medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
Epidemiology, Medical Statistics, Applied Mathematics

Keywords:
COVID-19, testing, combinatorics

Author for correspondence:
Lucas Böttcher
e-mail: lucasb@g.ucla.edu

A statistical model of COVID-19 testing in populations: effects of sampling bias and testing errors

Lucas Böttcher[1,2], Maria R. D'Orsogna[3,1] and Tom Chou[1,4]

[1]Dept. of Computational Medicine, University of California, Los Angeles, 90095-1766, Los Angeles, CA, United States
[2]Computational Social Science, Frankfurt School of Finance and Management, 60322 Frankfurt am Main, Germany
[3]Dept. of Mathematics, California State University at Northridge, Los Angeles, 91330-8313, CA, United States
[4]Dept. of Mathematics, University of California, Los Angeles, 90095-1766, Los Angeles, CA, United States

We develop a statistical model for the testing of disease prevalence in a population. The model assumes a binary test result, positive or negative, but allows for biases in sample selection and both type I (false positive) and type II (false negative) testing errors. Our model also incorporates multiple test types and is able to distinguish between retesting and exclusion after testing. Our quantitative framework allows us to directly interpret testing results as a function of errors and biases. By applying our testing model to COVID-19 testing data and actual case data from specific jurisdictions, we are able to estimate and provide uncertainty quantification of indices that are crucial in a pandemic, such as disease prevalence and fatality ratios.

© The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1. Introduction

Real-time estimation of the level of infection in a population is important for assessing the severity of an epidemic as well as for guiding mitigation strategies. However, inferring disease prevalence via patient testing is challenging due to testing inaccuracies, testing biases, and heterogeneous and dynamically evolving populations and severity of the disease.

There are two major classes of tests that are used to detect previous and current SARS-CoV-2 infections [1]. Serological, or antibody, tests measure the concentration of antibodies in infected and recovered individuals. Since antibodies are generated as a part of the adaptive immune system response, it takes time for detectable antibody concentrations to develop. Serological tests should thus not be used as the only method to detect acute SARS-CoV-2 infections. An alternative testing method is provided by viral-load or antigen tests, such as reverse transcription polymerase chain reaction (RT-PCR), enzyme-linked immunosorbent assay (ELISA), and rapid antigen tests, which are able to identify ongoing SARS-CoV-2 infections by directly detecting SARS-CoV-2 nucleic acid or antigen.

Test results are mainly reported as binary values (0 or 1, negative or positive) and often do not include further information such as the cycle threshold (Ct) for RT-PCR tests. The cycle threshold Ct defines the minimum number of PCR cycles at which amplified viral RNA becomes detectable. Large values of Ct indicate low viral loads in the specimen. An increase in Ct by a factor of about 3.3 corresponds to a viral load that is about one order of magnitude lower [2]. Cycle threshold cutoffs are not standardized across jurisdictions and range from values between $37-40$, making it difficult to compare RT-PCR test results [3]. Lower Ct cutoffs in the range of $30-35$ may be more reasonable to avoid classifying individuals with insignificant viral loads as positive [3].

Further uncertainty in COVID-19 test results arises from different type I errors (false positives) and type II errors (false negatives) that are associated with different assays. Note that inherent to any test, the threshold (such as Ct mentioned above) may be tunable. Therefore, besides intrinsic physical limitations, binary classification of "continuous-valued" readouts (*e.g.*, viral load) may also lead to an overall error of either type [4]. In this work, we will assume that there is a standardized threshold and the test readout is binary; if any virus is detected, the test subject is positive. We will not explicitly model the underlying statistics of the errors but assume that the test readouts are binary but can be erroneous at specified rates. Some uninfected individuals will be wrongly classified as infected with rate FPR and some infected individuals will be wrongly classified as uninfected with rate FNR. For serological COVID-19 tests, the estimated proportions of false positives and false negatives are relatively low, with $\mathrm{FPR}\approx 0.02-0.07$ and $\mathrm{FNR}\approx 0.02-0.16$ [5–8]. The FNRs of RT-PCR tests depend strongly on the actual assay method [9,10] and may be significantly larger than those of serological tests. Typical values of FNR for RT-PCR tests lie between 0.1 and 0.3 [11,12] but might be as high as $\mathrm{FNR}\approx 0.68$ if throat swabs are used [7,12]. False-negative rates may also vary significantly depending on the time delay between initial infection and testing [8]. According to a systematic review [13] that was conducted worldwide, the initial value of FNR is about 0.54, underlying the importance of retesting. Similar to serological tests, reported false-positive rates of RT-PCR tests are about $\mathrm{FPR}=0.05$ [7].

Estimates of disease prevalence and other surveillance metrics [14,15] need to account for FPRs and FNRs, in particular if reported positive-testing rates [16] are in the few percent range and potentially dominated by type I errors. In addition to type-I/II testing errors, another confounding effect is biased testing [17], that is, preferential testing of individuals that are expected to carry a high viral load (*e.g.*, symptomatic and hospitalized individuals). Biasing testing towards certain demographic and risk groups leads to additional errors in disease prevalence estimates that need to be corrected for.

To account for type-I/II errors, bias, retesting and exclusion after testing, we develop a corresponding framework for disease testing in Sec. 2. We apply our testing model to COVID-19 testing and case data in Sec. 3 and estimate testing bias by comparing random-sampling testing

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

data [18] with officially reported, biased COVID-19 case data in Sec. 4. We conclude our study in Sec. 5.

2. Statistical testing model

Here, and in the following subsections, we develop a general statistical model for estimating the number of infected individuals in a jurisdiction by testing a sample population. The relevant variables and parameters to be used in our derivations are listed and defined in Table 1.

symbol	definition
$N \in \mathbb{Z}^+$	population in jurisdiction
$Q : [0, N]$	number of tests administered
$Q^+ : [0, Q]$	recorded positives under error-free testing
$\tilde{Q}^+ : [0, Q]$	recorded positives under error-prone testing
$f : [0, 1]$	true proportion of infected individuals
$f_{\mathrm{b}} := Q^+/Q : [0, 1]$	fraction of positives under biased, error-free testing
$\tilde{f}_{\mathrm{b}} := \tilde{Q}^+/Q : [0, 1]$	fraction of positives under biased, error-prone testing
$b \in \mathbb{R}$	testing bias parameter
$\hat{b}, \hat{f} : [0, 1]$	estimates of bias and underlying infection fraction
$\mathrm{FPR} : [0, 1]$	false positive rate
$\mathrm{FNR} : [0, 1]$	false negative rate

Table 1. Overview of variables used in testing model. An overview of the main variables and parameters that will be used in developing our testing model. The sets $[0, N]$ and $[0, Q]$ contain all integers from 0 up to Q and N, respectively. The set $f, f_{\mathrm{b}}, \tilde{f}_{\mathrm{b}} : [0, 1]$ denotes all rational numbers between 0 and 1. For $\mathrm{FNR}, \mathrm{FPR}, [0, 1]$ represents all real numbers between 0 and 1. We assume that Q, Q^+, $\tilde{Q}^+$, $f_{\mathrm{b}} := Q^+/Q$, and $\tilde{f}_{\mathrm{b}} := \tilde{Q}^+/Q$ are determined by testing a population of known size N.

