of change of $f_{CE}(x, y)$ with respect to x. Since only the second term on the right-hand side of Eq. 22 depends on x,

$$\begin{aligned}\frac{\partial \hat{f}_{CE}(x,y)}{\partial x} &= [p_{CE} - y](-1)[1 - x - y]^{-2}(-1) \\ &= [p_{CE} - y][1 - x - y]^{-2} \\ &= \left[\frac{\hat{f}_{CE}(x,y)}{1 - x - y}\right], \text{ which is always } > 0\,. \end{aligned} \tag{23}$$

The intuition underlying this result is the following. The surface positivity is given. False negative tests mean that infected individuals are not counted as such, so that to obtain the data one is forced to estimate an incidence rate that is larger than the true incidence rate.

The rate of change of $\hat{f}_{CE}$ with p_{FP} is negative. The rate of change of $\hat{f}_{CE}$ with respect to p_{FP} is the same as the rate of change of $\hat{f}_{CE}(x, y)$ with respect to y. Since both terms on the right-hand side of Eq. 22 depend on y,

$$\begin{aligned}\frac{\partial f_{CE}(x,y)}{\partial y} &= -[1-x-y]^{-1} + [p_{CE}-y][1-x-y]^{-2} \\ &= \frac{1}{1-x-y}[-1+\hat{f}_{CE}(x,y)], \text{ which is always } < 0 \text{ since } \hat{f}_{CE} < 1. \quad (24)\end{aligned}$$

The intuition behind this result mirrors that in the previous section. False positive tests mean that uninfected individuals are counted in the infected pool. Thus, to obtain the data one estimates an incidence rate that is smaller than the true incidence rate.

The contours of $\hat{f}_{CE}$ are straight lines. When $\hat{f}_{CE}(x, y)$ is constant, we set $\hat{f}_{CE}(x, y) = c$, where $c < 1$ and solve the resulting equation,

$$\frac{p_{CE}-y}{1-x-y} = c\,,$$

for y in terms of x. Straightforward algebra gives

$$y = \frac{c}{1-c}x + \frac{p_{CE}-c}{1-c}\,.$$

Thus, the estimate for the fraction of coronavirus infections will be constant on lines with slope $c/(1 - c)$, which is always positive, since $c < 1$ because it is the estimate of the fraction of infected individuals.

Important Analytical Gains by Assuming That $p_{FP} = p_{FP_t}$

Using Eq. 19 in Eq. 20 allows us to write

$$\hat{f}_{CE} = \frac{f_t(1-p_{FN_t}) + (1-f_t)p_{FP_t} - p_{FP}}{1-p_{FN}-p_{FP_t}}\,. \quad (25)$$

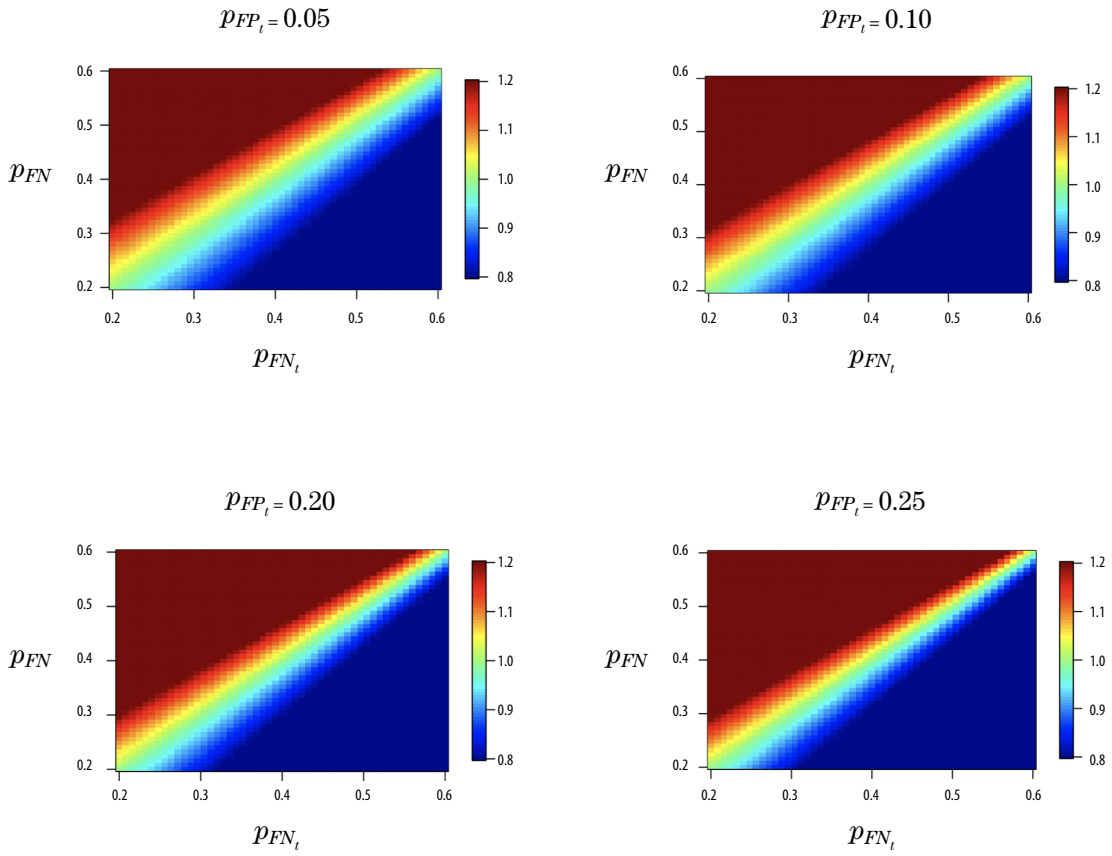
If we set $p_{FP} = p_{FP_t}$, this equation becomes

$$\hat{f}_{CE} = \frac{f_t(1-p_{FN_t}-p_{FP_t})}{1-p_{FN}-p_{FP_t}}\,. \quad (26)$$

Dividing both sides of Eq. 26 by f_t gives a measure of the error in incidence rate that is independent of the true incidence rate and depends only on the true test errors and the assumed value of p_{FP}:

$$\frac{f_{CE}}{f_t} = \frac{1 - p_{FN_t} - p_{FP_t}}{1 - p_{FN} - p_{FP_t}} . \tag{27}$$

