Title

Adjusting Incidence Estimates with Laboratory Test Performances: A Pragmatic Maximum Likelihood Estimation-Based Approach.

Journal

Epidemiology, 35(3)

Authors

Weng, Yingjie
Tian, Lu
Boothroyd, Derek
et al.

Publication Date

2024-05-01

DOI

10.1097/EDE.0000000000001725

Peer reviewed

OPEN

Adjusting Incidence Estimates with Laboratory Test Performances

A Pragmatic Maximum Likelihood Estimation-Based Approach

Yingjie Weng,[a] Lu Tian,[b] Derek Boothroyd,[a] Justin Lee,[a] Kenny Zhang,[a] Di Lu,[a] Christina P. Lindan,[c,d] Jenna Bollyky,[e] Beatrice Huang[f], George W. Rutherford,[c,d] Yvonne Maldonado,[g] Manisha Desai,[a,b] and on behalf of the California Pandemic Consortium

Submitted December 1, 2022; accepted January 28, 2024

From the [a]Quantitative Sciences Unit, Department of Medicine, Stanford University, Palo Alto, CA; [b]Biomedical Data Science, Department of Medicine, Stanford University, Palo Alto, CA; [c]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; [d]Institute for Global Health Sciences, University of California San Francisco, San Francisco, CA; [e]Division of Primary Care & Population Health, School of Medicine, Stanford University, Stanford, CA; [f]Department of Family and Community Medicine, University of California, San Francisco, CA; and [g]Division of Pediatric Infectious Diseases, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA.

M.D., Y.W., and L.T. designed the study and the main conceptual ideas, supervised the findings of work, and performed the material preparation, data collection and analysis. Y.W. and L.T. developed the statistical analysis plan, performed the statistical analysis and validation. D.B., J.L., K.Z., D.L., and B.H. contributed to data collection, data management and data cleaning. C.L., J.B., G.R, Y.M. designed and oversaw the clinical rationale of the study. All authors contributed to the study conception and design. The first draft of the manuscript was written by Y.W., and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

This work was supported by the Chan Zuckerberg Initiative (CZIF2020-003870). This manuscript was partially supported by the following National Institutes of Health (NIH) grants: the Biostatistics Shared Resource (B-SR) of the NCI-sponsored Stanford Cancer Institute (P30CA1244354), Stanford's Center for Clinical and Translational Education and Research award, under the Biostatistics, Epidemiology and Research Design (BERD) Program (UL1TR003142).

The authors report no conflicts of interest.

All study materials are available for request upon approval of the corresponding author.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Manisha Desai, Quantitative Sciences Unit, Department of Medicine, Stanford University, 3180 Porter Dr, Palo Alto, CA 94304. Email: manishad@stanford.edu.

The TrackCOVID study was designated as a public health surveillance study and not human subjects research under 45 CFR 46.102(l) by Stanford University School of Medicine Administrative Panel on Human Subjects in Medical Research and UCSF Institutional Review Board.

Informed consent was obtained from all study participants.

ISSN: 1044-3983/24/353-295307

DOI: 10.1097/EDE.0000000000001725

Abstract: Understanding the incidence of disease is often crucial for public policy decision-making, as observed during the COVID-19 pandemic. Estimating incidence is challenging, however, when the definition of incidence relies on tests that imperfectly measure disease, as in the case when assays with variable performance are used to detect the SARS-CoV-2 virus. To our knowledge, there are no pragmatic methods to address the bias introduced by the performance of labs in testing for the virus. In the setting of a longitudinal study, we developed a maximum likelihood estimation-based approach to estimate laboratory performance-adjusted incidence using the expectation-maximization algorithm. We constructed confidence intervals (CIs) using both bootstrapped-based and large-sample interval estimator approaches. We evaluated our methods through extensive simulation and applied them to a real-world study (TrackCOVID), where the primary goal was to determine the incidence of and risk factors for SARS-CoV-2 infection in the San Francisco Bay Area from July 2020 to March 2021. Our simulations demonstrated that our method converged rapidly with accurate estimates under a variety of scenarios. Bootstrapped-based CIs were comparable to the large-sample estimator CIs with a reasonable number of incident cases, shown via a simulation scenario based on the real TrackCOVID study. In more extreme simulated scenarios, the coverage of large-sample interval estimation outperformed the bootstrapped-based approach. Results from the application to the TrackCOVID study suggested that assuming perfect laboratory test performance can lead to an inaccurate inference of the incidence. Our flexible, pragmatic method can be extended to a variety of disease and study settings.

Keywords: COVID-19; Incidence; Maximum Likelihood Estimation; Sensitivity; Specificity

(Epidemiology 2024;35: 295–307)

It is often of interest to estimate disease incidence. For this purpose, we typically rely on repeated measurements of participants' disease status over time in a longitudinal framework, where disease status may be determined through lab tests. During the COVID-19 pandemic, accurate estimates of the incidence of infection with SARS-CoV-2 have been critical for public health policy decision-making.[1–3] There are multiple challenges to achieving this goal that include a sub-population

that is asymptomatic and a dynamic landscape of disease variants and diagnostic tests. Specifically, several SARS-CoV-2 test kits (rt-nucleic acid amplification, antigen, and serological tests) were approved under Emergency Use Authorization for clinical use without rigorous validation.[4,5] The SARS-CoV-2 literature demonstrated how critical it was to adjust for the varying performance of these laboratory tests (sensitivity and specificity) to draw an accurate inference on prevalence and incidence.[6,7] Indeed, for cross-sectional studies designed to estimate prevalence, the "standard correction" method based on Bayes' rules[6,8] was often applied to account for the laboratory test performance.[9–13] Both frequentist[8,9,14] and Bayesian approaches[11–13,15] were proposed for this purpose. For example, Bajema et al.[9] and Havers et al.[10] proposed a two-stage nonparametric bootstrapping approach that resampled the false positive and negative cases to account for the uncertainty of the sensitivity and specificity estimates. Meireles et al.[11] used a Bayesian approach with uniform prior distributions of sensitivity and specificity, whereas Sahlu and Whittaker[12] considered an informative beta prior and Meyer et al.[13] proposed more specific beta prior distributions that incorporated information by geographical regions. Under scenarios where knowledge of test performance was limited, Burstyn et al.[16] proposed a Bayesian approach to inform the sensitivity and specificity in the population using publicly available time-series data.

Because incidence measures the instantaneous probability of being infected at a given time point, it is considered a more relevant measure for informing timely decisions on public health and healthcare resource allocation to arrest SARS-CoV-2 transmission, as it provides additional insight into the current "momentum" of the epidemic.[17–20] In contrast to prevalence, incidence is often estimated with a longitudinal study design, which presents challenges in accounting for test performance.[17] Given the longitudinal nature of the study required to estimate incidence, participants need to be repeatedly tested over time, yielding a high volume of tests performed. This means that even a small percentage of laboratory testing errors can result in a high absolute number of false positive or negative test results, eventually leading to a biased estimate of incidence, while the direction and magnitude of the bias would depend on the underlying true incidence and prevalence of the disease. For the purposes of our study, we defined incidence as the initial infection for a participant, where participants were censored at a positive result with no further follow-up tests scheduled. The implications of censoring as a function of the test status are that the test performance of one assay will affect whether the patient will be censored at subsequent visits, affecting the participant's length of time considered at risk.

In the current literature, there are no guidelines or a "gold standard" approach to adjust for laboratory test errors when estimating incidence within a repeated measures longitudinal framework, which is a fairly common goal. For example, work by Becker and Britton (1999)[17] and by Gan and Bain (1998)[21] examined the estimation of the incidence risk under a similar longitudinal design. A key difference with the current article is our consideration of the testing error in identifying the true cases, quantified by imperfect sensitivity and specificity. Previous studies proposed methods to incorporate sensitivity and specificity by adjusting the number of positive or negative events.[22,23] However, in those studies, the incidence was modeled as a single probability with only one follow-up measurement. Further, the methods proposed assumed consistent follow-up times and nondifferential time at risk and loss to follow-up across participants. Novel methods that challenged or relaxed these assumptions were proposed in the field of HIV. For example, McDougal et al.[24] derived a correction factor for incidence adjustment—the probability of being infected divided by the probability of being in the window (P(T0)/P(W))—that was subsequently criticized because the correction factor took on the value of 1 under commonly occurring circumstances that did not make empirical sense.[24,25] Hargrove et al.,[26] therefore, proposed a modified version of a correction factor (ε)—the probability of being in the window period if infected at least twice the duration of the time interval (2μ) earlier.[26] Similar to McDougal's method, however, this approach could be misleading due to its underlying mathematical inconsistency that nonzero ε values could lead to anomalous results.[25] The methods mentioned above all failed to adequately leverage the repeated nature of testing by assuming a uniform distribution of testing across all time points. Thus, there is a critical gap in the current literature for addressing unbiased estimation of incidence in a typical longitudinal framework. Flexible methods are needed that adjust for laboratory test performance in the common scenario where incidence is estimated from repeated measurements.