Suppose we randomly administer Q tests within a particular short time period (*e.g.*, within one day or one week) to a total effective population of N previously untested individuals. This population of individuals is comprised of S susceptible, I infected, and R removed (*i.e.*, recovered or deceased) individuals, which are unknown. S, I, and R can dynamically change from one testing period to another due to transmission and recovery dynamics, as well as removal from the untested pool by virtue of being tested. The total population $N = S + I + R$ can also change through intrinsic population dynamics (birth, death, and immigration), but can assumed to be constant over the typical time scale of an epidemic that does not cause mass death.

We start the derivation of our statistical model by first fixing S, I and R, assuming both perfect error-free testing, considering a "testing with replacement" scenario, in which tested individuals can be retested within the same time window. Under these conditions, the probability that q tests

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

are returned positive and $Q^- = Q - q$ tests are returned negative is

$$P_{\text{true}}(Q^+ = q|Q, S, I, R) = \binom{Q}{q} f^q (1-f)^{Q-q}, \tag{2.1}$$

where the parameter

$$f \equiv \frac{(I+R)}{N} \quad \text{or} \quad \frac{I}{N}, \tag{2.2}$$

are simply the probabilities of identifying currently and previously infected individuals with tests such as serological (antibody) tests, or of detecting current infections with viral load tests, respectively. Note that testing with replacement renders P_{true} dependent only on Q and f, and not explicitly on I, S, R or N. The binomial expression 2.1 is accurate when the number of tests are much smaller than the population $(I+R)$ or I.

Equation 2.1 describes perfect error-free and random testing. However, if there is some prior suspicion of being infected, the administration of testing may be biased. For example, certain jurisdictions focus testing primarily on hospitalized patients and people with significant symptoms [17], thus biasing the tests to those that are infected. We quantify such testing biases through a biased-testing function $B(Q^+) \in \mathbb{R}_{\geq 0}$, leading to the following modification of Eq. 2.1:

$$P_{\text{true}}(Q^+ = q|Q, f) = \frac{\binom{Q}{q} f^q (1-f)^{Q-q} B(q)}{\sum_{k=0}^{Q} \binom{Q}{k} f^k (1-f)^{Q-k} B(k)}. \tag{2.3}$$

We define the biased-testing function $B(q)$ as a weighting over certain numbers q of positive tests. A convenient choice of this weighting is a "Boltzmann" functional form $B(q) = e^{qU_I} e^{(Q-q)U_S}$, for which $P_{\text{true}}(Q^+ = q|f, Q)$ becomes

$$\begin{aligned} P_{\text{true}}(Q^+ = q|Q, f, b) &= \frac{\binom{Q}{q} f^q (1-f)^{Q-q} e^{qb}}{\sum_{k=0}^{Q} \binom{Q}{k} f^k (1-f)^{Q-k} e^{kb}} \\ &= \frac{\binom{Q}{q} (fe^b)^q (1-f)^{Q-q}}{[1+(e^b-1)f]^Q}, \end{aligned} \tag{2.4}$$

where $-U_I, -U_S$ represent "costs" for testing infecteds, susceptibles, and $b \equiv U_I - U_S \in \mathbb{R}$ is a testing-bias parameter, which we employ below.

A Gaussian approximation to Eq. 2.4 can be found through the mean and variance of Q^+:

$$P_{\text{true}}(Q^+ = q|Q, f, b) \approx \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(q-\mu(f,b))^2}{2\sigma^2(f,b)}\right], \tag{2.5}$$

where

$$\mu(f,b) \equiv \frac{Qfe^b}{1+f(e^b-1)} \quad \text{and} \quad \sigma^2(f,b) \equiv \mu(f,b)\left(1 - \frac{\mu(f,b)}{Q}\right). \tag{2.6}$$

In addition to describing the probability distribution in Eq. 2.5 as a function of the number of positive tests Q^+, we can also express it in terms of the observed positive (and potentially sample-biased) testing fraction $f_{\text{b}} := Q^+/Q$:

$$P_{\text{true}}(f_{\text{b}} = x|Q, f, b) \approx \frac{1}{\bar{\sigma}(f,b)\sqrt{2\pi}} \exp\left[-\frac{(x-\bar{\mu}(f,b))^2}{2\bar{\sigma}^2(f,b)}\right], \tag{2.7}$$

where here,

$$\bar{\mu}(f,b) \equiv \mu/Q = \frac{fe^b}{1+f(e^b-1)} \quad \text{and} \quad \bar{\sigma}^2(f,b) \equiv \bar{\mu}(f,b)\left(1-\bar{\mu}(f,b)\right)/Q. \tag{2.8}$$

The expected value of f_{b}, $\bar{\mu}(f,b)$, can be understood as a product of the true underlying infected fraction f and a bias function $\tilde{B}(f,b)$ that depends on f and the bias parameter b, *i.e.*, $\bar{\mu}(f,b) =$

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

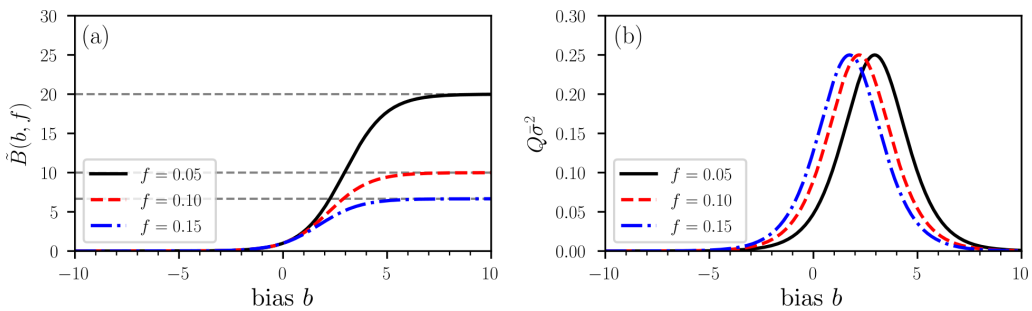

Figure 1. Illustration of a bias function $\tilde{B}(f,b)$. (a) The bias function $\tilde{B}(f,b)$ (Eq. 2.9) for three different fractions f of currently (and previously) infected individuals. Grey dashed lines indicate the asymptotic value $\tilde{B}(\infty,f)=1/f$. A value $b>0$ indicates a testing bias towards currently and/or previously infected individuals while susceptible and/or non-infectious individuals are preferentially tested for $b<0$. Unbiased testing corresponds to $b=0$ and $\tilde{B}(0,f)=1$. (b) The variance $\bar{\sigma}^2$ exhibits a maximum value of $1/(4Q)$ at a typical value of bias $b^*=\ln[(1-f)/f]$.