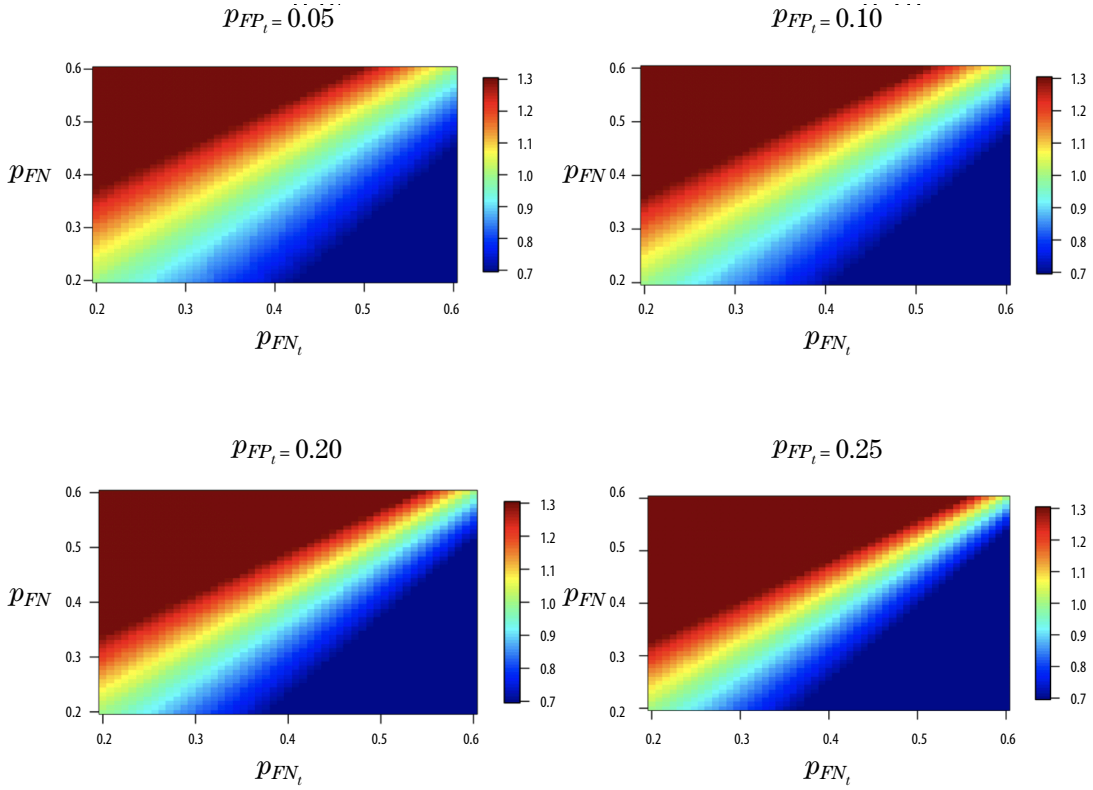
In Figure 12, we show regions in the p_{FN_t} (x axis)-p_{FN} (y axis) plane in which $\hat{f}_{CE}/f_t$ varies by no more than 20%. For ease of presentation, values of the ratio outside this region are colored brown for values greater than 1.2 and blue for values less than 0.8. In Figure 13, we show analogous results in which $\hat{f}_{CE}/f_t$ varies by no more than 30%.



Regions in the p_{FN_t} (x axis)-p_{FN} (y axis) plane in which $0.8 \leq \hat{f}_{CE}/f_t \leq 1.2$. Values of the ratio outside this region are colored brown for values greater than 1.2 and blue for values less than 0.8.

Figure 12. Regions That Vary By No More Than 20%

Figures 12 and 13 provide an operational context for how much one can be wrong in setting p_{FN} and still have estimates of the incidence rate that are not too bad.



Regions in the p_{FN_t} (x axis)-p_{FN} (y axis) plane in which $0.7 \leq \hat{f}_{CE}/f_t \leq 1.3$. Values of the ratio outside this region are colored brown for values greater than 1.3 and blue for values less than 0.7.

Figure 13. Regions That Vary By No More Than 30%

Accounting for Heterogeneity in Test Errors

To account for heterogeneity in test errors, we assume that the means, variances, and covariance of test errors are known and denote these by $\overline{p}_{FN}$ and $\overline{p}_{FP}$ (mean test errors), Vp_{FN}, Vp_{FP} (variance in test errors), and $Cov(p_{FN}, p_{FP})$ (covariance between test errors), respectively. For ease of notation, we let

$$\overline{f}_{CE}(\overline{p}_{FN}, \overline{p}_{FP}) = \frac{p_{CE} - \overline{p}_{FP}}{1 - \overline{p}_{FN} - \overline{p}_{FP}}.$$

We now demonstrate that

$$\mathcal{E}(\hat{f}_{CE}) = \overline{f}_{CE}(\overline{p}_{FN}, \overline{p}_{FP}) + \frac{\overline{f}_{CE}(\overline{p}_{FN}, \overline{p}_{FP})}{(1 - \overline{p}_{FN} - \overline{p}_{FP})^2} V_{p_{FN}} + \frac{\overline{f}_{CE}(\overline{p}_{FN}, \overline{p}_{FP}) - 1}{(1 - \overline{p}_{FN} - \overline{p}_{FP})^2} V_{p_{FP}} + \frac{2\overline{f}_{CE}(\overline{p}_{FN}, \overline{p}_{FP}) - 1}{(1 - \overline{p}_{FN} - \overline{p}_{FP})^2} Cov(p_{FN}, p_{FP}) \quad , \qquad (28)$$

where $\mathcal{E}(\cdot)$ denotes the expectation over the distribution of the test errors.

We suppress the subscript on p_{CE} and now write $f(x,y) = \frac{p-y}{1-x-y}$ but now assume that x and y have distributions with means $\overline{x}$ and $\overline{y}$,variances V_x and V_y, and covariance $Cov(x,y)$. These characterize the heterogeneity in the test errors.

Because $f(x,y)$ is a nonlinear function of x and y, we cannot simply insert the mean values. The delta-method[29] allows us to obtain a good approximation for the expectation, we which denote as $\mathcal{E}\{f(x,y)\}$.

We Taylor expand $f(x,y)$ to two terms around the mean values, using subscripts to denote partial derivatives (i.e., $f_x = \frac{\partial f}{\partial x}$ evaluated at the mean values), so that

$$\begin{aligned}\mathcal{E}\{f(x,y)\} = \mathcal{E}\{&f(\overline{x},\overline{y}) + f_x(x-\overline{x}) + f_y(y-\overline{y}) \\ &+ \frac{1}{2}[f_{xx}(x-\overline{x})^2 + 2f_{xy}(x-\overline{x})(y-\overline{y})^2 + f_{yy}(y-\overline{y})^2]\}. \end{aligned} \tag{29}$$

The first term in the first line on the right-hand side of Eq. 29 is a constant, so comes out of the expectation, and the next two terms are identically zero because they are linear in x and y, which have expectations $\overline{x}$ and $\overline{y}$. In the second line on the right-hand side of Eq. 29, $\mathcal{E}\{(x-\overline{x})^2\}$, $\mathcal{E}\{(y-\overline{y})^2\}$, and $\mathcal{E}\{(x-\overline{x})(y-\overline{y})\}$ are the variances of x and y and the covariance between them, so that:

$$\mathcal{E}\{f(x,y)\} = f(\overline{x},\overline{y}) + \frac{1}{2}f_{xx}V_x + 2f_{xy}Cov(x,y) + \frac{1}{2}f_{yy}V_y\,. \tag{30}$$

We now evaluate the relevant partial derivatives. From Eq. 23,

$$f_x = \frac{p-y}{(1-x-y)^2} = \frac{f}{1-x-y}\,. \tag{31}$$

Using this equation, the second partial derivative is

$$f_{xx} = \frac{f_x}{1-x-y} + \frac{f}{(1-x-y)^2} = \frac{2f}{(1-x-y)^2}\,. \tag{32}$$

The partial derivative with respect to y is (Eq. 24)

$$f_y = \frac{f-1}{1-x-y}\,. \tag{33}$$

Using this equation, the second partial derivative is

$$f_{yy} = \frac{f_y}{(1-x-y)} + \frac{f-1}{(1-x-y)^2} = \frac{2(f-1)}{(1-x-y)^2}. \tag{34}$$

[29] Hilborn and Mangel, *Ecological Detective*.