Parent Study: The TrackCOVID Study

Our work is motivated by a public health surveillance initiative to estimate the incidence and prevalence of SARS-CoV-2 infection and the associated risk factors in the San Francisco Bay Area (TrackCOVID study). The TrackCOVID study was conducted from July 2020 to March 2021, during the outbreak of the COVID-19 pandemic. The longitudinal study relied on a sampling framework to randomly select residents from six Bay Area counties using census tract data.[27] We requested participants in the study to come into the clinic for a baseline visit and, subsequently, for monthly visits for up to 6 months. Evidence of infection was defined as a positive test using two types of assays measured repeatedly: (1) an rt-PCR test of a nasopharyngeal swab and (2) serologic testing of blood. In this way, we can capture infections that may have been missed by one of the assays, and particularly those infections we may miss in between monthly visits. Thus, our definition of infection relied on the performances of multiple assays. Importantly, to account for (1) the sampling framework, (2) the probability of selection from the household, and (3) nonresponse bias, we estimated and incorporated a combined weight variable for each participant when calculating quantities of interest (such as incidence).[27]

The TrackCOVID study was designated as a public health surveillance study and not human subjects research under 45 CFR 46.102(l) by the Stanford University School of Medicine Administrative Panel on Human Subjects in Medical Research and the University of California, San Francisco Institutional Review Board.

In this article, we proposed a new statistical framework to address the problem of estimating incidence while adjusting for laboratory test performance in a longitudinal repeated measures framework using a maximum likelihood estimation (MLE)-based approach. Our method can be extended to a variety of study design scenarios under relaxed assumptions. We presented the likelihood function using the motivating example, assessed the properties of the method with a simulation study, and applied our approach to the real-world data collected in our parent study.

METHODS

To better understand the underlying principles of our method, we illustrated ideas through the TrackCOVID study.

Possible Trajectories of Observed Longitudinal Test Result

In the TrackCOVID study, each participant could have at most seven visits (including the baseline visit). Considering loss to follow-up and censoring after a given visit, we may observe a total of 14 possible test result trajectories as illustrated in Table 1. For example, scenario 4 reflects the situation where an individual has three negative tests followed by a positive test at visit 4, and scenario 12 represents a trajectory where there are four consecutive negative tests immediately following enrollment with loss to follow-up afterward. In practice, observed data could be represented by either an unweighted or weighted number of participants, whose test results fall into each of these possible trajectories or scenarios, denoted as $n_k, \quad k = 1, \cdots, 14$. If there is any intermittent missing value, we consider the observed data to be those preceding this value.

Possible Trajectories of True Longitudinal Infection Status

There are eight true infection status trajectories for the participants enrolled in our parent study (Table 2). Here, we assume that once a participant's underlying true disease status is positive, their status remains positive until the last follow-up visit. Participant's test status at each visit is a binary variable (positive or negative) that can be determined by either a single test result or multiple test results combined under prespecified rules. First, we define the following parameters:

1. π: prevalence at baseline (probability of being infected at baseline).
2. *p:* incidence at a visit assumed to be constant, (i.e., the conditional probability of infection at the current visit given no infection in the prior visit, $p=P$(infected at the current visit|uninfected in the prior visit)).
3. *Sen*: Sensitivity (conditional probability of testing positive given infected, or P(Test +|True +)), considered fixed and as prespecified by the literature in both the parent and simulation study.
4. *Spe*: Specificity (conditional probability of testing negative given uninfected, or P(Test −|True −)), considered fixed and as prespecified by the literature in both the parent and simulation study.
5. *r*: the rate of loss to follow-up at a visit (conditional probability that a participant's data are not available at the current visit given the data were observed in the prior visit).

TABLE 1. All Possible Trajectories of Observed Longitudinal Test Result at the Participant Level[a]

	Baseline	Follow-up Visit 1	Follow-up Visit 2	Follow-up Visit 3	Follow-up Visit 4	Follow-up Visit 5	Follow-up Visit 6
Test result scenario 1	+						
Test result scenario 2	−	+					
Test result scenario 3	−	−	+				
Test result scenario 4	−	−	−	+			
Test result scenario 5	−	−	−	−	+		
Test result scenario 6	−	−	−	−	−	+	
Test result scenario 7	−	−	−	−	−	—	+
Test result scenario 8	−	−	−	−	−	−	−
Test result scenario 9	−						
Test result scenario 10	−	−					
Test result scenario 11	−	−	−				
Test result scenario 12	−	−	−	−			
Test result scenario 13	−	−	−	−	−		
Test result scenario 14	−	−	−	−	−	−	

[a]+ indicates a positive test result; − indicates a negative test result; and blank indicates loss to follow-up or censoring.

We can express the probabilities of the eight true infection status trajectories—denoted as q_i *(i=1, 2, 3,..., 8)*—as a function of π and p (Table 2). Note that the incidence is assumed to be constant for simplicity. Later, we extend the method to allow nonconstant incidence risk.

Likelihood Function at the Participant Level

The values of sensitivity (*sen*) and specificity (*spe*) of the laboratory test are considered to be given, and the observed data are $\{n_1, \cdots, n_{14}\}$. Given this, our interest lies in estimating (π, p). We account for test performance via the maximum likelihood estimation, which requires expressing the likelihood function. According to Tables 1 and 2, we expect to have a total of $14 \times 8 = 112$ combinations of the test results and the true infection status trajectories for the study cohort. We express the likelihood function for each combination as a function of π, p, and r (Tables 3 and 4).

We first estimate the rate of loss to follow-up at a given visit, r, by considering the likelihood function below.

$$(1-r)^{(n_2+2n_3+3n_4+4n_5+5n_6+6n_7+6n_8)+(n_{10}+2n_{11}+3n_{12}+4n_{13}+5n_{14})} \times r^{n_9+n_{10}+n_{11}+n_{12}+n_{13}+n_{14}}. \quad (1)$$

The maximum likelihood estimator (MLE) of r is

$$\hat{r} = \frac{n_9 + n_{10} + n_{11} + n_{12} + n_{13} + n_{14}}{(n_2 + 2n_3 + 3n_4 + 4n_5 + 5n_6 + 6n_7 + 6n_8) + (n_{10} + 2n_{11} + 3n_{12} + 4n_{13} + 5n_{14})}, \quad (2)$$

where n_k is defined as the number of participants with test result scenario k.

We define $s_{ik}(\pi, p, \hat{r})$ to be the k^{th} test result scenario and i^{th} true infection status in Tables 3 and 4. Given the MLE of r ($\hat{r}$), the only model parameters left to estimate are the ones

TABLE 2. All Possible Trajectories of True Longitudinal Infection Status at the Participant Level[a]

	Baseline	Follow-up Visit 1	Follow-up Visit 2	Follow-up Visit 3	Follow-up Visit 4	Follow-up Visit 5	Follow-up Visit 6	Probability
True status 1	+	+	+	+	+	+	+	$q_1 = \pi$
True status 2	−	+	+	+	+	+	+	$q_2 = (1-\pi)\,p$
True status 3	−	−	+	+	+	+	+	$q_3 = (1-\pi)(1-p)\,p$
True status 4	−	−	−	+	+	+	+	$q_4 = (1-\pi)(1-p)^2 p$
True status 5	−	−	−	−	+	+	+	$q_5 = (1-\pi)(1-p)^3 p$
True status 6	−	−	−	−	−	+	+	$q_6 = (1-\pi)(1-p)^4 p$
True status 7	−	−	−	−	−	−	+	$q_7 = (1-\pi)(1-p)^5 p$
True status 8	−	−	−	−	−	−	−	$q_8 = (1-\pi)(1-p)^6$

[a]+ indicates a positive infection; − indicates no infection.