$f\tilde{B}(f,b)$, where $\tilde{B}$ is given by

$$\tilde{B}(f,b)=\frac{e^b}{1+f(e^b-1)}. \tag{2.9}$$

Note, that Eq. 2.9 implies that when $b>0$, the currently and/or previously infected population is favored to be tested, while for $b<0$, the non-infected and/or susceptible population is favored. The limits of $b\to\pm\infty$ indicate testing that is completely biased such that only infected and susceptible individuals are tested, respectively. Realistic values of our bias parameter b are positive and $\sim O(1)$.

Fig. 1(a) shows the bias function 2.9 as a function of b for different infection fractions f. For $\tilde{B}(f,b)>1$, the biased-testing fraction $f\tilde{B}$ is larger than the unbiased-testing fraction f. The opposite holds for $\tilde{B}(f,b)<1$. The variance $\bar{\sigma}^2$ is plotted as a function of b in Fig. 1(b) and exhibits a maximum value of $1/(4Q)$ at $b^*=\ln[(1-f)/f]$.

The probabilities P_{true} derived in Eqs. 2.1 and 2.3 correspond to "testing with replacement". The opposite limit is "testing without replacement"; once an individual is tested they are labeled as such and removed from the pool of test targets, at least within the specified testing period. This concept of sampling with and without replacement commonly arises in the measurement of diversity in ecological settings [19]. Without replacement, and still under conditions of perfect random testing, two slightly different forms for P_{true} arise for the different type of tests (*e.g.*, antibody *vs* PCR/viral load). For antibody tests that perfectly identifies recovered (or deceased) individuals as being previously infected, Eq. 2.1 is replaced by

$$\begin{aligned} P_{\text{true}}(Q^+=q|S,I,R,Q) &= \frac{\binom{I+R}{q}\binom{S}{Q-q}}{\binom{N}{Q}} \\ &= \binom{Q}{q}\prod_{n=0}^{Q-1}\left(\frac{1}{N-n}\right)\prod_{i=0}^{q-1}(I+R-i)\prod_{j=0}^{Q-q-1}(S-j), \end{aligned} \tag{2.10}$$

where the binomial coefficients for $q=0,Q$ are zero. On the other hand, if the perfect test only identifies individuals that currently have a viral load, the susceptible and recovered (or deceased)

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

individuals both test negative and $P_{\rm true}$ is described by

$$P_{\rm true}(Q^+ = q|S, I, R, Q) = \frac{\binom{I}{q}\binom{S+R}{Q-q}}{\binom{N}{Q}}$$
$$= \binom{Q}{q} \prod_{n=0}^{Q-1} \left(\frac{1}{N-n}\right) \prod_{i=0}^{q-1}(I-i) \prod_{j=0}^{Q-q-1}(S+R-j). \tag{2.11}$$

The expressions for $P_{\rm true}$ when tested subjects are not replaced, unlike in the case of testing with replacement, depend explicitly on S, I, R, and N.

To incorporate testing bias in into the probabilities $P_{\rm true}$ for testing without replacement, first consider Eqs. 2.11 and think of $\binom{I}{q}\binom{S+R}{Q-q}$ as the number of ways of distributing q positive tests among I infected individuals, and $Q-q$ negative tests among $S+R$ uninfected individuals. As in the biased-testing formulation of Eq. 2.3, we interpret the bias as a factor $B(q)$ that weights more tests in I or $S+R$ pools:

$$P_{\rm true}(Q^+ = q|Q, S, I, R, B) \equiv \frac{\binom{I}{q}\binom{S+R}{Q-q}B(q)}{\sum_{k=0}^{Q}\binom{I}{k}\binom{N-I}{Q-k}B(k)}. \tag{2.12}$$

To obtain the "testing without replacement" equivalent of Eq. 2.4, we again set $B(q) = e^{qU_I}e^{(Q-q)U_S} = e^{bq}e^{QU_S}$. By using the Chu–Vandermonde identity

$$\sum_{k=0}^{Q}\binom{I}{k}\binom{N-I}{Q-k} = \binom{N}{Q}, \tag{2.13}$$

we verify that when $b=0$ the expression 2.12 reduces to Eq. 2.11.

For the exponential form of B, the normalizing "partition function" becomes

$$\sum_{k=0}^{Q}\binom{I}{k}\binom{N-I}{Q-k}e^{kb} = \binom{N-I}{Q}\,{}_2F_1(-I, -Q; N-I-Q+1; e^b), \tag{2.14}$$

where ${}_2F_1$ denotes the (ordinary) hypergeometric function. Thus, the distribution of positive tests Q^+ under biased testing without replacement for viral load-type tests can thus be expressed as

$$P_{\rm true}(Q^+ = q|Q, S, I, R, b) \equiv \frac{e^{qb}\binom{Q}{q}\prod_{k=0}^{q-1}\left(\frac{I-k}{S+R-Q^- -k}\right)}{{}_2F_1(-I, -Q; N-I-Q+1; e^b)} \tag{2.15}$$

The distribution of positives under biased testing without replacement for antibody-type tests becomes (analogous to Eq. 2.10) is

$$P_{\rm true}(Q^+ = q|Q, S, I, R, b) \equiv \frac{e^{qb}\binom{Q}{q}\prod_{k=0}^{q-1}\left(\frac{I+R-k}{S-Q^- -k}\right)}{{}_2F_1(-I-R, -Q; N-I-R-Q+1; e^b)} \tag{2.16}$$

where $Q^- \equiv Q-q$. The above two expressions are equivalent except for the merging of the R pool with uninfected susceptibles S in one case, or with current infecteds I in the other. Because the testing decreases with the number of tests administered, Eqs. 2.15 and 2.16 cannot be reduced to functions of a simple positive-test fraction $f_{\rm b}$ or to universally accurate simple Gaussian forms.