The mixed partial derivative is found from either differentiating Eq. 31 or 33:

$$f_{xy} = \frac{f_y}{1-x-y} + \frac{f}{(1-x-y)^2} = \frac{f-1}{(1-x-y)^2} + \frac{f}{(1-x-y)^2} = \frac{2f-1}{(1-x-y)^2}\,. \quad (35)$$

Combining these in Eq. 29 we obtain

$$\mathcal{E}\{f(x,y)\} = f(\overline{x},\overline{y}) + \frac{f(\overline{x},\overline{y})}{(1-\overline{x}-\overline{y})^2}V_x + \frac{f(\overline{x},\overline{y})-1}{(1-\overline{x}-\overline{y})^2}V_y + \frac{2f(\overline{x},\overline{y})-1}{(1-\overline{x}-\overline{y})^2}Cov(x,y). \quad (36)$$

Replacing x and y in Eq. 36 by p_{FN} and p_{FP} gives Eq. 28, which is a beautiful equation (*sensu* Farmelo[30]).

To make use of Eq. 28, one requires the means, variances, and covariance of the test errors. To our knowledge, those data have not yet been published, but doing so is a future valuable endeavor.

The Coronavirus End Game: The Case of No Positive Tests

As the pandemic wanes, we expect that no positive results from T tests will become more common. It is clearly wrong to infer that surface positivity being 0 means that the incidence rate is 0. However, since $\tilde{P} = 0$ implies that the MLE for incidence rate is 0, a different approach is required.

Because testing involves the random processes generated by test errors, even when f is bounded away from 0 one may obtain $\tilde{P} = 0$. In Figure 14, we show four likelihoods chosen deliberately to show that the MLE can be both close to f_t or close to 0.

The probability that $\tilde{P} = 0$ when the incidence rate is f and T tests are given is

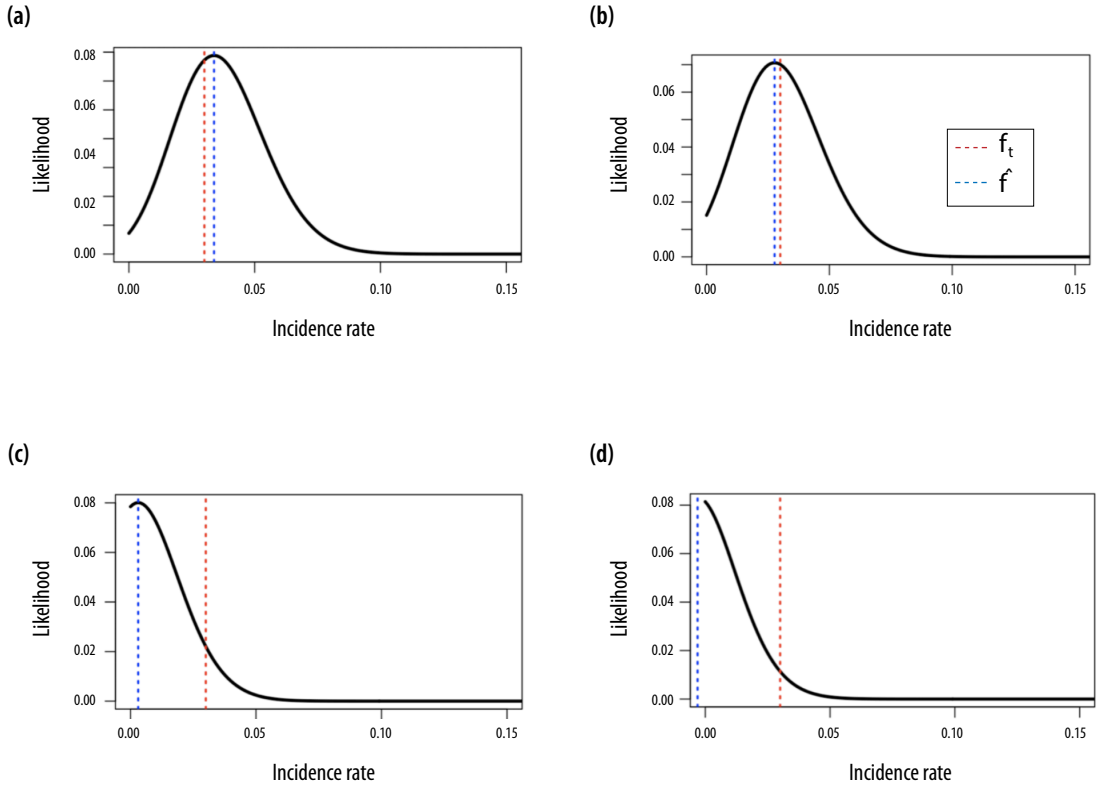
$$\Pr\{\tilde{P} = 0|T, f\} = (1 - p_+(f))^T\,. \quad (37)$$

This is also the likelihood of f given no positive results in T tests. Suppressing the dependence on $\tilde{P} = 0$, the log-likelihood is

$$\hat{\mathcal{L}}(f|T) = T \cdot log(1 - p_+(f)) = T \cdot log(1 - p_{FP} - f[1 - p_{FP} - p_{FN}])\,. \quad (38)$$

It is clear that the MLE for the incidence rate is 0, with maximum value of the likelihood $\hat{\mathcal{L}}_{max} = T \cdot log(1 - p_{FP})$.

[30] Farmelo, *It Must Be Beautiful.*



In all panels, $f_t = 0.03$ is the red dotted line and the MLE is the blue dotted line. Panels a and b show that the MLE can be close to f_t but on either side of it. Panels c and d show that the MLE can also be close to 0 and also on either side of it. (For purposes of illustration, in the lower right-hand panel we have left the MLE as a negative number, but as explained above, in all analyses we set such values to 0). For the vast majority of the simulations with the base-case parameters, the likelihoods are similar to panels a and b.