TABLE 3. Probability of Combination of Each Test Result and True Infection Status (True Status 1–4)

	True Status 1	True Status 2	True Status 3	True Status 4	Probability
Test result scenario 1	sen	$(1-spe)$	$(1-spe)$	$(1-spe)$	q_j
Test result scenario 2	$(1-sen)\,sen$	$spesen$	$spe\,(1-spe)$	$spe\,(1-spe)$	$(1-\hat{r})\,q_j$
Test result scenario 3	$(1-sen)^2 sen$	$spe\,(1-sen)\,sen$	$spe^2 sen$	$spe^2\,(1-spe)$	$(1-\hat{r})^2 q_j$
Test result scenario 4	$(1-sen)^3 sen$	$spe(1-sen)^2 sen$	$spe^2\,(1-sen)\,sen$	$spe^3 sen$	$(1-\hat{r})^3 q_j.$
Test result scenario 5	$(1-sen)^4 sen$	$spe(1-sen)^3 sen$	$spe^2(1-sen)^2 sen$	$spe^3\,(1-sen)\,sen$	$(1-\hat{r})^4 q_j$
Test result scenario 6	$(1-sen)^5 sen$	$spe(1-sen)^4 sen$	$spe^2(1-sen)^3 sen$	$spe^3(1-sen)^2 sen$	$(1-\hat{r})^5 q_j$
Test result scenario 7	$(1-sen)^6 sen$	$spe(1-sen)^5 sen$	$spe^2(1-sen)^4 sen$	$spe^3(1-sen)^3 sen$	$(1-\hat{r})^6 q_j$
Test result scenario 8	$(1-sen)^7$	$spe(1-sen)^6$	$spe^2(1-sen)^5$	$spe^3(1-sen)^4$	$(1-\hat{r})^6 q_j$
Test result scenario 9	$(1-sen)$	spe	spe	spe	$\hat{r} q_j$
Test result enario 10	$(1-sen)^2$	$spe\,(1-sen)$	spe^2	spe^2	$(1-\hat{r})\,r q_j$
Test result scenario 11	$(1-sen)^3$	$spe(1-sen)^2$	$spe^2\,(1-sen)$	spe^3	$(1-\hat{r})^2 r q_j$
Test result scenario 12	$(1-sen)^4$	$spe(1-sen)^3$	$spe^2(1-sen)^2$	$spe^3\,(1-senm)$	$(1-\hat{r})^3 r q_j$
Test result scenario 13	$(1-sen)^5$	$spe(1-sen)^4$	$spe^2(1-sen)^3$	$spe^3(1-sen)^2$	$(1-\hat{r})^4 r q_j$
Test result scenario 14	$(1-sen)^6$	$spe(1-sen)^5$	$spe^2(1-sen)^4$	$spe^3(1-sen)^3$	$(1-\hat{r})^5 r q_j$

Sen indicates sensitivity; *spe,* specificity.

TABLE 4. Probability of Combination of Each Test Result and True Infection Status (True Status 5–8)

	True Status 5	True Status 6	True Status 7	True Status 8	Probability
Test result scenario 1	$(1-spe)$	$(1-spe)$	$(1-spe)$	$(1-spe)$	q_j
Test result scenario 2	$spe\,(1-spe)$	$spe\,(1-spe)$	$spe\,(1-spe)$	$spe\,(1-spe)$	$(1-\hat{r})\,q_j$
Test result scenario 3	$spe^2\,(1-spe)$	$spe^2\,(1-spe)$	$spe^2\,(1-spe)$	$spe^2\,(1-spe)$	$(1-\hat{r})^2 q_j$
Test result scenario 4	$spe^3\,(1-spe)$	$spe^3\,(1-spe)$	$spe^3\,(1-spe)$	$spe^3\,(1-spe)$	$(1-\hat{r})^3 q_j$
Test result scenario 5	$spe^4 sen$	$spe^4\,(1-spe)$	$spe^4\,(1-spe)$	$spe^4\,(1-spe)$	$(1-\hat{r})^4 q_j$
Test result scenario 6	$spe^4\,(1-sen)\,sen$	$spe^5 sen$	$spe^5\,(1-spe)$	$spe^5\,(1-spe)$	$(1-\hat{r})^5 q_j$
Test result scenario 7	$spe^4(1-sen)^2 sen$	$spe^5\,(1-sen)\,sen$	$spe^6 sen$	$spe^6\,(1-spe)$	$(1-\hat{r})^6 q_j$
Test result scenario 8	$spe^4(1-sen)^3$	$spe^5(1-sen)^2$	$spe^6\,(1-sen)$	spe^7	$(1-\hat{r})^6 q_j$
Test result scenario 9	spe	spe	spe	spe	$\hat{r}q_j$
Test result scenario 10	spe^2	spe^2	spe^2	spe^2	$(1-\hat{r})\,rq_j$
Test result scenario 11	spe^3	spe^3	spe^3	spe^3	$(1-\hat{r})^2 rq_j$
Test result scenario 12	spe^4	spe^4	spe^4	spe^4	$(1-\hat{r})^3 rq_j$
Test result scenario 13	$spe^4\,(1-spe)$	spe^5	spe^5	spe^5	$(1-\hat{r})^4 rq_j$
Test result scenario 14	$spe^4(1-sen)^2$	$spe^5\,(1-sen)$	spe^6	spe^6	$(1-\hat{r})^5 rq_j$

Sen indicates sensitivity; *spe,* specificity.

of interest: prevalence (π) and incidence (p). The likelihood function for participants who fall into test result scenario k can be obtained by summing the probabilities of the combinations of true statuses (1–8). Specifically, the likelihood of observing test result k is expressed as

$$P_k\,(\pi,\ p) = \sum_{i=1}^{8} s_{ik}\,(\pi,\ p,\ \hat{r}),\quad k = 1,\cdots,\ 14. \tag{3.1}$$

Under a more general study design whereas there is a total of K test result scenarios and I true infection statuses, the likelihood of observing test result k can be generalized as

$$P_k\,(\pi,\ p) = \sum_{i=1}^{I} s_{ik}\,(\pi,\ p,\ \hat{r}),\quad k = 1,\cdots,\ K,. \tag{3.2}$$

where $s_{ik}\,(\pi,\ p,\ \hat{r})$ needs to be modified accordingly.

Likelihood Function for the Study Cohort

The likelihood function of the entire study cohort follows a multinomial distribution corresponding to 14 test result scenarios:

$$l\,(p,\ \pi) = \prod_{k=1}^{14} P_k(\pi,\ p,\ \hat{r})^{n_k}. \tag{4.1}$$

For the general case, the likelihood function becomes

$$l\,(p,\ \pi) = \prod_{k=1}^{K} P_k(\pi,\ p,\ \hat{r})^{n_k}. \tag{4.2}$$

Expectation-maximization Steps

The above likelihood function $l\,(p,\ \pi)$ can be conveniently maximized using the expectation-maximization algorithm.

Specifically, let ξ_{ik} be the number of individuals who has test result scenario $k \in \{1,\cdots,\ 14\}$ with true infection status $i \in \{1,\cdots,8\}$. Then it follows that $n_k = \sum_{i=1}^{8} \xi_{ik}$

- M step: If ξ_{ik} are observed, then let $m_i = \sum_{k=1}^{14} \xi_{ik}$. We can find the MLE of $p,\ \pi$ by considering the likelihood function

$$l_1\,(\pi) \times l_2\,(p), \tag{5}$$

where

$$l_1\,(\pi) = \pi^{m_1}(1-\pi)^{m_2+m_3+m_4+m_5+m_6+m_7+m_8}$$

$$l_2\,(p) = p^{m_2+m_3+m_4+m_5+m_6+m_7}(1-p)^{m_3+2m_4+3m_5+4m_6+5m_7+6m_8}.$$

Clearly, the likelihood function is maximized at

$$\hat{\pi} = \frac{m_1}{m},\ \hat{p} = \frac{m_2+m_3+m_4+m_5+m_6+m_7}{m_2+2m_3+3m_4+4m_5+5m_6+6m_7+6m_8}, \tag{6}$$

where $m = \sum_{i=1}^{8} m_i = \sum_{k=1}^{14} n_k$.

- E step: Given current estimate of p and π, we can find $E(\xi_{ik}|data,\ \ p,\ \ \pi)$ as

$$E\,(\xi_{ik}|data,\ p,\ \pi) = n_k \frac{s_{ik}\,(p,\ \pi,\ \hat{r})}{\sum_{i=1}^{8} s_{ik}\,(p,\ \pi,\hat{r})}. \tag{7}$$

In summary, the expectation-maximization algorithm can be implemented iteratively via the following steps:

1. Initialize $p,\ \pi$
2. Calculate $\hat{\xi}_{ik} = E\,(\xi_{ik}|data,\ p,\ \pi)$
3. Calculate $\hat{m}_i = \sum_{k=1}^{14} \hat{\xi}_{ik}$

4. Calculate $\hat{\pi} = \frac{\hat{m}_1}{m}$, $\hat{p} = \frac{\widehat{m_2}+\widehat{m_3}+\widehat{m_4}+\widehat{m_5}+\widehat{m_6}+\widehat{m_7}}{\widehat{m_2}+2\widehat{m_3}+3\widehat{m_4}+4\widehat{m_5}+5\widehat{m_6}+6\widehat{m_7}+6\widehat{m_8}}$
5. Repeat 2–4 until convergence.
6. Denote the final estimator by $(\hat{\pi}, \hat{p})$ with slightly abuse of notations.

