Epidemics 41 (2022) 100634

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Optimizing prevalence estimates for a novel pathogen by reducing uncertainty in test characteristics

Daniel B. Larremore [a,b,*,1], Bailey K. Fosdick [c,**,1], Sam Zhang [d], Yonatan H. Grad [e]

[a] Department of Computer Science, University of Colorado Boulder, Boulder, CO, 80309, USA
[b] BioFrontiers Institute, University of Colorado at Boulder, Boulder, CO, 80303, USA
[c] Department of Statistics, Colorado State University, Fort Collins, CO, 80523, USA
[d] Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO, 80309, USA
[e] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA

ARTICLE INFO

Keywords:
Bayesian inference
Prevalence
Sample size calculation
Sensitivity
Specificity

ABSTRACT

Emergence of a novel pathogen drives the urgent need for diagnostic tests that can aid in defining disease prevalence. The limitations associated with rapid development and deployment of these tests result in a dilemma: In efforts to optimize prevalence estimates, would tests be better used in the lab to reduce uncertainty in test characteristics or to increase sample size in field studies? Here, we provide a framework to address this question through a joint Bayesian model that simultaneously analyzes lab validation and field survey data, and we define the impact of test allocation on inferences of sensitivity, specificity, and prevalence. In many scenarios, prevalence estimates can be most improved by apportioning additional effort towards validation rather than to the field. The joint model provides superior estimation of prevalence, sensitivity, and specificity, compared with typical analyses that model lab and field data separately, and it can be used to inform sample allocation when testing is limited.

1. Introduction

Prevalence is traditionally estimated by analyzing the outcomes from diagnostic tests given to a subset of the population. During analysis of these outcomes, the sensitivity and specificity of the test, as well as the number of samples in the survey, are incorporated into point estimates and uncertainty bounds for the true prevalence. In many cases, sensitivity and specificity are taken to be fixed characteristics of the test (Flahault et al., 2005; Reiczigel et al., 2010). However, sensitivity and specificity are themselves estimated from test outcomes in validation studies. As a result, they, too, carry statistical uncertainty, and that statistical uncertainty should be carried forward into estimates of prevalence (Rogan and Gladen, 1978; Stringhini et al., 2020; Gelman and Carpenter, 2020). Since prevalence estimates may improve as sample size increases and with reduced uncertainty in the test characteristics, a fundamental study design question arises: Given limited testing capacity, how should one allocate tests between the field and validation lab? This question is especially pertinent with the emergence of a novel pathogen, when test availability is limited and diagnostic tests are not yet well validated.

Here, we review the derivation for a Bayesian joint posterior distribution for prevalence and test sensitivity and specificity based on sampling models for both the field survey data and validation data, originally introduced in Gelman and Carpenter (2020). While others have demonstrated how to estimate prevalence from this model or extensions of it (Stringhini et al., 2020; Levesque and Maybury, 2020; Nisar et al., 2021; Levin et al., 2022; Lopez et al., 2022), we highlight the utility of this model for addressing the problem of how to allocate a fixed number of tests between the field and the lab to produce the best prevalence estimates when the testing capacity is limited. We demonstrate that, when the sensitivity and specificity of a test have not yet been well established, the largest improvement in prevalence estimates could result from allocating samples to test validation rather than to the survey. Finally, we showcase how this joint model can produce improved estimates of sensitivity and specificity compared to models based only on the lab data.

* Correspondence to: Department of Computer Science, 111 Engineering Drive, ECOT 717, 430 UCB, Boulder, CO 80309, USA.
** Corresponding author.
E-mail addresses: daniel.larremore@colorado.edu (D.B. Larremore), bailey.fosdick@colostate.edu (B.K. Fosdick).
[1] Contributed equally.

https://doi.org/10.1016/j.epidem.2022.100634
Received 9 February 2022; Received in revised form 11 July 2022; Accepted 21 September 2022
Available online 27 September 2022
1755-4365/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2. Methods