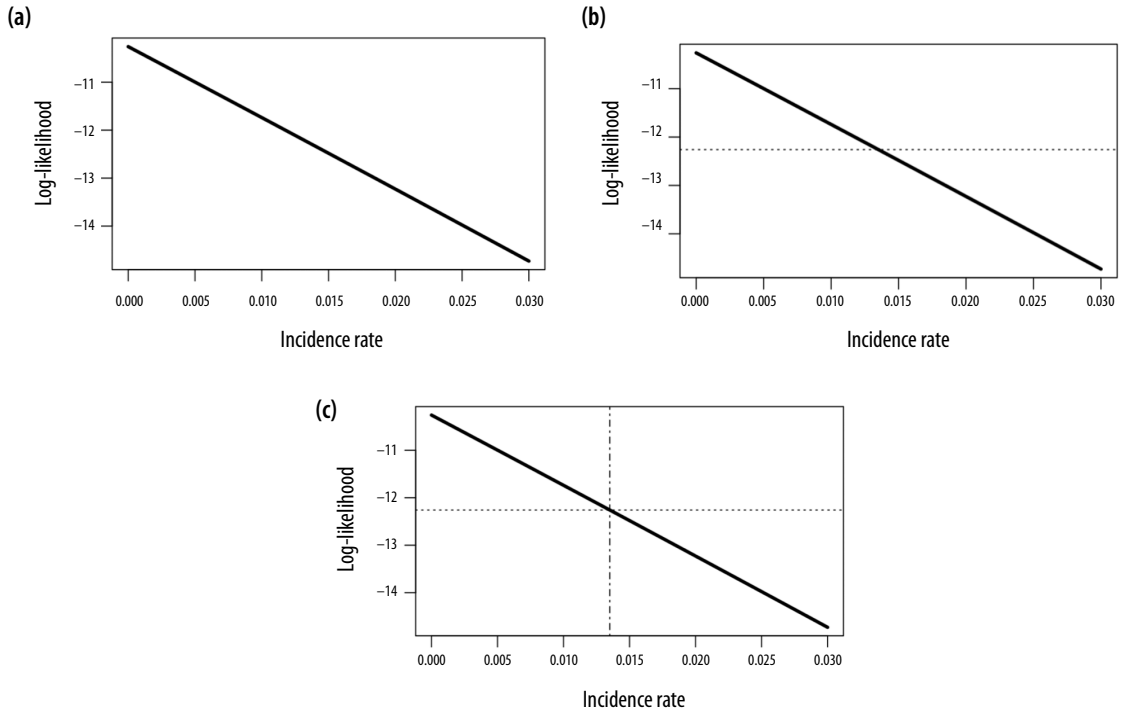
Figure 14. Four Deliberately Chosen Likelihoods for the Base-Case Parameters Showing How the MLE $\hat{f}$ Can Be Close to the True Incidence Rate f_t or Close to 0

It is also clear that the 95% CI will be bounded on the left at $f = 0$ and have a right-hand limit that depends on the number of tests and the test errors. We find it using the method of Hudson,[31] who shows that a good approximation to the 95.4% CI is the value f_H for which

$$\hat{\mathcal{L}}(f_H|T) = \hat{\mathcal{L}}_{max} - 2\,. \tag{39}$$

The solution of Eq. 39 can be obtained graphically (Figure 15), but with a few steps f_H can also be determined analytically.

[31] Hudson, "Interval Estimation."



In a graphical solution of Eq. 39, (a) one plots the log-likelihood, Eq. 38, as a function of incidence rate, (b) then draws a horizontal line at two units below the maximum value of the likelihood until it intersects the likelihood plot, and (c) draws a vertical line from that intersection to the *x* axis, which gives the value of f_H.

Figure 15. Graphical Illustration of Hudson's Method for Finding the 95% CI

Using Eq. 38, Eq. 39 becomes

$$T \cdot log(1 - p_{FP} - f_H[1 - p_{FP} - p_{FN}]) = T \cdot log(1 - p_{FP}) - 2\,, \tag{40}$$

from which it is straightforward to solve for f_H; we obtain

$$f_H = \frac{(1 - p_{FP})(1 - e^{-\frac{2}{T}})}{1 - p_{FP} - p_{FN}}. \tag{41}$$

Once f_H is obtained, we can apply Eq. 4 to estimate the risk of including an infected individual in groups of different sizes, given that no positive results were obtained in T tests (Figure 16).

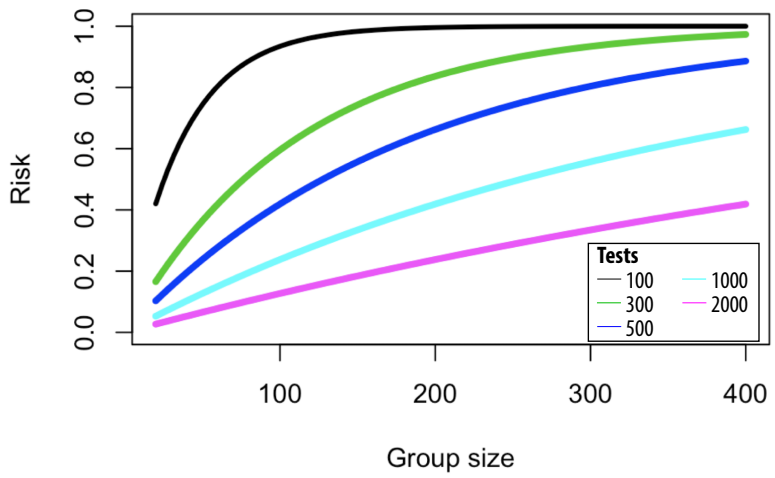


Figure 16. The Risk (Probability of Including At Least One Infected Individual) in Groups of Different Sizes, Given That No Positive Results were Obtained in T Tests for the Baseline Values of p_{FN} and p_{FP}

Discussion

We have provided supporting analysis for the recommendations for practice given in our companion report[32] and summarized in Eqs. 1–4 here. The key messages of our companion report and the additional recommendation for practice given in this report are as follows:

- Positivity rate is not incidence rate.
- Increasing the probability of a false negative test increases the estimate of incidence rate, and increasing the probability of a false positive test decreases the estimate of incidence rate.
- Risk (defined as the probability of including an infected individual in a group of a given size) is not either/or. That is, "safety is not binary"[33]—but is a function of both the incidence rate and the group size.

To assess the quality of the recommendations for practice in our companion report,[34] we modeled the process of generating test data in which the true state of the world (incidence rate f_t, probability of a false negative test p_{FN_t}, and probability of a false positive test p_{FP_t}) is known. This allows us to compare analytical predictions with the true state of the world.

[32] Brown and Mangel, *Recommendations for Practice*.

[33] Carroll, "Pandemic Decisions."

[34] Brown and Mangel, *Recommendations for Practice*.

When test errors are known, we showed that surface positivity is a very poor proxy for the underlying incidence rate (Figure 3) and that Eq. 2 is the MLE for the incidence rate (Eq. 11, illustrated in Figure 3). The test range given by Eq. 3 is a very good approximation to the true test range determined from the stochastic simulation (Figure 9). In particular, we demonstrated that with more than about 1,500 tests (for the base-case parameters; see the supplementary figures[35] for other parameter values) the Gaussian approximation to the 95% CI is virtually identical to the CI obtained from the stochastic simulation (Figure 8).