(a) Testing errors

The probability distributions $P_{\rm true}$ that we derived in Eq. 2.3 and in Eqs. 2.10–2.11 assume that testing is error-free, *i.e.*, that the false-negative rate $\mathrm{FNR} = 1 - \mathrm{TPR} = 0$ and false-positive rate $\mathrm{FPR} = 1 - \mathrm{TNR} = 0$, or equivalently that the true-positive rate $\mathrm{TPR} = 1$ and the true-negative rate $\mathrm{TNR} = 1$. To incorporate erroneous testing, we now construct the probability distribution of

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

error-generated deviation $P_{\rm err}(\tilde{Q}^+ = q|Q^+ = q')$ over the number of "apparent" positives q from tests that carry nonzero FPRs and FNRs, given that q' positives would be recorded if the tests were perfect. If q apparent positive tests are tallied, p of them might have been true positives drawn from the perfect-test positives q' in $\binom{q'}{p}$ ways, while the remaining $k = q - p$ apparent positives might have been erroneously counted as positives drawn from the $Q^- \equiv Q - q'$ true negatives. The remaining $q' - p$ true positive tests might have been erroneously tallied as false negatives, while the remaining $Q^- - k$ negative tests might have been correctly tallied as true negatives. Assuming nonzero FPR and FNR, we find that the probability distribution of finding $0 \le q \le Q$ apparent positive tests is

$$P_{\rm err}(\tilde{Q}^+ = q|Q^+ = q', Q, {\rm FPR}, {\rm FNR}) = \sum_{p=0}^{q} \binom{q'}{p} ({\rm TPR})^p ({\rm FNR})^{q'-p} \binom{Q-q'}{q-p} ({\rm FPR})^{q-p} ({\rm TNR})^{Q-q'-(q-p)}, \tag{2.17}$$

where we invoked the identities TPR + FNR = 1 and FPR + TNR = 1. The total error-prone distribution $P_{\rm TOT}(\tilde{Q}^+ = q|Q, S, I, R, b)$ of recording q positives after having administered Q tests under bias and/or testing errors is given by convolving the probability $P_{\rm err}$ of finding q apparent tests given q' true positive tests with the probability $P_{\rm true}$ of finding q' positive tests under perfect testing (Eq. 2.4 or Eq. 2.15):

$$P_{\rm TOT}(\tilde{Q}^+ = q|Q, S, I, R, b, {\rm FPR}, {\rm FNR}) = \sum_{q'=0}^{Q} P_{\rm err}(\tilde{Q}^+ = q|Q^+ = q', Q, {\rm FPR}, {\rm FNR}) P_{\rm true}(Q^+ = q'|Q, S, I, R, b). \tag{2.18}$$

This convolution can be further simplified by taking the Gaussian limit of the binomial distributions that appear in $P_{\rm err}$ and in $P_{\rm true}$. To be concrete, we use $P_{\rm true}$ as written in Eq. 2.3 under the testing with replacement scenario and use the approximation 2.5. The same Gaussian approximation can be used for all terms in $P_{\rm err}$ from Eq. 2.17. After approximating the summation over $p \in (0, q)$ in Eq. 2.17 by an integral over $p \in (-\infty, \infty)$ (provided q is sufficiently large), we find

$$P_{\rm err}(\tilde{Q}^+ = q|Q, Q^+ = q', {\rm FPR}, {\rm FNR}) \approx \frac{\exp\left[-\frac{(q - q'{\rm TPR} - (Q-q'){\rm FPR})^2}{2q'{\rm FNR}\,{\rm TPR} + 2(Q-q'){\rm FPR}\,{\rm TNR}}\right]}{\sqrt{2\pi}\sqrt{q'{\rm FNR}\,{\rm TPR} + (Q-q'){\rm FPR}\,{\rm TNR}}}. \tag{2.19}$$

We now convolve Eq. 2.5 with Eq. 2.19 as prescribed by Eq. 2.18 to find

$$P_{\rm TOT}(\tilde{Q}^+ = q|Q, f, b, {\rm FPR}, {\rm FNR}) \approx \frac{1}{\sigma_{\rm T}\sqrt{2\pi}} \exp\left[-\frac{(q-\mu_{\rm T})^2}{2\sigma_{\rm T}^2}\right], \tag{2.20}$$

where

$$\begin{aligned} \mu_{\rm T}(f, b, {\rm FPR}, {\rm FNR}) &\equiv Q\left[\bar{\mu}(f,b)(1-{\rm FNR}) + (1-\bar{\mu}(f,b)){\rm FPR}\right], \\ \sigma_{\rm T}^2(f, b, {\rm FPR}, {\rm FNR}) &\equiv Q(1-\bar{\mu}(f,b)){\rm FPR}(1-{\rm FPR}) + Q\bar{\mu}(f,b){\rm FNR}(1-{\rm FNR}) \\ &\quad + Q\bar{\mu}(f,b)(1-\bar{\mu}(f,b))(1-{\rm FNR}-{\rm FPR})^2. \end{aligned} \tag{2.21}$$

with $\bar{\mu}(f, b)$ the mean value of the fraction of observed positives under biased, but perfect testing (see Eq. 2.8). The mean number of apparent positive tests $\mu_{\rm T}$ is given by the sum of the expected value of true positive tests (*i.e.*, $Q\bar{\mu}(f,b)(1-{\rm FNR}) = Q\bar{\mu}(f,b){\rm TPR}$) and the expected value of false positive tests (*i.e.*, $Q(1-\bar{\mu}(f,b)){\rm FPR}$). Based on the derived expressions for $\mu_{\rm T}$ and $\sigma_{\rm T}^2$, we define the random variable as the fraction of observed positive tests $\tilde{f}_{\rm b} \equiv \tilde{Q}^+/Q$ under biased and