The details of the expectation-maximization algorithm are provided in eAppendix 1; http://links.lww.com/EDE/C117.

Confidence Intervals

Nonparametric Bootstrapped-based Approach

The confidence intervals (CIs) of p, π can be constructed using the nonparametric bootstrapped-based method. Specifically, we create 500 bootstrap cohorts with the same sample size as the original study cohort via the standard nonparametric bootstrap method so that each bootstrap cohort would randomly select participants with the replacement from the original study cohort.[28] We then compute the MLE of (p, π) for each of the 500 bootstrap cohorts. The distribution of $(\hat{p}, \hat{\pi})$ is then approximated by the empirical distribution from the 500 bootstrapped-based estimates. Lastly, we calculate the 2.5 and 97.5 percentiles of the 500 bootstrapped-based estimates as the lower and upper limits for the 95% empirical CIs of p and π.

Large Sample Interval Estimator Approach

As an alternative to the bootstrapped-based approach, we also estimate the variance of $\hat{p}$ and $\hat{\pi}$ based on the likelihood function directly. To this end, we first calculate the 2nd-order numerical derivatives of the log-likelihood function at $(\hat{p}, \hat{\pi})$ to construct the Fisher's information matrix.[29,30] The diagonal elements of the inverse of the matrix are used to estimate the variances of $\hat{p}$ and $\hat{\pi}$. The 95% CIs are then computed via equation (8) and equation (9) below, leveraging the large-sample approximation to the distribution of MLE. The CI is constructed with logit transformation to ensure that the CI is always within 0 and 1. More details of the calculations are provided in eAppendix 2; http://links.lww.com/EDE/C117.

$$\text{expit}\left\{\text{logit}(\hat{p}) \pm 1.96 \times \frac{\sqrt{\text{var}(\hat{p})}}{\hat{p}(1-\hat{p})}\right\} \quad (8)$$

$$\text{expit}\left\{\text{logit}(\hat{\pi}) \pm 1.96 \times \frac{\sqrt{\text{var}(\hat{\pi})}}{\hat{\pi}(1-\hat{\pi})}\right\} \quad (9)$$

Important Considerations of Underlying Statistical Assumptions

The validity of our proposed method based on our motivating study relies on the following assumptions.

1. *Constant incidence*: We assume the same probability (P) of having a new infection at each follow-up visit for all participants, which aligns with the original design in TrackCOVID. In the presence of nonconstant incidence over time, our estimate of the incidence parameter (p) is essentially an "average" of incidences across all study visits.
2. Discrete and finite number of follow-up visits with a fixed length of time between visits (e.g., monthly or weekly visits). This assumption can be relaxed, as we described in the next section.
3. Laboratory test performance or properties of the diagnostic tests (sensitivity and specificity) are constant across visits and prespecified based on the literature or previous studies. We did not consider laboratory test performance metrics as parameters to be estimated either in the simulation study or in the parent study. However, with appropriate modifications of our proposed algorithm, the uncertainty in the reported laboratory performance metrics can be considered.
4. Constant rate of loss to follow-up over time: we assumed the same probability (r) of loss to follow-up at each follow-up visit. In the presence of nonconstant rates of loss to follow-up over study visits, we may consider estimating the parameters of loss to follow-up as a function of study visits or calendar time.

Potential Extensions of the Method

There are many possible extensions of the method. Below we describe three key extensions.

Incorporate Sampling Weights

Oftentimes, when over- or undersampling of some patient subgroups is implemented, weighted inference is needed. It is straightforward to incorporate weights in our expectation-maximization algorithm. Specifically, instead of using n_k as our input data, we can use the weighted counts of a number of participants under the k^{th} test result scenario. We can then update n_k as $\sum_{i=1}^{n_k} w_{ik}$, where w_{ik} represents the weighted number of participants in the k^{th} test result scenario.

Incorporate Factors Associated with Incidence

Knowing potential factors that are associated with incidence, we can update our model and the corresponding likelihood function on the participant level to relax the constant incidence assumption. For example, in the case of SARS-CoV-2 which has a rapidly changing landscape over time, it may be reasonable to expect that incidence would depend on calendar months. One way to relax this assumption is to model the incidence rate at a given visit as a function of calendar time. For example, we may assume that

$$p(s;\theta) = \frac{\exp(\alpha_0 + \beta_0 s)}{1 + \exp(\alpha_0 + \beta_0 s)},$$

where s is the calendar month of the visit, and (α_0, β_0) are unknown model parameters to be estimated. Under this model, we may test if the incidence rate is constant over time by examining the point and interval estimates of β_0. This model can be

estimated via the expectation-maximization algorithm as well. Specifically, we may organize the observed data into multiple waves $\{n_{1,s}, \cdots, n_{14,s}\}$, $s = 0, \cdots, S$, where $n_{j,s}$ represents the observed frequency of the j^{th} scenario among patients who were enrolled at calendar time s, that is, the s^{th} wave. Let ξ_{iks} be the number of individuals who had test result scenario $k \in \{1, \cdots, 14\}$ with true infection status $i \in \{1, \cdots, 8\}$ enrolled at the s^{th} wave. Then it follows that $n_{ks} = \sum_{i=1}^{8} \xi_{iks}$. The new expectation-maximization algorithm consists of the following two steps:

1. Maximization step: If ξ_{iks} is observed, then let

$$m_{is} = \sum_{k=1}^{14} \xi_{iks}, \quad m_s = \sum_{i=1}^{8} m_{is}, \quad m_i = \sum_{s=0}^{S} m_{is}, \quad m = \sum_{i=1}^{8} m_i.$$

Then the number of new incidences at calendar month t becomes

$$\gamma_t^+ = \sum_{s=0}^{\min(t-1,\ S)} m_{(t+1-s)s}$$

while the number of patients at calendar month t is

$$\gamma_t = \sum_{s=0}^{\min(t-1,\ S)} \sum_{i=t+1-s}^{8} m_{is}$$

We can find the MLE of $(\alpha, \beta,\ \pi)\,\pi$ by considering the likelihood function

$$l_1(\pi) \times l_2(\alpha,\ \beta),$$

where

$$l_2(\alpha, \beta) = \prod_{t \geq 1} \frac{\exp\left\{(\alpha + \beta t)\,\gamma_t^+\right\}}{\left\{1 + \exp(\alpha + \beta t)\right\}^{\gamma_t}}.$$

To be noted, maximizing the likelihood function $l_2(\alpha, \beta)$ is the same task as fitting a logistic regression with observation $\{(t,\ 1)\}$ weighted by γ_t^+, and $\{(t,\ 0)\}$ weighted by $\gamma_t - \gamma_t^+$, for $t \geq 1$.

2. Expectation step: Given the current estimate of (α, β) and π, we can find $E(\xi_{iks}|data,\ \alpha,\ \beta,\ \pi)$ as

$$E(\xi_{iks}|data,\ \alpha, \beta,\ \pi) = n_{ks} \frac{q_{iks}(\alpha, \beta,\ \pi,\ \hat{r})}{\sum_{k=1}^{14} q_{iks}(\alpha, \beta,\ \pi, \hat{r})},$$

where $q_{iks}(a, \beta,\ \pi, r)$ is the probability of being at the i^{th} true status trajectory, k^{th} observed test trajectory for participants enrolled at time s. Specifically, $q_{iks}(\alpha, \beta, \pi,\ r)$ can be calculated by replacing q_j in Table 2 by

$$q_1 = \pi,$$

$$q_2 = (1 - \pi) \frac{\exp(\alpha + \beta(s+1))}{1 + \exp(\alpha + \beta(s+1))},$$

$$q_j = (1 - \pi) \times \prod_{l=1}^{\max(j-2)} \frac{1}{1 + \exp(\alpha + \beta(s+l))} \times \frac{\exp(\alpha + \beta(s+j-1))}{1 + \exp(\alpha + \beta(s+j-1))},\ 3 \leq j \leq 7,$$

and

$$q_8 = (1 - \pi) \times \prod_{l=1}^{6} \frac{1}{1 + \exp(\alpha + \beta(s+l))}.$$

This expectation-maximization algorithm is used to find the MLEs of $(\alpha,\ \beta)$. The variance of the MLE is estimated using the nonparametric bootstrap method. In principle, we can account for more complex factors associated with incidence using similar extension. When the number of incidences is small, for example, as in the TrackCOVID study, the expectation-maximization algorithm may generate outliers for some bootstrapped datasets, which may exaggerate the variance estimate for $\left(\widehat{\alpha,}\ \widehat{\beta}\right)$. To overcome this, instead of using the empirical standard error of the bootstrapped estimates, we use appropriately normalized median absolute deviations as the estimates for the standard error of the corresponding parameters. The reported 95% CIs for the TrackCOVID study were constructed based on this robust standard error estimate.