By sweeping over the number of tests, we showed that the mean of the MLE is an accurate estimate for even a modest number of tests, but that its variance is so great for a modest number of tests that the MLE is very imprecise (Figure 6). However, the standard deviation of the MLEs declines as the reciprocal of the square root of the number of tests (Figure 7) and the test range has the same characteristic (Figure 8). This result has important implications for distributing tests in space since it means that it is possible to oversample regions by allocating too many tests.

We then turned to the case in which the test errors are not known. To investigate this case, we generated data using f_t, p_{FN_t}, and p_{FP_t} but applied Eq. 3 assuming test errors that may differ from the true ones and computed the relative error between the mean of the MLEs and the true incidence rate (Eq. 18). Using simulation, we showed (Figures 10 and 11) that when the choice of p_{FN} used in computing $\hat{f}$ exceeds the true value, one overestimates the true incidence rate and that when the choice of p_{FP} used in computing $\hat{f}$ exceeds the true value, one underestimates the true incidence rate.

Using a certainty equivalent (CE) approximation, in which we replace stochastic surface positivity by its mean, we confirmed the observations from the simulation analytically (Eqs. 23 and 24). In addition, using the CE approximation, we showed how to include distributions of test errors in the construction of $\hat{f}$ (Eq. 29); we previously reported this in our companion report (appendix).[36]

We also explicitly considered the case in which surface positivity is 0, which is likely to happen during testing as the pandemic wanes (although, as shown in Figure 14, stochastic fluctuations may lead to $\tilde{P} = 0$ even when f_t is far from 0). In this case the MLE for incidence rate is 0, and although the 95% CI can be computed from the stochastic simulation, that method is not appropriate as a recommendation for practice. When $\tilde{P} = 0$, the likelihood function simplifies considerably (Eq. 38), and using a method for approximating the 95% CI directly from the likelihood, we derived a formula for the maximum incidence rate consistent with no positive tests (i.e., the right-hand limit of the 95% CI; Figure 15, Eq. 41). This can then be used in risk calculations relating group size and the probability of including an infected individual when $\tilde{P} = 0$ (Figure 16).

[35] Available at https://www.jhuapl.edu/Content/figures/Mangel_Brown_SuppFigs.pdf.

[36] Brown and Mangel, *Recommendations for Practice*.

There is a remaining question about how to develop recommendations for practice when surface positivity $P/T \leq p_{FP}$, in which case Eq. 2 can no longer be applied since the right-hand side is 0, which we interpret as the MLE being 0 (Figure 14d). As in the case of surface positivity equal to 0, the 95% CI can be computed from the stochastic simulation, but a simpler recommendation for practice still needs to be discovered, unless a test with no false positive errors is developed.

Supplementary Figures

In the main text, we used base-case parameters $f_t = 0.03$, $p_{FN} = 0.3$, and $p_{FP} = 0.05$. In this section, we reproduce Figures 6–9 for different parameters that bracket the base-case parameters of the text: $f_t = 0.01, 0.03, 0.09$; $p_{FN} = 0.2, 0.3, 0.4$; and $p_{FP} = 0.025, 0.05, 0.1$. Including the base-case parameters, there are 27 combinations of parameters. For ease of presentation, we show one four-panel figure (corresponding to Figures 6–9) for each of the parameter combinations.[37]

Examination of these figures shows that

- All of the qualitative patterns described for the base-case parameters are present for each of the alternative parameter sets.
- The methods we developed perform most poorly when $f_t = 0.01$. This is the result of the infection being rarer, and consequently harder to detect.
- Similarly, the method performs most poorly for the largest values of false negative and false positive tests because the information obtained via testing is most degraded.

However, the main conclusions remain intact.

[37] Available at https://www.jhuapl.edu/Content/figures/Mangel_Brown_SuppFigs.pdf.

Appendix The Binomial Probability Distribution and Likelihood and the Gaussian Approximation to the Binomial

The Binomial Distribution

The binomial distribution arises in coin flipping. Imagine that a coin, for which the probability of landing on heads is p, is flipped N times. The number of times the coin lands on heads is then 0, 1, 2, ... N; the binomial distribution gives the probability of obtaining k heads in the N flips.

We denote this probability by $\mathcal{B}(k, N, p)$.

The analogy with coronavirus testing is clear: instead of the number of flips, we have the number of tests T; the probability of a positive test given by Eq. 1; and instead of the number of times landing on heads, we have the number of positive tests P.

For the binomial distribution in coin flipping, we let K denote a random variable corresponding to the number of times the coin lands on heads. The probability that $K = k$ is

$$\Pr\{K = k|N, p\} = \binom{N}{k} p^k (1-p)^{N-k}, \tag{42}$$

where $\binom{N}{k} = \frac{N!}{k!(N-k)!}$. We write $K \sim \mathcal{B}(\cdot, N, p)$ to denote that K, which may take specific values from 0 to N, has probabilities generated by a binomial distribution with parameters N and p.

Since $\mathcal{B}(k, N, p)$ is the probability of k heads in N flips and the outcome of each flip must be either heads or tails, when we sum these probabilities over all values of k we must obtain 1. That is, $\sum_{k=0}^{N} \mathcal{B}(k, N, p) = 1$.

The expected or mean number of times the coin lands on heads is $\mathcal{E}(K|N, p) = Np$, and the variance in the number of times it lands on heads is $Var(K \mid N, p) = Np(1 - p)$.[38] Thus, the coefficient of variation (standard deviation divided by the mean) in the number of flips is

$$\begin{aligned} CV(K|N,p) &= \frac{\sqrt{Var(K|N,p)}}{\mathcal{E}(K|N,p)} \\ &= \frac{\sqrt{Np(1-p)}}{Np} \\ &= \frac{1}{\sqrt{N}} \cdot \sqrt{\frac{1-p}{p}} \quad . \end{aligned} \tag{43}$$

The coefficient of variation of K, a measure of dispersion around the mean value, scales as $N^{-0.5}$; this is relevant in our discussion of test precision.

[38] Feller, *Probability Theory*.

The Binomial Likelihood

Our second analytical tool is the binomial likelihood. Instead of asking about the chances of a certain number of times a coin will land on heads given the number of flips and the chance of landing on heads on a single flip, we begin with the observation of K heads in N flips and ask what inferences can be made about the probability that the coin will land on heads on a single flip.

The likelihood of p given K and N, which we write as $\mathcal{L}(p \mid K, N)$ where the vertical line is used to separate the quantity about which we wish to make inference (p) and the observations (K and N), is

$$\mathcal{L}(p|K, N) = \binom{N}{k} p^k (1-p)^{N-k}. \tag{44}$$

Intuition suggests that a good estimate for the probability of heads is K/N; the likelihood allows us to quantify that intuition and characterize the uncertainty of the estimate.[39]

The right-hand side of Eq. 44 is exactly the same as the right-hand side of Eq. 42 except that now p is the variable and K is fixed. Three important points about the binomial likelihood are (1) p is a continuous variable ranging from 0 to 1;[40] (2) since the binomial coefficient is independent of p we can write $\mathcal{L}(p|K, N) \propto p^k(1-p)^{N-k}$; and (3) unlike the binomial distribution that sums to 1 as k ranges between 0 and N, the binomial likelihood does not integrate to 1 over p ranges from 0 to 1.