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

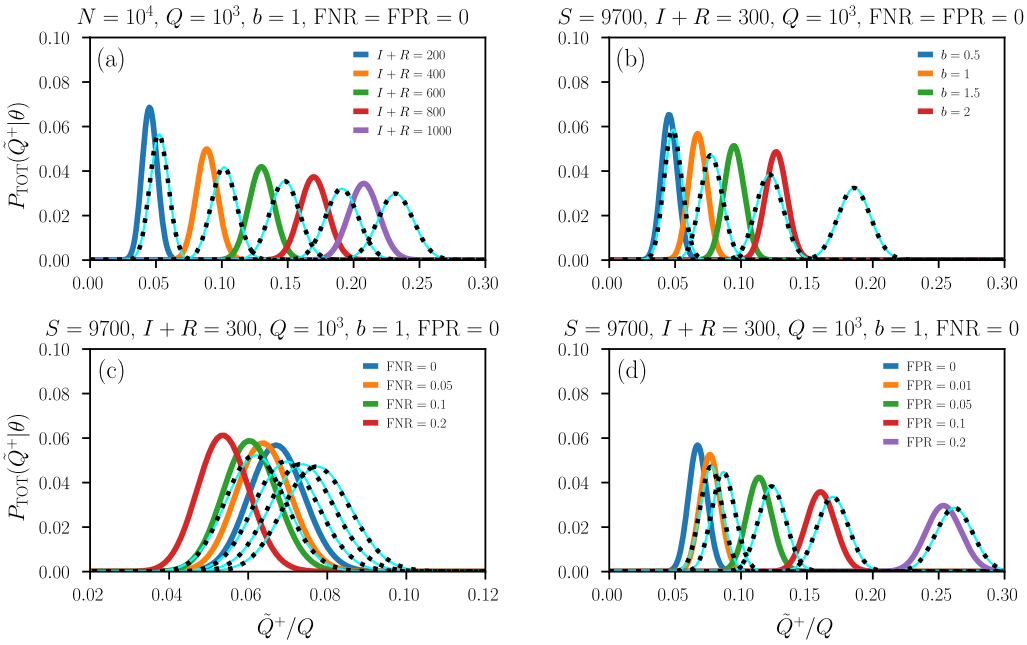

Figure 2. Distribution of apparently positive tests. Plots of $P_{\mathrm{TOT}}(\tilde{Q}^+ = q|Q, I, S, R, b)$ with $N = S + I + R = 10^4$, $Q = 10^3$, and different (a) values of $I + R$, (b) testing biases b, (c) FNRs, and (d) FPRs. The Gaussian approximation (solid light blue lines) of Eq. 2.22 provides an accurate approximation of P_{TOT}. Dashed black lines correspond to distributions with replacement and the remaining solid colored lines correspond to those without replacement.

error-prone testing and obtain

$$P_{\mathrm{TOT}}(\tilde{f}_{\mathrm{b}} = x|Q, f, b, \mathrm{FPR}, \mathrm{FNR}) \approx \frac{e^{-(x-\bar{\mu}_{\mathrm{T}})^2/(2\bar{\sigma}_{\mathrm{T}}^2)}}{\bar{\sigma}_{\mathrm{T}}\sqrt{2\pi}}, \tag{2.22}$$

where

$$\begin{aligned} \bar{\mu}_{\mathrm{T}}(f, b, \mathrm{FPR}, \mathrm{FNR}) &\equiv [\bar{\mu}(f, b)(1 - \mathrm{FNR}) + (1 - \bar{\mu}(f, b))\mathrm{FPR}]\,, \\ Q\bar{\sigma}_{\mathrm{T}}^2(f, b, \mathrm{FPR}, \mathrm{FNR}) &\equiv (1 - \bar{\mu}(f, b))\mathrm{FPR}(1 - \mathrm{FPR}) + \bar{\mu}(f, b)\mathrm{FNR}(1 - \mathrm{FNR}) \\ &\quad + \bar{\mu}(f, b)(1 - \bar{\mu}(f, b))(1 - \mathrm{FNR} - \mathrm{FPR})^2. \end{aligned} \tag{2.23}$$

The Gaussian approximation 2.23 is quite accurate provided that (i) the number of positive and apparent positive tests, $Q^+ = q'$ and $\tilde{Q}^+ = q$, are sufficiently large, and (ii), the quantities $\bar{\mu}$, FNR, and FPR are not too close to 0 or 1. In Table 1, we provide an overview of the main variables used in the statistical testing model (2.17)–(2.23).

Fig. 2 shows the distribution of apparently infected individuals P_{TOT} for different numbers of infected/recovered individuals [Fig. 2 (a)], testing biases [Fig. 2 (b)], and testing sensitivities (*i.e.*, true positive rates, TPRs) and specificities (*i.e.*, true negative rates, TNRs) [Fig. 2 (c–d)]. Solid light blue lines represent the Gaussian approximation 2.22 and dashed black lines and the remaining colored lines are calculated by directly evaluating Eq. 2.18 with replacement (Eq. 2.3) and without replacement (Eq. 2.16), respectively. The FNRs that we consider in Fig. 2 (c) are chosen in accordance with reported sensitivities of serological and RT-PCR tests for SARS-CoV-2 [9–12]. We observe that an increase in the FNR slightly shifts the distribution P_{TOT} towards smaller values of apparently infected individuals, which is consistent with the FNR dependence of the mean $\bar{\mu}_{\mathrm{T}}$ (Eq. 2.23). For serological and RT-PCR tests, the FPR $= 1 -$ TNR is about 5%. A smaller specificity would lead to larger FPRs and a shift of P_{TOT} towards larger values of

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

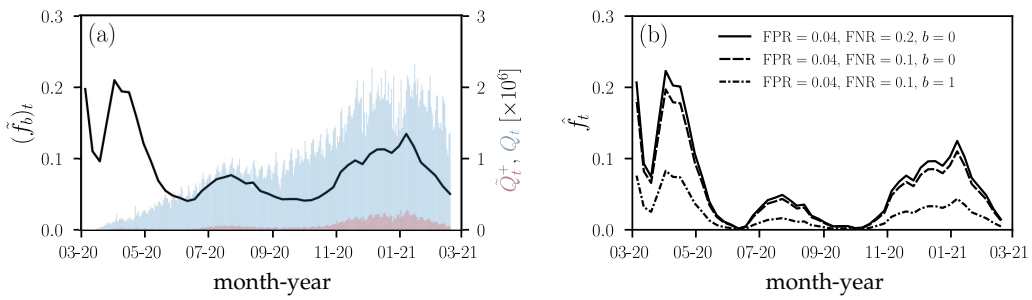

Figure 3. Observed and corrected proportions of positive tests in the United States. (a) The solid black line represents the 7-day average of the proportion of positive tests $(\tilde{f}_\mathrm{b})_t = \tilde{Q}_t^+/Q_t$ in the United States. Blue and red bars show the corresponding total number of daily tests Q and apparent positive tests $\tilde{Q}_t^+$, respectively. (b) The corrected proportion of positive tests $\hat{f}_t$, found by inverting Eq. 3.1, for different FPR, FNR, and bias combinations.

$\tilde{Q}^+$ and $\tilde{f}_\mathrm{b}$ [Fig. 2 (d)]. Our results show that P_TOT is more affected by variations in the testing specificity than by variations in testing sensitivity.

(b) Temporal variations and test heterogeneity

Up to now, we have discussed single viral-load and antibody tests (with and without replacement) but have not considered temporal variations in the number of tests Q, the number of returned positives $\tilde{Q}^+$, and heterogeneity in FNR and FPR that are associated with different classes (types, manufacturing batches, etc.) of assays. To make our model applicable to empirical time-varying testing data, we use S_t, I_t, R_t to denote the number of susceptible, infected, and removed individuals at time t (or in successive time windows labeled by t), respectively. If $K > 1$ test classes are present, we also include an additional index $c \in \{1, \ldots, K\}$ in all relevant model parameters. The testing bias and the total number of tests may be both test-class and time-dependent. That is, $b = b_{t,c}$ and $Q = Q_{t,c}$. Test specificity and sensitivity mainly depend on the assay type and not on time. We thus set $\mathrm{FPR} = \mathrm{FPR}_c$ and $\mathrm{FNR} = \mathrm{FNR}_c$.