Our goal is to estimate population prevalence (θ), test sensitivity (se), and test specificity (sp) by learning from the field survey data X and the validation data V. The field survey data X contains the number of positive tests (n_+) out of N_{field} samples. The validation data V contains the number of true positives (tp) resulting from N_{pos} positive control samples, providing information on test sensitivity, and the number of true negatives (tn) resulting from N_{neg} negative control samples, providing information about test specificity. We employ Bayes' rule for estimation

$$\Pr(\theta, se, sp \mid X, V) \propto \Pr(X, V \mid \theta, se, sp) \,, \tag{1}$$

where we assume independent uniform priors on each of the parameters $\{\theta, se, sp\}$. Note that informative priors could also be used for se and sp, were data from other studies or from the manufacturer are available. Survey data X and validation data V are collected independently via different processes but share the test's sensitivity and specificity. We rewrite Eq. (1) as

$$\Pr(\theta, se, sp \mid X, V) \propto \Pr(X \mid \theta, se, sp)\ \Pr(V \mid se, sp) \,. \tag{2}$$

Given the parameter values $\{\theta, se, sp\}$, the probability that a single random field test is positive is equal to the probability of obtaining a true positive or a false positive: $p = \theta se + (1-\theta)(1-sp)$. Assuming the field sample represents a random sample from the population and uniform prevalence across the population, the number of positive outcomes after N_{field} independent tests is binomially distributed, so the probability of observing n_+ positive tests is

$$\Pr(n_+ \mid \theta, se, sp) = \binom{N_{\text{field}}}{n_+} p^{n_+}(1-p)^{N_{\text{field}}-n_+} \,. \tag{3}$$

If the true test sensitivity is se, then the probability that a known positive sample produces a positive test outcome – a true positive – is se, while the probability of a false negative is $1 - se$. The number of true positives tp in a set of N_{pos} independent positive validation tests is also binomially distributed:

$$\Pr(tp \mid se) = \binom{N_{\text{pos}}}{tp} se^{tp}(1-se)^{N_{\text{pos}}-tp} \,. \tag{4}$$

A parallel argument for true specificity sp and the outcomes of N_{neg} independent negative validation tests leads to the probability of observing tn true negatives:

$$\Pr(tn \mid sp) = \binom{N_{\text{neg}}}{tn} sp^{tn}(1-sp)^{N_{\text{neg}}-tn} \,. \tag{5}$$

Substituting the probabilities in Eqs. (3), (4), and (5) into Eq. (2) and absorbing constants into the proportion, we obtain

$$\begin{aligned}\Pr(\theta, se, sp \mid X, V) \propto \Big(& [1 - sp + \theta(se + sp - 1)]^{n_+} \\ & \times [sp - \theta(se + sp - 1)]^{N_{\text{field}}-n_+} \Big) \\ \times \quad & sp^{tn}(1-sp)^{N_{\text{neg}}-tn} se^{tp}(1-se)^{N_{\text{pos}}-tp} \,.\end{aligned} \tag{6}$$

Eq. (6) provides the form of the *joint* posterior distribution of θ, se, and sp, allowing one to learn simultaneously about these quantities and see how they depend on the data. Although the joint posterior distribution is not amenable to analytic computations (e.g., calculating expectations and variances), it is easily sampled using a Markov chain Monte Carlo (MCMC) algorithm. These posterior samples can then be used to estimate any summary statistics of interest, including point estimates (e.g., posterior means and modes) and credible intervals for the parameters. Posterior samples can further be passed as inputs into subsequent modeling tasks to account for uncertainty in prevalence (Larremore et al., 2021; Kissler et al., 2020). (See an in-browser javascript calculator for computing the posterior distribution https://larremorelab.github.io/covid19testgroup and open-source code in R and Python https://github.com/LarremoreLab/bayesian-joint-prev-se-sp. Gelman and Carpenter (2020) also provide Stan code for estimating this model.)

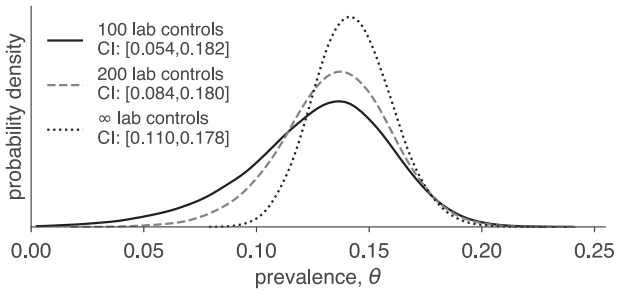

Fig. 1. Increased validation effort decreases prevalence uncertainty. Prevalence estimates from 75 (n_+) positives in 500 (N_{field}) field samples, using validation outcomes of $\{tp, tn\} = \{47, 49\}$ based on $N_{\text{neg}} = N_{\text{pos}} = 50$ samples (solid line), $\{94, 98\}$ based on $N_{\text{neg}} = N_{\text{pos}} = 100$ samples (dashed line). Widths of 95% credible intervals decreased by 24% (prevalence), 32% (sensitivity), and 34% (specificity) due to increased validation efforts. Dotted line shows a Bayesian analysis of the same data using point estimates of 94% sensitivity and 98% specificity, equivalent to infinite lab validation data, for reference.