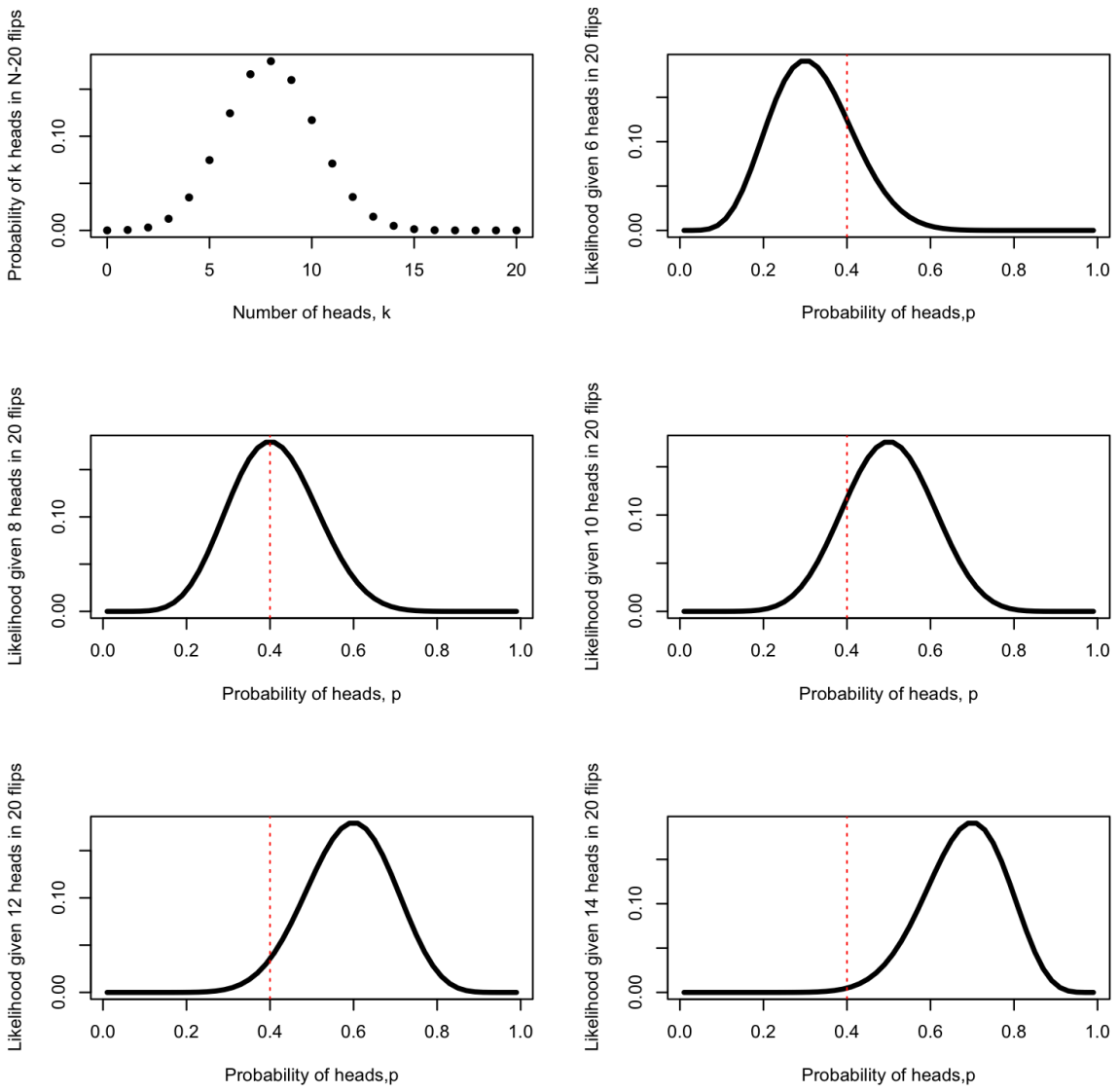
To convert the binomial likelihood to a probability distribution, we normalize the likelihood by its weighted sum. If all values of p are considered equally likely, the appropriate probability density, denoted by $\phi(p|K, N)$, is

$$\phi(p|K, N) = \frac{\mathcal{L}(p|K, N)}{\sum_{p'=0}^{1} \mathcal{L}(p'|K, N)}, \tag{45}$$

where we have replaced what is more properly an integral (since p is continuous) by the summation that we used in the computations, and we use p' to reduce confusion about where p appears on the right-hand side of Eq. 45. That is, the denominator in this equation is a normalization constant.

[39] It is straightforward to show (Mangel, *Theoretical Biologist's Toolbox*) that the value of p that makes the likelihood its maximum is $p_{MLE}(K, N) = K/N$, according to the intuition described above.

[40] Thus, instead of summing, one should integrate. For the results presented here, we discretized the interval between 0 and 1 into increments of 0.001 and replaced integration by summation.



In the upper left-hand panel we show the probability of a coin landing on heads k times in $N = 20$ flips when the probability of heads is $p_{true} = 0.4$. Because we can observe only an integer number of the times the coin lands on heads, we show the probabilities as points. The most likely number of heads, the value of k giving the peak of the points, is 8, which accords with our intuition of having 40% of the 20 flips as heads and is the maximum likelihood estimate (MLE). As k moves away from this most likely value of 8, the probability of observing a given number of heads declines. In the remaining five panels we show the normalized likelihood of different values of the probability of heads p when the observed number of heads is 6 (upper right), 8 (middle left), 10 (middle right), 12 (lower left), or 14 (lower right). Each of these values of the observation is possible, but they have different probabilities of occurring. But once they have occurred, those are the data and we must make inferences about the possible values of p from them. In the panels showing likelihood, we have plotted a vertical line at the true value of the probability of heads. Note that as the value of K increases from 6 to 14, the value of p giving the maximum value of the likelihood increases and accords with our intuition that $p_{MLE}(K, N) = K/N$. If we were to repeat the process of flipping the coin 20 times over and over again, because of the upper left-hand panel we expect that curves similar to the that in the middle left-hand panel will occur more frequently than some of the others.

Figure A-1. Illustration of the Binomial Distribution and Associated Likelihood

If one had prior information about possible values of p, denoted by $\phi_0(p)$, then according to Bayes' theorem of conditional probability, the probability density of p given the data is

$$\phi(p|K,N) = \frac{\mathcal{L}(p|K,N)\phi_0(p)}{\sum_{p'=0}^{1}\mathcal{L}(p'|K,N)\phi_0(p')}. \tag{46}$$

In Figure A-1, we give an example of the binomial distribution and associated normalized binomial likelihood.