3. Inference of prevalence and application to COVID-19 data

For a certain time window, one often wishes to infer $I_t + R_t$ and S_t, or I_t and $S_t + R_t$ from values of $b_{t,c}$, $Q_{t,c}$, and $q_{t,c}$. Alternatively, since $b_{t,c}$ is difficult to independently ascertain, one may only be able to infer $(f_b)_{t,c} = f(S_t, I_t, R_t, b_{t,c})$. For a single test result $\tilde{q}_{t,c}$ (or $(\tilde{f}_\mathrm{b})_{t,c}$), we can generate the maximum likelihood estimate (MLE) of the bias-modified prevalence $(\hat{f}_b)_{t,c}$ by setting the measured value $(\tilde{f}_\mathrm{b})_{t,c} = \bar{\mu}_\mathrm{T}[(\hat{f}_b)_{t,c}]$ to find

$$(\tilde{f}_\mathrm{b})_{t,c} = (\hat{f}_b)_{t,c}(1 - \mathrm{FNR}_c) + (1 - (\hat{f}_b)_{t,c})\mathrm{FPR}_c \tag{3.1}$$

and

$$(\hat{\sigma}_\mathrm{T})^2_{t,c} \approx (\bar{\sigma}_\mathrm{T})^2_{t,c}[(\hat{f}_b)_{t,c}]. \tag{3.2}$$

Since $(\hat{f}_b)_{t,c} = \hat{f}_{t,c}\tilde{B}(\hat{f}_{t,c}, b_{t,c})$, Eq. 3.1 can be solved for $\hat{f}_{t,c}$:

$$\hat{f}_{t,c} = \frac{(\tilde{f}_\mathrm{b})_{t,c} - \mathrm{FPR}_c}{e^b[1 - \mathrm{FNR}_c - (\tilde{f}_\mathrm{b})_{t,c}] + (\tilde{f}_\mathrm{b})_{t,c} - \mathrm{FPR}_c}. \tag{3.3}$$

The posterior distribution P_post over values of f can be found through Bayes' theorem:

$$P_\mathrm{post}(f|\tilde{Q}^+, Q, \mathrm{FPR}, \mathrm{FNR}) = \frac{P_\mathrm{TOT}(\tilde{Q}^+|Q, f, \mathrm{FPR}, \mathrm{FNR})P_0(f)}{\int_0^1 P_\mathrm{TOT}(\tilde{Q}^+|Q, f', \mathrm{FPR}, \mathrm{FNR})P_0(f')\,\mathrm{d}f'}, \tag{3.4}$$

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

where $P_0(f)$ is a prior distribution over the underlying infection fraction f in the population. For notational brevity, we did not include the indices c and t in Eq. 3.4. We can again simplify the analysis by using the Gaussian approximation and a simple initial uniform prior, $P_0(f < f_{\max} \leq 1) = 1/f_{\max}$.

As an example, we collected US testing data [20] from March 2020 to March 2021. Figure 3(a) shows the daily number of observed positive tests $\tilde{Q}_t^+$ (red bars) and the corresponding total daily number of tests Q_t (blue bars). The 7-day average of the observed positive testing rate $(\tilde{f}_{\rm b})_t = \tilde{Q}_t^+/Q_t$ is indicated by the black solid line. The first drop in $(\tilde{f}_{\rm b})_t$ in March 2020 was associated with the initially very limited number of available SARS-CoV-2 testing infrastructure followed by the ramping up of testing capacity. After new cases surged by the end of March and in April 2020, different types of stay-at-home orders and distancing policies with different durations were implemented across the United States [21]. In June and July 2020, reopening plans were halted and reversed by various jurisdictions to limit the resurgence of COVID-19 [22].

In Fig. 3(b), we show the corrected proportion of positive tests $\hat{f}_t$, found by numerically inverting Eq. 3.1 for different FPR, FNR, and bias combinations. We observe that a small $\mathrm{FPR} = 0.04$ shifts values $(\tilde{f}_{\rm b})_t \approx 0.05$ towards zero such that the corrected positive testing rate $\hat{f}_t \approx 0$. Reducing the FNR from 0.2 to 0.1 has only little effect on the corrected proportion of positive tests $\hat{f}_t$ [solid black and dashed lines in Fig. 3(b)]. Accounting for a positive testing bias of $b = 1$ (*i.e.*, preferential testing of infected and symptomatic individuals by a factor of e), however, markedly changes the inferred $\hat{f}_t$ [dashed-dotted black line in Fig. 3(b)].

4. Inference of bias b

One way to estimate the testing bias b is to identify a smaller subset of control tests within a jurisdiction that is believed to be unbiased and compare it with the reported fraction of positive tests obtained via standard (potentially biased) testing procedures.

We can derive a rather complete methodology to estimate bias by formally comparing the statistics of two sets of tests applied to the same population. The first set of control tests with testing parameters $\theta_0 = \{Q_0, f, \mathrm{FPR}_0, \mathrm{FNR}_0\}$ is known to be unbiased (has prior distribution $\delta(b)$), while the second set is taken with known parameters $\theta = \{Q, \mathrm{FPR}, \mathrm{FNR}\}$, but unknown testing bias b. For example, the control set may consist of a smaller number Q_0 of tests that are administered completely randomly, while the second set may be the scaled-up set of tests with $Q > Q_0$. Since both sets of tests are applied roughly at the same time to the same overall population, the underlying positive fraction f is assumed to be the same in both test sets. We can then use Bayes' rule on the first unbiased test set to infer f:

$$P_{\rm post}(f|\tilde{f}_0, \theta_0, b=0) = \frac{P_{\rm TOT}(\tilde{f}_0|f, \theta_0, b=0)P_0(f)}{\int_0^1 P_{\rm TOT}(\tilde{f}_0|f', \theta_0, b=0)P_0(f')\,{\rm d}f'}. \tag{4.1}$$

The probability distribution over b for a specified value of f can also be constructed from Bayes' rule

$$P_{\rm post}(b|\tilde{f}_{\rm b}, f, \theta) = \frac{P_{\rm TOT}(\tilde{f}_{\rm b}|f, \theta, b)P_0(b)}{\int_{-\infty}^{\infty} P_{\rm TOT}(\tilde{f}_{\rm b}|f, \theta, b')P_0(b')\,{\rm d}b'}, \tag{4.2}$$

where $P_0(\cdot)$ are prior distributions over the relevant parameters. The final distribution over the bias factor, given the two measurements $\tilde{f}_0$ and $\tilde{f}_{\rm b}$ derived from the two sets of tests with testing parameters θ_0 and θ, can be found using

$$P_{\rm b}(b|\tilde{f}_{\rm b}, \tilde{f}_0, \theta, \theta_0) = \int_0^1 P_{\rm post}(b|\tilde{f}_{\rm b}, f, \theta)P_{\rm post}(f|\tilde{f}_0, \theta_0, b=0){\rm d}f. \tag{4.3}$$