This model framework assumes that each diagnostic test is independent of the others and that the conditions in the field and validation are sufficiently similar that the diagnostic test has the same sensitivity and specificity in both. This assumption, and relaxations thereof, are considered in the Discussion.

3. Results

To demonstrate the impact of conducting additional validation tests, we computed the posterior distributions in Eq. (6) for two scenarios in which field survey data consisting of 75 positive and 425 negative tests were analyzed using two sets of validation data. The first set was based on 100 validation tests and the other based on 200 validation tests, with tests split equally between positive and negative controls in both cases. Both validation data sets contained 94% true positives and 98% true negatives. The increase in validation samples resulted in a change in the 95% posterior credible interval for prevalence from $[0.054, 0.182]$ to $[0.084, 0.180]$ (Fig. 1), corresponding to a 24% reduction in the credible interval width from additional validation data alone.

To compare the results of finite validation efforts to the theoretical optimum of infinite validation tests, we computed a posterior distribution for prevalence assuming known values of sensitivity and specificity of 94% and 98%, respectively. This results in a decrease in the width of the posterior credible interval by an additional 30% (Fig. 1). The marginal impact of each additional validation test on posterior prevalence uncertainty decreases as this theoretical limit is approached.

When there is a limit on the number of tests that a prevalence study can use, due to budget, time, throughput, or other constraints, it may be tempting to deploy as many tests as possible to the field. This follows an intuition that additional field samples will decrease uncertainty in estimates of θ. However, additional validation samples will also indirectly decrease uncertainty in θ by reducing uncertainty around sensitivity and/or specificity. By taking posterior uncertainty of prevalence as the quantity to be minimized, we can search over combinations of N_{field}, N_{neg}, and N_{pos}, representing the numbers of field, negative control, and positive control tests, respectively. When the total number of tests $N = N_{\text{field}} + N_{\text{neg}} + N_{\text{pos}}$ is fixed, only two sample sizes can be specified freely, which means that this sample allocation problem becomes a minimization over a two-dimensional grid.

To demonstrate the use of this approach, we considered the allocation of $N = 1000$ tests in a setting where sensitivity and specificity are

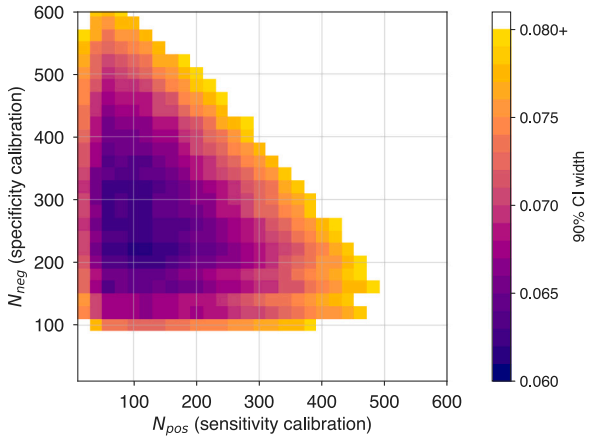

Fig. 2. Optimized allocation of tests. Uncertainty in prevalence estimates, represented as 95% credible interval width, is shown as a heatmap for various allocations of $N = 1000$ tests, when prevalence is suspected to be 0.15, sensitivity 0.93, and specificity 0.98. Each pixel represents a choice of N_{neg} and N_{pos}, where $N_{\text{field}} = N - N_{\text{neg}} - N_{\text{pos}}$. Widths are indicated by color (see colorbar) with values larger than 0.09, or invalid choices of N_{pos} and N_{neg}, in white. Each pixel was computed based on data equal to the expected test results for that allocation and using posterior samples from Eq. (6). Optimal allocations for the studied scenario favor allocation to negative controls over positive controls, with only 600–700 samples allocated to the field survey. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