Allow Differential Cadence and Number of Follow-ups

To extend our method to allow varying lengths of time between visits, we can simply update the probability of each true infection trajectory by attributing a differential probability of infection based on time since the last visit. For example, suppose we were to re-design our study with the first three follow-up visits at a biweekly cadence and the last three visits at a monthly cadence. By denoting the incidence of the first three follow-up visits as p, we could have the incidence of the last three visits equal to $1 - (1-p)^2 \approx 2p$ under the constant incidence assumption. As a result, we can update the probability of the true infection trajectories accordingly (eTable 1; http://links.lww.com/EDE/C117).

Theoretically, it is possible to extend our methods so that they allow for varying cadence and frequency of the follow-up visits across participants. This could be implemented by writing out the likelihood function for each individual participant separately, with the infection rate being $p = \exp(\lambda L_t)$, where L_t is the time from the previous visit to the current visit and λ is the infection rate, which can be further modeled as, for example, a function of calendar time as above.

Simulation Study Design

We conducted an extensive simulation study to investigate the finite sample performance of the proposed method based on our TrackCOVID study design. We considered different combinations of the five key parameters to mimic relevant contexts: true prevalence (π), true incidence (p), probability of loss to follow-up (r), sensitivity (sen), and specificity (spe).

For each simulation scenario, we simulated 500 datasets with a sample size of 4000. For each simulated dataset, we used the proposed expectation-maximization algorithm to find the MLEs for $(\pi,\ p)$. In the expectation-maximization algorithm, we set the initial values of π and p as the observed prevalence and incidence, respectively, without considering the sensitivity and specificity, and the maximum number of iterations was 1000. The convergence was reached if consecutive changes in log-likelihood were less than 10^{-6}. For each simulated dataset, we recorded whether the algorithm converged within 1000 iterations and the number of iterations at the convergence. We also reported median bias and median percent bias and the empirical coverage level of constructed 95% CIs of p. Specifically, we calculated the median bias as the median of the differences between the estimated incidence $\hat{p}$ and true incidence p; the median percent bias was calculated as the ratio of the median bias to p; and the empirical coverage level was the proportion of constructed CIs containing the true incidence p.

Scenario for TrackCOVID Study

The following values of our key parameters were selected as follows to closely mimic the motivating TrackCOVID study:

$$\pi = 0.015,\ p = 0.01,\ r = 0.3,\ sen = 0.728,\ spe = 0.997$$

Scenarios Under Varying Assumptions of $\pi,\ p,\ r$

To investigate the impact of π, p, r upon the performance of proposed methods, we considered additional plausible scenarios with varying assumptions of π, p, r under fixed $sen = 0.728$, $spe = 0.997$. All pairwise combinations of π, p, r below were considered in the simulations, resulting in $3 \times 3 \times 2 = 18$ simulation scenarios. For each of the 18 simulation scenarios, we simulated 500 datasets with a sample size of 4000.

$$\pi = 0.015,\ 0.02,\ 0.03$$

$$p = 0.01,\ 0.1,\ 0.5$$

$$r = 0.01,\ 0.3$$

Scenarios Under Varying Assumptions of $sen,\ spe$

To further explore the impact of *sen*, *spe* upon our algorithm performance, we performed simulation scenarios with varying *sen*, *spe* at fixed assumptions of $\pi,\ p,\ r\ (\pi = 0.015,\ p = 0.01,\ r = 0.3)$. The following possible *sen*, *spe* were considered.

$$sen = 0.7,\ 0.8,\ 0.9$$

$$spe = 0.99,\ 0.992,\ 0.994,\ 0.996,\ 0.998$$

All pairwise combinations of *sen*, *spe* were considered in this set of simulations, resulting in $3 \times 5 = 15$ simulation scenarios.

Scenarios Under Varying Incidence Rates over Calendar Time

To investigate the impact of varying incidence, we considered at month s, incidence would follow the function below.

$$p\,(s;\theta) = \frac{\exp\,(\alpha_0 + \beta_0 s)}{1 + \exp\,(\alpha_0 + \beta_0 s)}.$$

Specifically, we assumed that

$$\pi = 0.015,\ \alpha = -5,\ \beta = 0.2,\ r = 0.3,\ sen = 0.728,\ spe = 0.997$$

and patients were enrolled at $s = 0,\ 1,\ 2,$ and 3, that is, there were four waves of patients with the sample size being 500, 1000, 1000, and 1500, respectively, mimicking the parent study.

Software

We coded our method using the R programming language (The R Project for Statistical Computing, Vienna, Austria) and created an open source package ("mlelabperform") in a github repository (install_github("isabelweng/mlelabperform")). Scripts of the simulations were made available under the same repository.

RESULTS

Results from Simulation Study

Scenario Corresponding to TrackCOVID Study

($\pi = 0.015,\ p = 0.01,\ r = 0.3,\ sen = 0.728,\ spe = 0.997$) The median bias for estimating incidence (p) was low: −0.000124 (95% CI = −0.0029, 0.0031) or −0.15 per 100 person-year (95% CI = −3.5, 3.7) (to be noted, $p = 0.01$, or 12 per 100 person year). The median percent bias for estimating incidence was −1.24% (95% CI = −29%, 31%). The empirical coverage level of large-sample interval estimated 95% CIs was 98.6% and higher than the coverage level obtained with bootstrapped-based 95% CIs (92.2%). We were able to estimate π and p for all 500 simulated datasets under the TrackCOVID scenario, where the median number of iterations needed to reach convergence was 81, (interquartile range: 71, 92).

Scenarios Under Varying Assumptions of $\pi,\ p,\ r$

The expectation-maximization algorithm successfully converged within 1000 iterations for all the simulated datasets across 18 simulated scenarios with close to zero median bias (-2.02×10^{-4} to 1.00×10^{-4}, or −0.24 to 0.12 per 100 person-year) and percent median bias (−1.05% to 5.00%) for p (Figure 1A,B). It took more iterations for the algorithm to converge with lower incidence (p). Higher missing rate (r) was associated with requiring more iterations for convergence. Baseline prevalence π did affect the convergence performance under our simulation scenarios (Figure 1C). As expected, a higher missing rate (r) led to wider or more conservative CIs for p. Both large-sample interval estimated or

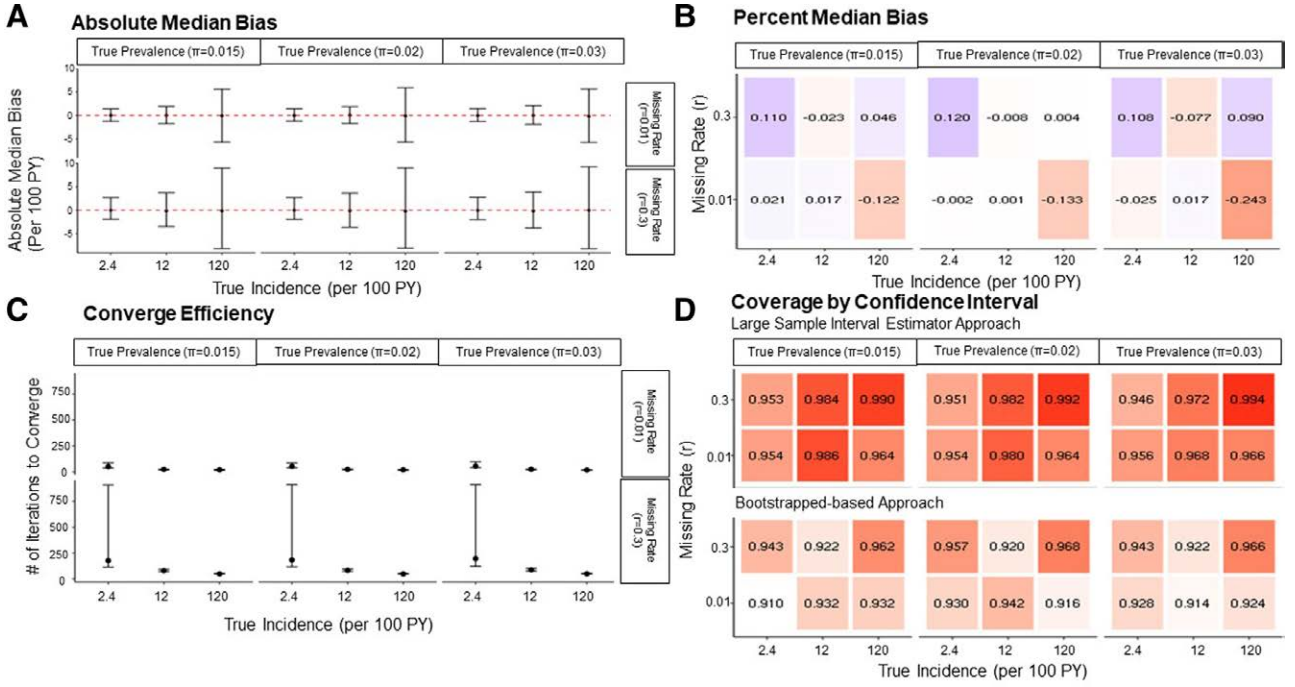

FIGURE 1. Impact of prevalence (π), incidence(p), missing rate (r) upon algorithm performance (sensitivity = 0.728 and specificity = 0.997). A, Absolute median bias: the median of the differences between the estimated incidence $\hat{p}$ and true incidence p. B, Percent median bias: ratio of the median bias to p. C, Converge efficiency: the number of iterations at the convergence. D, Coverage by confidence interval: the proportion of constructed CIs containing the true incidence p.

bootstrapped-based CIs reached nominal coverage levels with the large-sample interval being slightly more conservative (Figure 1D).