Of course, a simpler maximum likelihood estimate can also be applied to data by first inferring the most likely value of f from the control test set. We can use the number of positive tests in the control sample Q_0^+ to define the variable $\tilde{f}_0 = Q_0^+/Q_0$. One can then maximize $P_{\rm TOT}(\tilde{f}_0|f, \theta_0, b =$

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

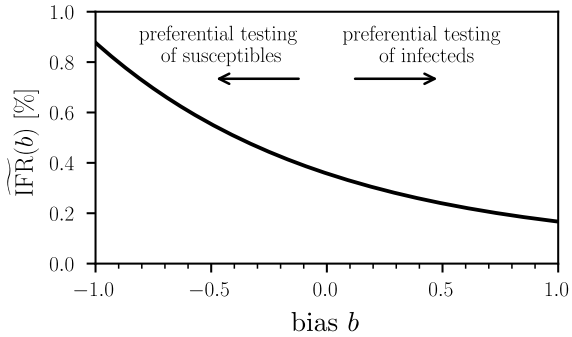

Figure 4. Dependence of the observed IFR on testing bias. The observed infection fatality ratio $\widetilde{\mathrm{IFR}}$ (Eq. (4.4)) as a function of testing bias b. We used the example of the German jurisdiction Gangelt and set $\hat{f} = 0.1553$, $D = 7$, and $N = 12{,}597$. A value $b > 0$ indicates a testing bias towards currently and/or previously infected individuals while susceptible and/or non-infectious individuals are preferentially tested for $b < 0$. Unbiased testing corresponds to $b = 0$.

0) with respect to f and use this value $\hat{f}$ in $P_{\mathrm{TOT}}(\tilde{f}_{\mathrm{b}}|\hat{f}, \theta, b)$. Maximizing $P_{\mathrm{TOT}}(\tilde{f}_{\mathrm{b}}|\hat{f}, \theta, b)$ with respect to b then gives the MLE estimate $\hat{b}$. We can use random and unbiased sampling results obtained in the German jurisdiction of Gangelt, North Rhine-Westphalia [18]. A total of 600 adult persons with different last names were randomly selected from a population of 12,597 and asked to participate in the study together with their household members. The resulting study comprised of $Q_0 = 919$ subjects who underwent serological and PCR testing between March 31-April 6, 2020. The specificity and sensitivity corrected, unbiased positive test fraction was determined to be $f = 15.53\%$ (95% CI; 12.31%–18.96%). Thus, we use this value as an estimate for the true underlying positivity rate $\hat{f}$. The larger sample taken across North Rhine-Westphalia between March 30–April 5, 2020 was measured ($Q \approx 25,000$) to be $\tilde{f}_{\mathrm{b}} \approx 0.1$ [23]. Assuming that this value is also error-corrected, an estimate of the bias $\hat{b}$ in this main testing set can be found by solving $\tilde{f}_{\mathrm{b}} \approx 0.1 = \bar{\mu}_{\mathrm{T}}(\hat{f} = 0.1553, \hat{b}, \mathrm{FPR} = \mathrm{FNR} = 0) = \bar{\mu}(\hat{f} = 0.1553, \hat{b})$ for $\hat{b}$. We find that the difference between the unbiased positive testing rate of 15.53% and 10% corresponds to a bias of $\hat{b} = -0.50$. This negative bias likely arises because Gangelt was a an infection hotspot within the entire North Rhine-Westphalia region, so the control sample was probably not unbiased. For comparison, a higher biased positive testing rate of 20% would lead to an estimated testing bias $\hat{b} = 0.31$.

The number of total deaths on April 6, 2020 amounted to 7. Hence, the corresponding estimate of the infection fatality ratio (IFR), the number of disease-induced deaths D_t divided by the total number of cases N_t at time t, in this jurisdiction on April 6, 2020 was $7/(0.1553 \times 12{,}597) = 0.36\%$ (95% CI; 0.29%–0.45%) [18]. If only a biased estimate of the proportion of positive cases is known and not the true value f, we can use our framework to distinguish between the true $\mathrm{IFR}_t = D_t/(N_t - S_t) = D_t/(f_0 N_t)$ and the observed infection fatality ratio

$$\widetilde{\mathrm{IFR}}_t = \frac{D_t}{f N_t} = \frac{D_t}{f \tilde{B}(f, b) N_t} \,. \tag{4.4}$$

Figure 4 shows the observed $\widetilde{\mathrm{IFR}}$ as a function of testing bias b for the aforementioned example of the German jurisdiction of Gangelt. Values of $b > 0$ correspond to preferential testing of infected individuals and thus lead to an apparently lower $\widetilde{\mathrm{IFR}}$. The opposite holds for $b < 0$.

5. Summary and Conclusions

Radiological testing methods such as chest computed tomography (CT) are used sporadically to identify COVID-19-induced pneumonia in patients with negative tests [24], However, the overwhelming majority of COVID-19 tests are based on serological (or antibody) tests and rapid antigen tests, ELISA, and RT-PCR assay [1]. These tests are designed to subsequently output

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

a binary signal, either infected or not. The population statistics of this output are affected by testing errors and bias. False positive and false negative rates of serological tests are generally smaller than those of rapid antigen tests and RT-PCR tests. However, serological tests are unable to identify early-stage infections since they are measuring antibody titres that usually develop a few days up to a few weeks after infection. In addition to the occurrence of false positives and false negatives (*i.e.*, type-I and type-II errors), certain demographic groups (*e.g.*, elderly people or those with comorbidities such as heart and lung diseases) may be overrepresented in testing statistics.

To quantify the impact of both type-I/II errors and testing bias on reported COVID-19 case and death numbers, we developed a mathematical framework that describes erroneous and biased sampling (both with and without replacement) from a population of susceptible, infected, and removed (*i.e.*, recovered or deceased) individuals. We identify a positive testing bias $b > 0$ with an overrepresentation of previously or currently infected individuals in the study population. Conversely, a negative testing bias $b < 0$ corresponds to an overrepresentation of susceptible and/or non-infectious individuals in the study population. We derived maximum likelihood estimates of the testing-error and testing-bias-corrected fraction of positive tests. Our methods can be also applied to infer the full distribution of corrected positive testing rates over time and for different types of tests across different jurisdictions.