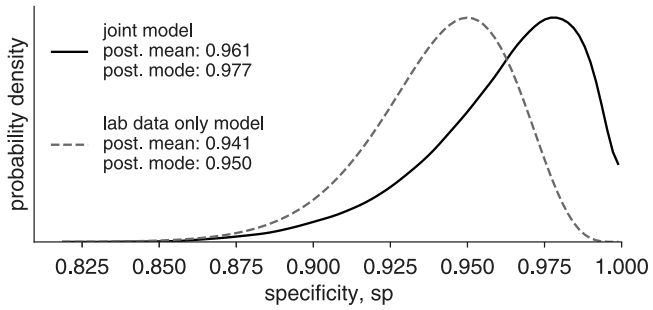

Fig. 3. Test outcomes from the field affect estimates of sensitivity and specificity. Specificity estimates are shown for validation outcomes $\{tp, tn\} = \{100, 95\}$ based on $N_{\text{pos}} = N_{\text{neg}} = 100$ controls analyzed independently of field data (dashed line; Beta posterior distribution) or jointly with $n_+ = 10$ positives in $N_{\text{field}} = 1000$ field samples (solid line). While Fig. 1 illustrates the influence of lab validation data on prevalence estimates, this figure illustrates the less intuitive influence of field survey data on specificity estimates. This effect of field data is strongest on specificity when prevalence is low, and strongest on sensitivity when prevalence is high.

suspected to be around $se = 0.93$ and $sp = 0.98$ and in a population with suspected prevalence of 0.15. We allocated N_{pos} and N_{neg} to positive and negative controls, respectively, with the remainder allocated to N_{field}. We then sampled from the posterior distribution for θ in Eq. (6) conditional on data equal to the expected counts of tp, tn, and n_+. From these posterior samples, we computed the width of the 90% credible interval, and recorded it, before continuing to a new choice of sample allocation. Through this process, we found that at least twice as many samples should be allocated to specificity validation (N_{neg}) as compared to sensitivity validation (N_{pos}), and that around 1/3 of the 1000 total tests should be used for validation instead of for the field study (Fig. 2).

An additional consequence of jointly modeling the validation and field data is that estimates of sensitivity and specificity may be affected by field survey data. Mathematically, this is because sensitivity and specificity appear in the probabilities of both the field and lab data sets in Eqs. (3), (4), and (5). To illustrate this point, we considered a scenario in which 95 of 100 negative controls were found to be negative during validation, resulting in a point estimate of specificity of 0.95, followed by a large study in a low prevalence area that resulted in only 10 positive tests out of 1000 samples. Such field data would appear inconsistent with the validation data, because even if prevalence were zero, one would expect 50 positives from 1000 field tests. However, an analysis based on Eq. (6) resolves this apparent inconsistency by inferring that the test's specificity is likely to be higher than 0.95, with a posterior mean of 0.961 and posterior mode of 0.977 (Fig. 3, solid line). For comparison, we also analyzed the validation data separately using a uniform prior on specificity, which produced a beta posterior distribution with a posterior mean of 0.941 and a posterior mode at 0.95 (Fig. 3, dashed line).

4. Discussion

The sensitivity and specificity of a diagnostic test are inferred from a finite number of validations tests. As a consequence, sensitivity and specificity themselves carry uncertainty, which affects the statistical interpretation of prevalence surveys in the field. As shown in Gelman and Carpenter (2020), studies that use only point estimates of test characteristics can dramatically underestimate uncertainty around prevalence (e.g., Fig. 1). Here, we reviewed how this issue can be ameliorated by jointly modeling field data and validation data using standard Bayesian techniques. Our results provides insight for two general lines of intuition. First, when prevalence is low, validation samples should be preferentially allocated to specificity over sensitivity, with the opposite recommendation for high prevalence scenarios. Second, without strong prior information about the sensitivity and specificity of a diagnostic test, substantial validation efforts are required.