Scenarios Under Varying Assumptions of *sen*, *spe*

The expectation-maximization algorithm successfully converged for all the simulated datasets across 15 simulated scenarios. Sensitivity and specificity (*sen*, *spe*) did not have a big impact on bias (Figure 2A). It took fewer iterations for the algorithm to converge with increasing sensitivity or specificity (Figure 2B). All 95% CIs attained coverage levels above 90% and often close to 95%. As we saw previously, the large-sample interval estimated approach was slightly more conservative relative to the bootstrapped-based approach (Figure 2C).

Scenarios Under Varying Incidence Rates over Calendar Time

The expectation-maximization algorithm yielded estimates of α and β with low median bias of -0.005 (0.1% where $\alpha = -5$) and 0.008 (4.1% where $\beta = 0.2$), respectively. The empirical coverage level of the 95% CIs based on nonparametric bootstrapping was 94.2% for α and 94.6% for β. In addition, we rejected the hypothesis that the incidence rate is constant over time in 69.4% of the simulated data sets at the two-sided significance level of 0.05.

Results from Application to the TrackCOVID Study

We applied our method to the TrackCOVID study to estimate the prevalence and incidence with and without weights that accounted for stratified sampling. TrackCOVID was comprised of a total of 3860 participants enrolled in the six Bay Area counties, with a median follow-up of 3.03 months (interquartile range: 1.97–4.33 months). Table 5 shows the crude incidence, weighted incidence, and assay performance-adjusted weighted incidence in the entire cohort and by county, where we calculated the sensitivity of our case definition as 0.728 and its specificity as 0.997 (eAppendix 3; http://links.lww.com/EDE/C117).

The expectation-maximization algorithm converged within 600 iterations for all analyses. As expected, the assay performance-adjusted weighted incidence estimate was lower than the corresponding incidence rate, which does not adjust for assay performance. With a low underlying incidence, even a few false positive cases can substantially bias the unadjusted naive incidence estimate upward, as observed in our simulation study. Aligned with our simulation results, sensitivity, on the other hand, was not expected to have a large impact on our incidence estimates due to relatively low underlying incidence. To further illustrate this in our case study, we explored the trend of assay performance-adjusted incidence estimates under various values of sensitivity and specificity. When setting sensitivity at a fixed value of 0.728,

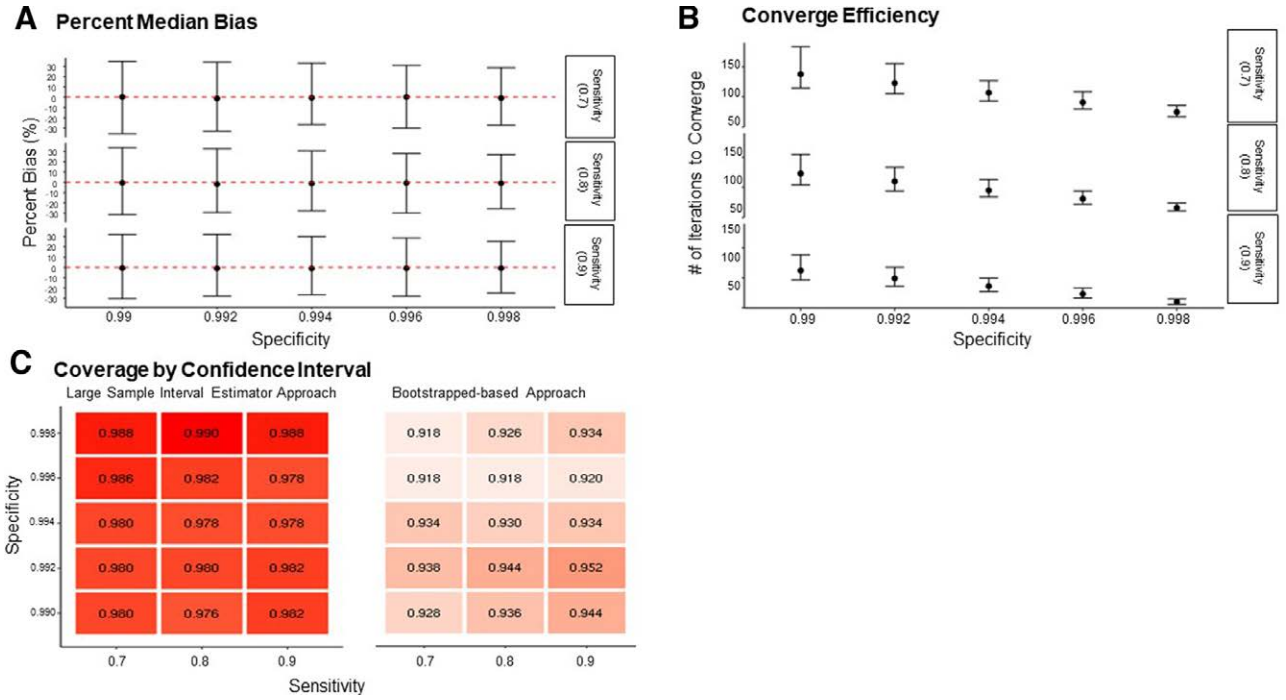

FIGURE 2. Impact of sensitivity, specificity upon algorithm performance (π = 0.015, p = 0.01, r = 0.3). A. Percent median bias: ratio of the median bias to p. B, Converge efficiency: the number of iterations at the convergence. C, Coverage by confidence interval: the proportion of constructed CIs containing the true incidence p.

the adjusted incidence estimate demonstrated wide variation (range: 0–12.6 per 100 person-years [PY]) for varying specificity levels (Figure 3A). In contrast, when specificity was fixed at 0.997, adjusted incidence estimate did not vary substantially (range: 8.9–9.2 per 100 PY) under different values of sensitivity (Figure 3B).

The CIs were constructed when there were more than 10 cases observed in the county. For San Mateo County, the laboratory performance-adjusted incidence was close to 0, which means it is likely that all seven incident cases identified in San Mateo County could be a result of false positives given the volume of lab tests at that county and the specificity of laboratory performance ($spe = 0.997$). This finding suggested that the true parameter of interest may be close to the boundary of the parameter space, where the large-sample approximation may provide misleading conclusions.

Last, we repeated our analysis assuming a time-varying incidence rate as

$$p(s;\theta) = \frac{\exp(\alpha_0 + \beta_0 s)}{1 + \exp(\alpha_0 + \beta_0 s)}.$$

The results are summarized in Table 5. For the entire cohort, the 95% CI of β_0 was 0.36 (0.041, 0.72), suggesting that the incidence rate indeed increased with time. Since $p(s, \theta) \approx \exp(\alpha_0 + \beta_0 s)$ for small α_0, we may conclude that on average the incidence rate increased by $\exp(\beta_0) - 1 = 43.7\%$ per month during the period from September 2020 to December 2021 (95% CI = 5.7%, 95.6%). $\beta_0 > 0$ suggested that there was a steep increase in COVID-19 incidence in the Bay Area during the study period, consistent with the initial outbreak of COVID-19 in the USA. When evaluating trends within the two biggest counties (Alameda and Santa Clara) and the other four counties combined (Contra Consta, Marin, San Francisco, and San Mateo), we observed nuances. Specifically, we observed an increasing trend for Alameda (51.1% monthly increase in the incidence, 95% CI = 12.6%, 102.6%), while analyses of Santa Clara and the other four counties combined showed increasing trends of incidence by 20.7% and 54.8% per month with CIs that included the value of zero. Overall, however, the results indicated an increasing trend of the incidence rate during the study period.

DISCUSSION

In this study, we proposed to use maximum likelihood estimation to estimate the prevalence and incidence with an adjustment for measurement error of incident cases due to laboratory performance. We evaluated the properties of our proposal via extensive simulation studies and applied them to a real study. Our algorithm demonstrated solid performance in most scenarios investigated. We demonstrated that both nonparametric bootstrapped-based and large-sample interval estimation methods worked well in constructing CIs with a reasonable sample size. If the sample size is small or the incidence rate is low, it is desirable to develop exact inference procedures for which the validity

TABLE 5. Comparisons of Crude Incidence, Weighted Incidence and Assay Performance-Adjusted Weighted Incidence in TrackCOVID study

						Assay Performance Adjusted Weighted Incidence (per 100 PY, Constant Incidence)		Assay Performance Adjusted Weighted Incidence (per 100PY, Time-varying Incidence Assuming Linear Trend over Time, Bootstrap CI with Robust Standard Errors)[d]	
County	**Cases**	**Weighted Cases**	**Risk Time (PY)**	**Weighted Risk Time (PY)**	**Incidence (per 100 PY)**	**Weighted Incidence (per 100 PY, Bootstrap CI)**	**LSIE CI**	**Bootstrapped-based CI**	**Incidence before and at September 2020 (per 100PY)[b]**	**Percent monthly increase of the incidence[c]**
Overall	86	124.94	992.51	936.54	8.66 (6.99, 10.59)	13.34 (11.23, 15.69)	9.04 (6.55, 12.48)	9.04 (5.17, 13.5)	4.12 (0.87, 19.35)	43.7% (5.7%, 95.6%)
Alameda	22	40.72	213.75	223.2	10.29 (6.56, 15.17)	18.24 (13.41, 23.95)	14.3 (8.40, 24.25)	14.3 (4.05, 25.7)	5.2 (1.18, 22.4)	51.1%(12.6%, 102.6%)
Santa Clara	25	40.45	231.26	272.82	10.81 (7.12, 15.54)	14.83 (10.83, 19.61)	10.5 (6.02, 18.12)	10.5 (2, 18.8)	10.2 (0.47, 204.0)	20.7% (−38.7%, 137.6%)
Contra Costa	13	21.71	101.62	135.91	12.75 (6.97, 20.81)	15.97 (10.25, 23.23)	10.7 (5.22, 21.98)	10.7 (1.48, 21.9)	1.95 (0.062, 60.1)	54.8% (−24.1%, 215.7%)
Marin	6	2.98	119.14	41.6	5.04 (1.87, 10.64)	7.17 (1.49, 19.6)	3.52 NA[a]	3.52 NA[a]		
San Francisco	13	13.95	191.61	152.4	6.78 (3.66, 11.32)	9.15 (5.09, 14.89)	6.22 (2.59, 14.88)	6.22 (0, 16.9)		
San Mateo	7	5.13	135.13	110.68	5.18 (2.11, 10.38)	4.63 (1.55, 10.38)	0 NA[a]	0 NA[a]		

[a]It is not applicable since <10 cases were identified.
[b]Estimated as $\frac{2400\times\exp(\alpha_0)}{1+\exp(\alpha_0)}$.
[c]Estimated as $exp\,(\beta_0) - 1$
[d]We combined the 4 counties for the time-varying adjusted weighted incidence analysis due to insufficient power with small number of cases in each county.
CI indicates confidence interval; LSIE, large-sample interval estimation; PY, person-years.

does not require large-sample approximations. Under the scenario where the underlying incidence was low, such as the TrackCOVID study, we found that even a few false positive cases (or slightly imperfect specificity), can substantially bias the unadjusted naive incidence estimate upward. This suggests that assuming perfect assay performance can lead to considerable inaccurate estimates of the incidence, particularly when the true prevalence and incidence are low. For example, when the latter is true, ignoring a test with specificity even as high as 99.7% led to an overestimation of incidence by 3.6 per 100 person-years, equivalent to an overestimation of the true incidence by 39.8%.

Our method relied on the assumption of constant incidence, which is a common assumption for most longitudinal cohort studies with similar goals. The original design of TrackCOVID assumed constant incidence by follow-up visits with incidence being estimated as the total number of new cases across all follow-up visits over total years of follow-up across all participants. We developed the algorithm based on this assumption and additionally provided guidance on how to relax this assumption. If this assumption is violated, our MLE should be interpreted as an "average" incidence, which may not always be meaningful. For example, in our TrackCOVID study, we noticed during the study implementation that the public health dashboard data suggested that the community incidence rate was much lower between July 2020 and November 2020 than that between December 2020 and March 2021 (the 2nd surge of COVID-19 pandemic).[31] It may very well be possible that most of the incident cases were coming from the last three follow-up visits during the latter half of the study (December 2020–March 2021). This may challenge the constant incidence assumption and would require a cautious interpretation of the estimated incidence rate. To test this assumption, we considered an important case where the incidence may depend on the calendar time. As a result, we found an increase in trend of incidence over calendar time for TrackCOVID. For studies where the disease landscape is changing or studied over a long period of time where preventive treatment and/or behaviors may vary and affect the incidence rate, we propose to use the extensions of our method discussed in the article so that the incidence rate can be estimated as a function of follow-up time or other factors of interest. With the fitted model, predictions for future incidences can be made. However, one may want to avoid predicting incidence over a long time horizon, since the simple extrapolation of the current model may not be valid. The model can also be extended to incorporate regression analysis of associations between other factors of interest and the underlying incidence. Furthermore, for a binary exposure of interest, we may use the proposed method to estimate the incidence rate among exposed and unexposed populations (e.g., disease status and vaccination status) after appropriate matching or weighting based on propensity scores. Thus, our method may prove useful for examining causal effects

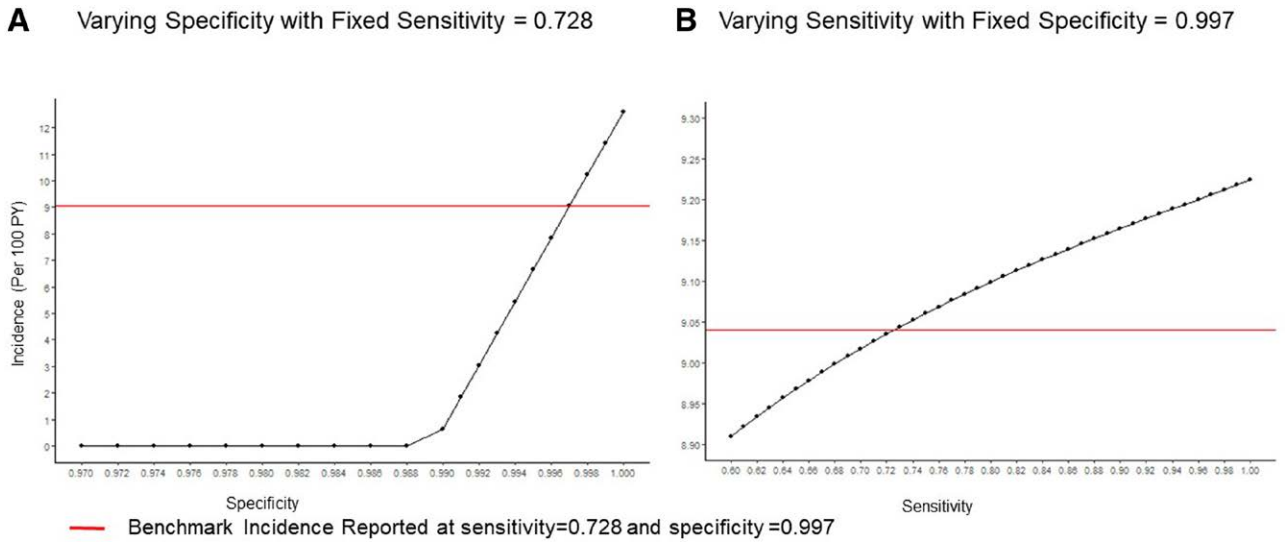

FIGURE 3. Lab performance-adjusted weighted incidence under different scenario of sensitivity and specificity (data from TrackCOVID study). Red horizontal line: benchmark incidence reported at sensitivity = 0.728 (A) and specificity = 0.997 (B).

if information on individual patients is available. This direction warrants further research. Finally, the constant incidence rate assumption can be examined empirically by applying our proposed method in subcohorts of participants, which allows us to compare the incidences by subpopulation. The key for our approach and corresponding extensions of the approach is to account for the sensitivity and specificity in relevant analyses.