The Gaussian Approximation to the Binomial Distribution

When the mean of a binomial distribution Np is bounded away from 0 and N, a normal distribution with mean Np and variance $Np(1-p)$ provides a very good approximation to the binomial distribution.[41] That is, we can then write

$$\mathcal{B}(k,N,p) \approxeq \frac{1}{\sqrt{2\pi Np(1-p)}} exp\left[-\frac{(k-Np)^2}{2Np(1-p)}\right]. \tag{47}$$

[41] Feller, *Probability Theory*, 179–186.

Bibliography

Aguilar, Jacob B., and Juan B. Gutierrez. "An Epidemiological Model of Malaria Accounting for Asymptomatic Carriers." *Bulletin of Mathematical Biology* 82 (2020): 42. https://doi.org/10.1007/s11538-020-00717-y.

Allen, Danielle, Sharon Block, Joshua Cohen, Peter Eckersley, M. Eifler, Lawrence Gostin, Darshan Goux, et al. *Roadmap to Pandemic Resilience. Massive Scale Testing, Tracing, and Supported Isolation (TTSI) as the Path to Pandemic Resilience for a Free Society*. Cambridge, MA: Edmond J. Safra Center for Ethics at Harvard University, April 20, 2020. https://ethics.harvard.edu/files/center-for-ethics/files/roadmaptopandemicresilience_updated_4.20.20_0.pdf.

Aris, R. *Discrete Dynamic Programming: An Introduction to the Optimization of Staged Processes*. Waltham, MA: Blaisdell, 1964.

Atkins, Benjamin D., Chris P. Jewell, Michael C. Runge, Matthew J. Ferrari, Katriona Shea, William J. M. Probert, and Michael J. Tildesley. "Anticipating Future Learning Affects Current Control Decisions: A Comparison between Passive and Active Adaptive Management in an Epidemiological Setting." *Journal of Theoretical Biology* 506 (2020): 110380. https://doi.org/10.1016/j.jtbi.2020.110380.

Auger, Katherine A., Samir S. Shah, Troy Richardson, David Hartley, Matthew Hall, Amanda Warniment, Kristen Timmons, Dianna Bosse, Sarah A. Ferris, Patrick W. Brady, Amanda C. Schondelmeyer, and Joanna E. Thomson. "Association between Statewide School Closure and COVID-19 Incidence and Mortality in the US." *Journal of the American Medical Association (JAMA)* 324, no. 9 (2020): 859–870. https://doi.org/10.1001/jama.2020.14348.

Bi, Q., Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A. Truelove, Tong Zhang, Wei Gao, Cong Cheng, Xiujuan Tang, Xiaoliang Wu, Yu Wu, Binbin Sun, Suli Huang, Yu Sun, Juncen Zhang, Ting Ma, Justin Lessler, and Tiejian Feng. "Epidemiology and Transmission of COVID-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study." *The Lancet* 20, no. 8 (2020): 911–919. https://doi.org/10.1016/S1473-3099(20)30287-5.

Booeshaghi, A. Sina, Fayth Tan, Benjamin Renton, Zackary Berger, and Lior Pachter. "Markedly Heterogeneous COVID-19 Testing Plans among US Colleges and Universities." medRxiv, August 11, 2020. https://doi.org/10.1101/2020.08.09.20171223.

Brown, Alan, and Marc Mangel. *Operational Analysis for Coronavirus Testing: Recommendations for Practice*. National Security Report NSAD-R-21-014. Laurel, MD: Johns Hopkins University Applied Physics Laboratory, 2021. https://www.jhuapl.edu/Content/documents/OperationalAnalysisCoronavirusTesting.pdf.

Carroll, Aaron E. "You Can't Rely on the C.D.C. to Make Your Pandemic Decisions." *New York Times*, July 7, 2021. https://www.nytimes.com/2021/06/07/opinion/pandemic-decision-making.html.

Delamater, Paul L., Erica J. Street, Timothy F. Leslie, Y. Tony Yang, and Kathryn H. Jacobsen. "Complexity of the Basic Reproductive Number (R_0)." *Emerging Infectious Diseases* 25, no. 1 (2019): 1–4. https://doi.org/10.3201/eid2501.171901.

Farmelo, Graham. *It Must Be Beautiful. Great Equations of Modern Science*. London and New York: Granta Books, 2002.

Feller, William. *An Introduction to Probability Theory and Its Applications*. Vol. 1. New York: John Wiley & Sons, 1968.

Green, Daniel A., Jason Zucker, Lars F. Westblade, Susan Whittier, Hanna Rennert, Priya Velu, Arryn Craney, Melissa Cushing, Dakai Liu, Magdalena E. Sobieszczyk, Amelia K. Boehme, and Jorge L. Sepulveda. "Clinical Performance of SARS-CoV-2 Molecular Tests." *Journal of Clinical Microbiology* 58, no. 8 (2020): e00995-20. https://doi.org/10.1128/JCM.00995-20.

He, Jian-Long, Lin Luo, Zhen-Dong Luo, Jian-Xun Lyu, Ming-Yen Ng, Xin-Ping Shen, and Zhibo Wen. "Diagnostic Performance between CT and Initial Real-Time RT-PCR for Clinically Suspected 2019 Coronavirus Disease (COVID-19) Patients outside Wuhan, China." *Respiratory Medicine* 168, article 105980 (2020): 1–5. https://doi.org/10.1016/j.rmed.2020.105980.

He, Xi, Eric H. Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yujuan Guan, Xinghua Tan, Xiaoneng Mo, Yanqing Chen, Baolin Liao, Weilie Chen, Fengyu Hu, Qing Zhang, Mingqiu Zhong, Yanrong Wu, Lingzhai Zhao, Fuchun Zhang, Benjamin J. Cowling, Fang Li, and Gabriel M. Leung. "Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19." *Nature Medicine* 26 (2020): 672–675. https://doi.org/10.1038/s41591-020-0869-5.

Hilborn, Ray, and Marc Mangel. *The Ecological Detective: Confronting Models with Data*. Princeton, NJ: Princeton University Press, 1997.

Hudson, D. J. "Interval Estimation from the Likelihood Function." *Journal of the Royal Statistical Society. Series B (Methodological)* 33, no. 2 (1971): 256–262. https://www.jstor.org/stable/2985006.

Jones, Nicholas R., Zeshan U. Qureshi, Robert J. Temple, Jessica P. J. Larwood, Trisha Greenhalgh, and Lydia Bourouiba. "Two Metres or One: What Is the Evidence for Physical Distancing in covid-19?" *BMJ* 370 (2020): m3223. https://doi.org/10.1136/bmj.m3223.

Klompas, Michael, Meghan A. Baker, and Chanu Rhee. "Airborne Transmission of SARS-CoV-2. Theoretical Considerations and Available Evidence." *Journal of the American Medical Association (JAMA)* 324, no. 5 (2020): 441–442. https://doi.org/10.1001/jama.2020.12458.