Bayesian frameworks such as the one utilized here can be used even when no validation data is available (Joseph et al., 1995; Branscum et al., 2005; Toft et al., 2005; Diggle, 2011). Furthermore, they can easily incorporate prior information about prevalence, sensitivity, or specificity from other pilot or validation studies – either directly, as validation data, or indirectly, through the considered use of informative priors – and can jointly model the application of multiple diagnostic tests with different performance characteristics simultaneously (Joseph et al., 1995) or scenarios when multiple lab validation data sets are available (Gelman and Carpenter, 2020). These methods also avoid the need to rely on asymptotic approximations (Flahault et al., 2005) in the process of calculating confidence intervals.

The direct inclusion of validation tests in prevalence estimation not only allows uncertain sensitivity and specificity to affect prevalence estimates (Figs. 1 and 2), but also allows field data to affect sensitivity and specificity estimates (Fig. 3). This underscores the importance of reporting the raw outcomes from validation tests. The outcomes of validation tests should be included directly in publications that analyze field data whenever possible, motivated by statistical and reproducibility requirements. In some cases, the teams developing the diagnostic test in the lab and those deploying tests in the field are different. This work highlights the value of feedback between these groups, as collaborative efforts could improve test assay development and refinement, especially as new variants emerge and test performance may change. This study's results rely on a number of assumptions. First, we assumed that validation samples are representative of the populations surveyed in the field. However, in the rapid deployment of SARS-CoV-2 seroprevalence surveys, for instance, positive control samples were restricted to only symptomatic and virologically confirmed COVID-19 cases, leading to validation samples that do not fully represent the antibody responses of asymptomatic individuals (Long et al., 2020). Second, we assumed identical diagnostic protocols for lab and field samples, as well as identical test performance in both settings. While test performance can vary across these settings in practice (Greiner and Gardner, 2000), during the emergence of a new pathogen there is likely not time for ample test validation before test deployment, thus making

this assumption necessary. Together, these assumptions allowed us to jointly model lab and field data in Eq. (6), which makes explicit the often implicit mathematical link between the field and the lab.

By highlighting the marginal value of additional validation effort, joint models like Eq. (6) expose the tradeoff between collecting validation and field data when testing capacity is limited. This simulation-informed approach to sample allocation allows a finite number of samples to be maximally utilized via strategic study design (Hens et al., 2012; Blaizot et al., 2019; Larremore et al., 2021). Differing costs between lab and field test deployment may heavily influence decisions about test allocation. While such costs are not explicitly considered here, the current framework could be extended to include cost by redefining the objective as the minimization of a combination of both prevalence uncertainty and total cost.

CRediT authorship contribution statement

Daniel B. Larremore: Conceptualization, Methodology, Software, Writing. **Bailey K. Fosdick:** Conceptualization, Methodology, Software, Writing. **Sam Zhang:** Methodology, Software. **Yonatan H. Grad:** Conceptualization, Methodology, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

DBL and YHG were supported in part by the SeroNet program of the National Cancer Institute, USA (1U01CA261277-01). The work of YHG was also supported in part by the Morris-Singer Fund for the Center for Communicable Disease Dynamics at the Harvard T.H. Chan School of Public Health and contract 200-2016-91779 with the Centers for Disease Control and Prevention. Disclaimer: The findings, conclusions, and views expressed are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC).

References

Blaizot, Stephanie, Herzog, Sereina A., Abrams, Steven, Theeten, Heidi, Litzroth, Amber, Hens, Niel, 2019. Sample size calculation for estimating key epidemiological parameters using serological data and mathematical modelling. BMC Med. Res. Methodol. 19 (51).

Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prevent. Vet. Med. 68 (2–4), 145–163.

Diggle, Peter J., 2011. Estimating prevalence using an imperfect test. Epidemiol. Res. Int. 2011.

Flahault, Antoine, Cadilhac, Michel, Thomas, Guy, 2005. Sample size calculation should be performed for design accuracy in diagnostic test studies. J. Clin. Epidemiol. 58 (8), 859–862.

Gelman, Andrew, Carpenter, Bob, 2020. Bayesian analysis of tests with unknown specificity and sensitivity. J. R. Stat. Soc. Ser. C. Appl. Stat. 69 (5), 1269–1283.

Greiner, Matthias, Gardner, I.A., 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. Prev. Vet. Med. 45 (1–2), 3–22.

Hens, Niel, Shkedy, Ziv, Aerts, Marc, Faes, Christel, Van Damme, Pierre, Beutels, Philippe, 2012. Modeling Infectious Disease Parameters Based on Serological and Social Contact Data: A Modern Statistical Perspective, Vol. 63. Springer Science & Business Media.