Our study has limitations. First, we acknowledged that test performances were likely to vary by the type of tests/assays, immunity/vaccination status, and exposure status. Our method was used to estimate laboratory performance-adjusted incidence under the assumption that the assay performances (sensitivity and specificity) were well-established or prespecified based on prior literature or other studies. In practice, however, we often do not have reliable data on lab performance. As an imperfect remedy, we recommend performing analyses under a range of sensitivity and specificity measures to characterize their influence on the conclusions. To be noted, the reported sensitivity and specificity of the lab test or assay are often coming from the empirical data of limited sample size. When those data are available, it is desirable to consider the uncertainty in the sensitivity and specificity values when drawing statistical inferences on incidence. One practical approach is to bootstrap both the observed longitudinal cohort data and the empirical laboratory data that provides the sensitivity and specificity estimates. Second, in the TrackCOVID study, we imputed the intermittently missing data conditional on the outcome of the next nonmissing follow-up visit. Our method can be extended to allow general intermittent missing patterns without this ad-hoc imputation, and new scenarios beyond those in Table 1. However, in practice, we recommend concentrating on a moderate number of simple missing patterns that account for the majority of the patients to avoid complex likelihood function computation. Note that the true infection status trajectory of interest consists of 0s followed by 1s, where the first appearance of a 1 indicates the timing of the first positive test. The analysis of such outcomes can fit the framework of a discrete-time survival analysis. Finally, when the loss to follow-up rate is high, it may be important to consider the time-specific loss to follow-up instead of a constant rate assumed in the current method. An extension to accommodate this is straightforward, particularly when there is a sufficient number of patients with loss to follow-up at each visit and a reasonable point estimate of the missing rate can be obtained. Note that in our study and in many longitudinal studies such as ours, we expect to observe some loss to follow-up. In our study, we assumed that bias due to loss to follow-up was negligible, as we assumed that the loss to follow-up was missing at random (conditional on case status). That said, the impact and underlying reasons of loss to follow-up should be considered when applying these methods. To that end, we highly recommend devising additional analyses to understand the impact of loss to follow-up on the interpretation of findings from the primary analysis.

CONCLUSIONS

In this article, we proposed an MLE-based approach that allows estimation of incidence and prevalence that adjusts for measurement error of incident cases due to imperfect lab performance. Our method is flexible and has the potential to be extended to scenarios with varying incidence over time and time gaps between visits, as well as other external factors.

ACKNOWLEDGMENTS

The authors thank all project staff for their dedicated work on this study, the altruism of participants in both the TrackCOVID study, and community partners who collaborated on the recruitment of participants for the TrackCOVID cohort.

California Pandemic Consortium: Lloyd Minor, MD; Anna Graber; Christopher Leung; Anthony Bet; Matthew Sklar; Cole Holderman; Amanda Kempema; Hilary Tang; Jo-seph Lohmann; Dan Lowenstein, MD; David Glidden; Rodolfo Villa; Charles Craik; Jessica Chao; Aida Julien; Marcus Paoletti; Sravya Jaladanki; Guntas Padda; Daisy Valdivieso; Lillian Brown; Carina Marquez; Jacob Ghahremani; Emerald Wan; Audrey Mustoe; Steve Miller; Ralph Gonzalez; Joe DeRisi; Sabrina Mann; Emily Crawford; CLIA HUB consortium; Parul Bhargava.

REFERENCES

1. Subramanian R, He Q, Pascual M. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc Natl Acad Sci*. 2021;118:e2019716118.
2. Mercer TR, Marc S. "Testing at scale during the COVID-19 pandemic." *Nat Rev Genet*. 2021;22:415–426.
3. Pollock AM, Lancaster J. "Asymptomatic transmission of covid-19." *BMJ*. 2020;371:m4851.
4. *Centers for Disease Control and Prevention: CDC's Diagnostic Test for COVID-19 Only and Supplies*, https://www.cdc.gov/coronavirus/2019-ncov/lab/virus-requests.html. Accessed 11 Feb 2022.
5. *U.S. Food & Drug Administration: EUA Authorized Serology Test Performance*, https://archive.cdc.gov/#/details?q=virus-requests&start=0&rows=10&url=https://www.cdc.gov/coronavirus/2019-ncov/downloads/lab/multiplex-primers-probes-printer.pdf. Accessed 27 Feb 2024.
6. Sempos CT, Lu T. "Adjusting coronavirus prevalence estimates for laboratory test kit error." *Am J Epidemiol*. 2021;190:109–115.
7. Zhang Z, Bi Q, Fang S, Wei L, Wang X, He J, et al. Insight into the practical performance of RT-PCR testing for SARS-CoV-2 using serological data: a cohort study. *Lancet Microbe*. 2021;2:e79–e87.
8. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
9. Bajema KL, Wiegand RE, Cuffe K, Patel SV, Iachan R, Lim T, et al. Estimated SARS-CoV-2 seroprevalence in the US as of September 2020. *JAMA Intern Med*. 2021;181:450–460.
10. Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, et al. Seroprevalence of antibodies to SARS-CoV-2 in 10 sites in the United States, March 23-May 12, 2020. *JAMA Intern Med*. 2020;180:1576–1586.
11. Meireles P, Amaro J, Da Costa JP, Lopes MM, Varandas T, Norton P, et al. Prevalence of SARS-CoV-2 antibodies among workers of the public higher education institutions of Porto, Portugal: a cross-sectional study. *Occup Environ Med*. 2021;78:648–653.
12. Sahlu I, Whittaker AB. Obtaining prevalence estimates of coronavirus disease 2019: a model to inform decision-making. *Am J Epidemiol*. 2021;190:1681–1688.
13. Meyer MJ, Yan S, Schlageter S, Kraemer JD, Rosenberg ES, Stoto MA. Adjusting COVID-19 seroprevalence survey results to account for test sensitivity and specificity. *Am J Epidemiol*. 2022;191:681–688.
14. Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Stat Med*. 2014;33:4141–4169.
15. Speybroeck N, Devleesschauwer B, Joseph L, Berkvens D. Misclassification errors in prevalence estimation: Bayesian handling with care. *Int J Public Health*. 2013;58:791–795.
16. Burstyn I, Goldstein ND, Gustafson P. Towards reduction in bias in epidemic curves due to outcome misclassification through Bayesian analysis of time-series of laboratory test results: case study of COVID-19 in Alberta, Canada and Philadelphia, USA. *BMC Med Res Methodol*. 2020;20:1–10.
17. Becker NG, Britton T. "Statistical studies of infectious disease incidence.". *J Roy Statistical Society*. 1999;61:287–307.
18. Caruana EJ, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. *J Thorac Dis*. 2015;7:E537–E540.
19. Morgenstern H, Kleinbaum DG, Kupper LL. Measures of disease incidence used in epidemiologic research. *Int J Epidemiol*. 1980;9:97–104.
20. Bhopal RS. *Concepts of epidemiology: integrating the ideas, theories, principles, and methods of epidemiology*. Oxford University Press, 2016.
21. Gan G, Bain LJ. Some results for type I censored sampling from geometric distributions. *J Stat Plan Inference*. 1998;67:85–97.
22. Couris CM, Polazzi S, Olive F, et al. "Breast cancer incidence using administrative data: correction with sensitivity and specificity.". *J Clin Epidemiol*. 2009;62:660–666.
23. Moore CL, Amin J, Gidding HF, Law MG. "A new method for assessing how sensitivity and specificity of linkage studies affects estimation." *PLoS One*. 2014;9:e103690.
24. McDougal JS, Parekh BS, Peterson ML, et al. Comparison of HIV-1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. *AIDS Res Hum Retroviruses*. 2006;22:945–952.
25. Brookmeyer R. "Should biomarker estimates of HIV incidence be adjusted?." *AIDS*. 2009;23:485–491.
26. Hargrove JW, Humphrey JH, Mutasa K, et al. Improved HIV-1 incidence estimates using BED capture enzyme immunoassay. *AIDS*. 2008;22:511–518.
27. Lindan CP, Desai M, Boothroyd D, et al. Design of a population-based longitudinal cohort study of SARS-CoV-2 incidence and prevalence among adults in the San Francisco Bay Area. *Ann Epidemiol*. 2021;67:81–100.
28. Efron B. "Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods." *Biometrika*. 1981;68:589–599.
29. Abt M, Welch WJ. Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes. *Canadian Journal of Statistics*. 1998;26:127–137.
30. Scott, William A. "Maximum likelihood estimation using the empirical fisher information matrix." *J Stat Comput Simul*. 2002;72:599–611.
31. *California Open Data Portal: COVID-19 Time-Series Metrics by County and State*, Available at: https://data.ca.gov/dataset/covid-19-time-series-metrics-by-county-and-state-archived. Accessed 27 Feb 2024.