The mathematical quantity that underlies most of our analysis is the proportion of apparent positive tests. As pointed out in [25], the *absolute* number of positive tests may not capture the actual growth of an epidemic due to limitations in testing capacity. Still, many jurisdictions report absolute case numbers without specifying the total number of tests or additional information about test type, date of test, and duplicate tests [26], rendering interpretation and application to epidemic surveillance challenging. For a reliable picture of COVID-19 case numbers, more complete testing data, including total number of tests, number of positive tests, test type, and date of test, has to be reported and made publicly available at online data repositories. To correct for false positive, false negatives, and testing bias in testing statistics (Fig. 3), it will be also important to further improve estimates of FPR, FNR, and b through in field studies. In particular, estimating the testing bias b requires random sampling studies similar to that carried out in [18]. Finally, while we have presented our analysis in the context of the COVID-19 pandemic, the general results presented in this paper apply to testing and estimation of severity of any infectious disease afflicting a population.

Data Accessibility. Our source codes are publicly available at `https://github.com/lubo93/disease-testing`.

Authors' Contributions. LB, MRD, and TC contributed equally to the study design, data analyses, and manuscript writing.

Competing Interests. The authors declare no competing interests.

Funding. LB acknowledges financial support from the Swiss National Fund (P2EZP2_191888). The authors also acknowledge financial support from the Army Research Office (W911NF-18-1-0345), the NIH (R01HL146552), and the National Science Foundation (DMS-1814364, DMS-1814090).

References

1. CDC, "Overview of Testing for SARS-CoV-2 (COVID-19)," https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html, 2020, accessed: 2021-01-26.
2. M. R. Tom and M. J. Mina, "To interpret the SARS-CoV-2 test, consider the cycle threshold value," pp. 2252–2254, 2020.
3. A. Mandavilli, "Your coronavirus test is positive. Maybe it shouldn't be," *The New York Times*, vol. 17, 2020.
4. P. N. Patrone and A. J. Kearsley, "Classification under uncertainty: Data analysis for diagnostic antibody testing," *arXiv:2012.10021*, p. arXiv:2012.10021, 2021.
5. U.S. Food and Drug Administration, "EUA authorized serology test performance," https://www.fda.gov/medical-devices/emergency-situations-medical-devices/

medRxiv preprint doi: https://doi.org/10.1101/2021.05.22.21257643; this version posted May 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

eua-authorized-serology-test-performance, 2020.
6. A. N. Cohen, B. Kessel, and M. G. Milgroom, "Diagnosing COVID-19 infection: the danger of over-reliance on positive test results," *medRxiv*, 2020.
7. J. Watson, P. F. Whiting, and J. E. Brush, "Interpreting a COVID-19 test result," *BMJ*, vol. 369, 2020.
8. M. L. Bastos, G. Tavaziva, S. K. Abidi, J. R. Campbell, L.-P. Haraoui, J. C. Johnston, Z. Lan, S. Law, E. MacLean, A. Trajman *et al.*, "Diagnostic accuracy of serological tests for COVID-19: systematic review and meta-analysis," *BMJ*, vol. 370, 2020.
9. R. Lassaunière, A. Frische, Z. B. Harboe, A. C. Nielsen, A. Fomsgaard, K. A. Krogfelt, and C. S. Jørgensen, "Evaluation of nine commercial SARS-CoV-2 immunoassays," *medRxiv*, 2020.
10. J. D. Whitman, J. Hiatt, C. T. Mowery, B. R. Shy, R. Yu, T. N. Yamamoto, U. Rathore, G. M. Goldgof, C. Whitty, J. M. Woo *et al.*, "Test performance evaluation of SARS-CoV-2 serological assays," *medRxiv*, 2020.
11. Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology*, p. 200432, 2020.
12. W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, and W. Tan, "Detection of SARS-CoV-2 in different types of clinical specimens," *JAMA*, vol. 323, no. 18, pp. 1843–1844, 2020.
13. I. Arevalo-Rodriguez, D. Buitrago-Garcia, D. Simancas-Racines, P. Zambrano-Achig, R. del Campo, A. Ciapponi, O. Sued, L. Martinez-Garcia, A. Rutjes, N. Low *et al.*, "False-negative results of initial RT-PCR assays for COVID-19: a systematic review," *medRxiv*, 2020.
14. L. Böttcher, M. D'Orsogna, and T. Chou, "Using excess deaths and testing statistics to improve estimates of COVID-19 mortalities," *European Journal of Epidemiology*, 2021.
15. L. Böttcher, M. Xia, and T. Chou, "Why case fatality ratios can be misleading: individual-and population-based mortality estimates and factors influencing them," *Physical Biology*, vol. 17, p. 065003, 2020.
16. CDC, "Summary of Laboratory Testing Results Reported to CDC," https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html, 2020, accessed: 2020-04-18.
17. N. E. Fenton, M. Neil, M. Osman, and S. McLachlan, "COVID-19 infection and death rates: the need to incorporate causal explanations for the data and avoid bias in testing," *Journal of Risk Research*, pp. 1–4, 2020.
18. H. Streeck, B. Schulte, B. M. Kümmerer, E. Richter, T. Höller, C. Fuhrmann, E. Bartok, R. Dolscheid-Pommerich, M. Berger, L. Wessendorf *et al.*, "Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.
19. S. Xu, L. Böttcher, and T. Chou, "Diversity in biology: definitions, quantification and models," *Physical Biology*, vol. 17, no. 3, p. 031001, 2020.
20. "The COVID Tracking Project," https://covidtracking.com/, accessed: 2021-02-20.
21. A. Moreland, C. Herlihy, M. A. Tynan, G. Sunshine, R. F. McCord, C. Hilton, J. Poovey, A. K. Werner, C. D. Jones, E. B. Fulmer *et al.*, "Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement – United States, March 1–May 31, 2020," *Morbidity and Mortality Weekly Report*, vol. 69, no. 35, p. 1198, 2020.
22. "A Timeline of COVID-19 Developments in 2020," https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020, 2021, accessed: 2021-04-17.
23. "Laborbasierte Surveillance von SARS-CoV-2–24.04.2020," https://ars.rki.de/Docs/SARS_CoV2/Archiv/Wochenberichte/20200424_Wochenbericht.pdf, 2020, accessed: 2021-04-12.
24. X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing," *Radiology*, vol. 296, no. 2, pp. E41–E45, 2020.
25. R. Omori, K. Mizumoto, and G. Chowell, "Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate," *International Journal of Infectious Diseases*, vol. 94, pp. 116–118, 2020.
26. J. G. McGinty, "Covid-19 Positivity Rate Might Not Mean What You Think It Does," https://www.wsj.com/articles/covid-19-positivity-rate-might-not-mean-what-you-think-it-does-11599211800, 2020, accessed: 2021-04-18.