Joseph, Lawrence, Gyorkos, Theresa W., Coupal, Louis, 1995. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am. J. Epidemiol. 141 (3), 263–272.

Kissler, Stephen M., Tedijanto, Christine, Goldstein, Edward, Grad, Yonatan H., Lipsitch, Marc, 2020. Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period. Science.

Larremore, Daniel B., Fosdick, Bailey K., Bubar, Kate M., Zhang, Sam, Kissler, Stephen M., Metcalf, C. Jessica E., Buckee, Caroline, Grad, Yonatan, 2021. Estimating SARS-CoV-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. Elife 10, e64206.

Levesque, Jérôme, Maybury, David W., 2020. A note on COVID-19 seroprevalence studies: a meta-analysis using hierarchical modelling. MedRxiv.

Levin, Andrew T., Owusu-Boaitey, Nana, Pugh, Sierra, Fosdick, Bailey K., Zwi, Anthony B, Malani, Anup, Soman, Satej, Besançon, Lonni, Kashnitsky, Ilya, Ganesh, Sachin, McLaughlin, Aloysius, Song, Gayeong, Uhm, Rine, Herrera-Esposito, Daniel, de los Campos, Gustavo, Peçanha Antonio, Ana Carolina, Tadese, Enyew Birru, Meyerowitz-Katz, Gideon, 2022. Assessing the burden of COVID-19 in developing countries: systematic review, meta-analysis and public policy implications. BMJ Glob. Health 7 (5).

Long, Quan-Xin, Tang, Xiao-Jun, Shi, Qiu-Lin, Li, Qin, Deng, Hai-Jun, Yuan, Jun, Hu, Jie-Li, Xu, Wei, Zhang, Yong, Lv, Fa-Jin, Su, Kun, Zhang, Fan, Gong, Jiang, Wu, Bo, Liu, Xia-Mao, Li, Jin-Jing, Qiu, Jing-Fu, Chen, Juan, Huang, Ai-Long, 2020. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat. Med..

Lopez, Cesar A., Cunningham, Clark H., Pugh, Sierra, Brandt, Katerina, Vanna, Usaphea P., Delacruz, Matthew J., Guerra, Quique, Goldstein, Samuel Jacob, Hou, Yixuan J., Gearhart, Margaret, Wiethorn, Christine, Pope, Candace, Amditis, Carolyn, Pruitt, Kathryn, Newberry-Dillon, Cinthia, Schmitz, John, Premkumar, Lakshmanane, Adimora, Adaora A., Emch, Michael, Boyce, Ross, Aiello, Allison E., Fosdick, Bailey K., Larremore, Daniel B., de Silva, Aravinda M., Juliano, Jonathan J, Markmann, Alena J., 2022. Ethnoracial disparities in SARS-CoV-2 seroprevalence in a large cohort of individuals in central North Carolina from April to December 2020. mSphere 7 (3), e00841–21.

Nisar, Muhammad Imran, Ansari, Nadia, Khalid, Farah, Amin, Mashal, Shahbaz, Hamna, Hotwani, Aneeta, Rehman, Najeeb, Pugh, Sierra, Mehmood, Usma, Rizvi, Arjumand, et al., 2021. Serial population-based serosurveys for COVID-19 in two neighbourhoods of Karachi, Pakistan. Int. J. Infect. Dis. 106, 176–182.

Reiczigel, J., Földi, J., Ózsvári, L., 2010. Exact confidence limits for prevalence of a disease with an imperfect diagnostic test. Epidemiol. Infect. 138 (11), 1674–1678.

Rogan, Walter J., Gladen, Beth, 1978. Estimating prevalence from the results of a screening test. Am. J. Epidemiol. 107 (1), 71–76.

Stringhini, Silvia, Wisniak, Ania, Piumatti, Giovanni, Azman, Andrew S, Lauer, Stephen A, Baysson, Hélène, De Ridder, David, Petrovic, Dusan, Schrempft, Stephanie, Marcus, Kailing, et al., 2020. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. Lancet 396 (10247), 313–319.

Toft, Nils, Jørgensen, Erik, Højsgaard, Søren, 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev. Vet. Med. 68 (1), 19–33.