Koopman, Bernard Osgood. *Search and Screening. General Principles with Historical Applications*. New York: Pergamon Press, 1980.

Kucirka, Lauren M., Stephen A. Lauer, Oliver Laeyendecker, Denali Boon, and Justin Lessler. "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time since Exposure." *Annals of Internal Medicine* 173 (2020): 262–267. https://doi.org/10.7326/M20-1495.

Li, Ruiyun, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. "Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV2)." *Science* 368, no. 6490 (2020): 489–493. https://doi.org/10.1126/science.abb3221.

Mangel, Marc. *Decision and Control in Uncertain Resource Systems*. New York: Academic Press, 1985.

———. *The Theoretical Biologist's Toolbox*. Cambridge: Cambridge University Press, 2006.

McElreath, Richard. *Statistical Rethinking. A Bayesian Course with Examples in R and Stan*. 2nd ed. CRC Press, Boca Raton, FL, 2020.

Morey, Richard D., Rink Hoekstra, Jeffrey N. Rouder, Michael D. Lee, and Eric-Jan Wagenmakers. "The Fallacy of Placing Confidence in Confidence Intervals." *Psychonomic Bulletin & Review* 23 (2016): 103–123. https://link.springer.com/article/10.3758/s13423-015-0947-8.

National Academies of Sciences, Engineering, and Medicine. *Reopening K–12 Schools during the COVID-19 Pandemic: Prioritizing Health, Equity, and Communities*. Washington, DC: National Academies Press, 2020. https://doi.org/10.17226/25858.

Oran, Daniel P., and Eric J. Topol. "Prevalence of Asymptomatic SARS-CoV-2 Infection. A Narrative Review." *Annals of Internal Medicine* 173, no. 5 (2020): 362–367. https://doi.org/10.7326/M20-3012.

Shelton, Andrew O., and Marc Mangel. "Reply to Sugihara et al: The Biology of Variability in Fish Populations." *Proceedings of the National Academy of Sciences* 108, no. 48 (2011): E1126. https://doi.org/10.1073/pnas.1115765108.

Shental, Noam, Shlomia Levy, Vered Wuvshet, Shosh Skorniakov, Bar Shalem, Aner Ottolenghi, Yariv Greenshpan, Rachel Steinberg, Avishay Edri, Roni Gillis, Michal Goldhirsh, Khen Moscovici, Sinai Sachren, Lilach M. Friedman, Lior Nesher, Yonat Shemer-Avni, Angel Porgador, and Tomer Hertz. "Efficient High-Throughput SARS-CoV-2 Testing to Detect Asymptomatic Carriers." *Science Advances* 6, no. 37 (2020): 1–8. https://doi.org/10.1126/sciadv.abc5961.

Stone, Lawrence D., Johannes O. Royset, and Alan R. Washburn. *Optimal Search for Moving Targets*. Heidelberg: Springer Verlag, 2016.

Watson, Jessica, Penny F. Whiting, and John E. Brush. "Interpreting a Covid-19 Test Result." *BMJ* 369 (2020): m1808. https://doi.org/10.1136/bmj.m1808.

Wilson, Erica, Catherine V. Donovan, Margaret Campbell, Thevy Chai, Kenneth Pittman, Arlene C. Seña, Audrey Pettifor, David J. Weber, Aditi Mallick, Anna Cope, Deborah S. Porterfield, Erica Pettigrew, and Zack Moore. "Multiple COVID-19 Clusters on a University Campus—North Carolina, August 2020." *Morbidity and Mortality Weekly Report (MMWR)* 69, no. 39 (2020): 1416–1418. http://dx.doi.org/10.15585/mmwr.mm6939e3.

Zuckerman, Solly. *Beyond the Ivory Tower. The Frontiers of Public and Private Science*. New York: Taplinger Publishing Company, 1970.

———. *Scientists and War. The Impact of Science on Military and Civil Affairs*. New York: Harper & Row, 1967.

Acknowledgments

Marc Mangel thanks Bethann Pflugeisen both for discussing these ideas during their development and for inspiring the notion of a separate report (the companion to this one) on recommendations for practice, and Spiro Stefanou for reminding him of using the Aris quotation in a graduate class in 1980. We thank Joseph Travis (Florida State University) for a careful reading of the manuscript, many helpful comments, and thoughtful questions that led to the new recommendation for practice.

About the Authors

Dr. Marc Mangel is distinguished professor emeritus at the University of California Santa Cruz and professor emeritus at the University of Bergen. He has applied operations and systems analysis widely, including search theory in agricultural pest control and fisheries, stochastic optimization for management of fisheries and understanding the life history patterns of salmon and southern ocean krill, and nonlinear dynamical systems in the study of disease. After receiving his PhD, Dr. Mangel worked in the Operations Evaluation Group (OEG) of the Center for Naval Analyses (CNA), including in a field position at Naval Air Station Whidbey Island. From there, he moved to the University of California Davis with the goal of doing OEG-style work with applications in fisheries and agriculture. In collaboration with colleagues at APL, he is now bringing ideas from population biology to describe variability, compromise, and recovery of cyber systems. During his career, Dr. Mangel has provided leadership and advice to many national and international organizations, most recently the Scientific Review Board of the International Pacific Halibut Commission. In 2013, he was the principal scientific expert for Australia in the International Court of Justice case *Whaling in the Antarctic: Australia v. Japan. New Zealand Intervening*. Dr. Mangel has authored and co-authored numerous journal articles and books, including *Decision and Control in Uncertain Resource Systems*, *Dynamic Modeling in Behavioral Ecology*, *The Ecological Detective: Confronting Models with Data*, and *The Theoretical Biologist's Toolbox*. He holds a BSc (physics) and an MSc (biophysics) from the University of Illinois and a PhD (applied mathematics and statistics) from the University of British Columbia; he was awarded a doctor of science *honoris causa* from the University of Guelph in 2014.

Dr. Alan Brown researches operational issues in APL's Asymmetric Operations Sector (AOS). Recently, he supported the National Response Coordination Center at FEMA, leading the future operations analysis cell during the first months of the COVID-19

pandemic. Currently, he leads analysis in support of development and execution of a serologic testing program for Howard County, Maryland. In addition to supporting these health-related analysis efforts, Dr. Brown conducts research on cyber issues as a member of the AOS Transformational Projects Office, serves as chief scientist for the AOS Mission Analysis Group, and was recently technical lead for systems engineering analysis and operational testing for the deputy secretary of defense special program Missile Defeat Time Critical Targeting. He joined APL in 2015 after thirty years at the Center for Naval Analyses (CNA), where his work combined analyses of capabilities, operations, resource requirements, and policies to inform senior-level decision-makers as they grappled with complex challenges. Dr. Brown led CNA's Operational Policy Team, which analyzed real-world operations and the policies that govern them, including mission-focused organization; war gaming; tactic, technique, and procedure development; and training for operational units. He holds a PhD from Brown University and a BS from Worcester Polytechnic Institute, both in physics.

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY