Bias in (sero)prevalence estimates

Sarah R Haile (sarah.haile@uzh.ch)
ORCID: 0000-0002-4704-6570
Affiliation:
Epidemiology, Biostatistics and Prevention Institute (EBPI)
University of Zurich
Hirschengraben 84
8001 Zurich

Abstract

Thousands of studies of SARS-CoV-2 seroprevalence have been published since the beginning of the pandemic. Researchers have reported a range of methods used to estimate seroprevalence from study data. Because diagnostic tests are imperfect, false negatives and false positives can be expected, as typically described by a test's sensitivity and specificity. A number of methods exist in the statistical literature to correctly estimate disease prevalence in the presence of test misclassification, but these methods seem to be less known and not routinely used in the epidemiology literature. Using straightforward calculations, we estimated the amount of bias introduced when reporting the proportion of positive test results instead of using sensitivity and specificity to estimate disease prevalence. We also derived a range of disease prevalence, given sensitivity and specificity, where this bias is within desired tolerances. Expected bias is often more than is desired in practice, ranging from 1% to 10%. We apply these results a recent school-based cohort study of COVID-19 in children in Switzerland. We conclude with some tips on software packages which may be used to compute less biased seroprevalence estimates.

Keywords: Prevalence, Seroprevalence, Diagnostic Tests, Statistical Methods

1 Introduction

Since the beginning of the SARS-CoV-2 pandemic, thousands of papers have been published detailing seroprevalence estimates in various populations (1). As a reader and sometimes reviewer of such publications, I noticed that while some researchers used simple approaches such as proportions or logistic regression, others used complicated methods like Bayesian hierarchical models. This made me wonder what, if any, degree of bias was introduced by using one method or the other.

As diagnostic tests are not 100% correct, it is expected that some small number of test results will be either false positives or false negatives. Using a simple proportion of the number of positive diagnostic tests over the total number of tests ignores any misclassification inherent to the test. In the case where there are similar numbers of true positives and true negatives in the population, the bias introduced by using the proportion of positive tests to estimate the proportion of subjects with the disease may not be very high. However, if the rate of false positives differs greatly from that of false negatives, the bias may be quite large.

For example, in Table 1, 25.6% (256/1000) of subjects had a positive test result, but the true disease prevalence in this population is 19.6% (196/1000). So there is a bias of 6% because there are many more false positive test results (n = 80) than false negatives (n = 20). Statisticians often talk about sensitivity and specificity of the diagnostic in relation to these quantities (described in more detail below), but it is accepted that without an accepted "gold standard" diagnostic tool, it is difficult to accurately assess disease prevalence.

Accounting for such misclassification in the interpretation of diagnostic tests is certainly not new in the literature. A straightforward method of adjusting observed prevalence is available (2) and has been extended to compute confidence intervals (3,4). Bayesian approaches have also been developed (for example, 5,6,7). A comparison of Bayesian and frequentist methods (8) showed that Bayesian methods are to be preferred, or the method of (2) with confidence intervals of (4).

Despite this extensive treatment of the misclassification problem in the statistical literature, many researchers appear to not realize they may be publishing biased results or know what to do about it. In what follows, my aim is to quantify the bias introduced in using the proportion of positive tests to estimate the proportion of subjects with disease in order to emphasize the need for methods other than standard logistic regression (a naive approach, assuming that all subjects with a positive test are also disease positive). I will describe key concepts, and derive an estimate of the bias, as well as a range of prevalences where such naive estimates show low bias. Bias estimates will be described according to test sensitivity and specificity, and I will apply these results to a real example of SARS-CoV-2 seroprevalence in children. A brief discussion will follow.

2 Methods

To start, I introduce some notation. Disease status, D, is denoted 1 if a subject has the disease in question (or for the case of seroprevalence, has antibodies for it), and 0 otherwise. Similarly, the result of the diagnostic test, Y, is given as 1 if the subject tests positive for the disease, and 0 otherwise. FP is often used to refer to false positive test results, and similarly FN for false negatives, TN for true negatives and TP for true positives.

Prevalence is the probability of having the disease, $P = Pr(D = 1)$. Sensitivity, denoted Se, sometimes also called the true positive fraction (TPF), is the probability of having a positive test result, given that the subject has the disease, $Pr(Y = 1|D = 1)$ (9). On the other hand, specificity, Sp is the probability of having a negative test result when a subject does not have the disease, $Pr(Y = 0|D = 0)$ (sometimes 1 - specificity is discussed, which is often referred to as false positive fraction, or FPF (10)). In real settings where true disease status is known via another method, sometimes referred to as the "gold standard", Se can be computed as $TP/(TP + FN)$, where TP is the number of true positives and FN is the number of true negatives. Similarly, Sp can be computed as $1 - FP/(FP + TN)$.

The proportion of positive tests can be expressed as $Pr(Y = 1) = (FP + TP)/(FP + TP + TN + FN)$, while the disease prevalence in the sample can be expressed as $Pr(D = 1) = (FN + TP)/(FP + TP + TN + FN)$. The difference between these two quantities is simply $(FP - FN)/(FP + TP + TN + FN)$.

According to Bayes' Theorem (11), the proportion of false positives can be written as $Pr(Y = 1, D = 0) = Pr(Y = 1|D = 0)Pr(D = 0)$, which simplifies to $(1 - P)(1 - Sp)$. In a similar fashion, the proportion of false negatives can be written as $Pr(Y = 0, D = 1) = Pr(Y = 0|D = 1)Pr(D = 1)$, which simplifies to $P(1 - Se)$. The bias when using the proportion of positive tests ($Pr(Y = 1)$) to estimate the proportion with disease ($Pr(D = 1)$) is therefore $(1 - P)(1 - Sp) - P(1 - Se)$ or equivalently $1 - Sp + P(Sp + Se - 2)$.

Suppose we want to guarantee that the bias is no larger than, say, $\delta = 0.02$. We can solve

$$-\delta \leq 1 - Sp + P(Sp + Se - 2) \leq \delta$$

for P, to get:

$$\max\left(\frac{\delta + Sp - 1}{Sp - Se - 2}, 0\right) \leq P \leq \min\left(\frac{-\delta + Sp - 1}{Sp + Se - 2}, 1\right).$$

3 Results

Using the result $bias = 1 - Sp + P(Sp + Se - 2)$ described above, I calculated the expected bias for a range of reasonable combinations of sensitivity, specificity and disease prevalence (Table 2). When sensitivity and

specificity were both 90%, bias was as high as 10%, especially near prevalences of 0% or 100%. When specificity was 90%, a bias of 10% could be expected with small prevalences near 0% even if sensitivity was 99%. The least bias, 1%, could be expected where sensitivity and specificity were both 99%.

Next, using the bounds of prevalence as derived above, I explored where the maximum tolerated bias is limited to 1%, 2.5% and 5%. When Se and Sp are each 90%, bias is within a tolerance of 1% only very close to 50% disease prevalence, within 2.5% tolerance in the range of 38% - 62% disease prevalence and to within 5% tolerance as long as disease prevalence is between 25% and 75%. When the desired tolerance is 1%, the range of disease prevalence where a naive approach will yield unbiased results is fairly narrow in all cases, unless Se and Sp are each 99%.

As an example of this, take the Ciao Corona study (12), a school-based longitudinal study of seroprevalence in Swiss school children with 5 rounds of SARS-CoV-2 antibody testing between June 2020 and June 2022, covering a range of seroprevalences in the population (Trial Registration: ClinicalTrials.gov NCT04448717). The study was approved by the Ethics Committee of the Canton of Zurich, Switzerland (2020-01336). All participants provided written informed consent before being enrolled in the study. The antibody test used has a sensitivity of 94% in children, and a specificity of 99.2%. In June 2020, 98 / 2473 (4.0%) of subjects showed as seropositive, compared to 154 / 2500 (6.2%) in October 2021, 17.3% (426 / 2453) in March 2021, 48.5% (910 / 1876) in November 2021, and 94.5% (2008 / 2125) in June 2022. Given the diagnostic test characteristics, absolute bias can be expected to be less than 1% in the range of 0% - 26.5% disease prevalence, and less than 2% for disease prevalence of up to 41.2%. These results imply that reported seroprevalence estimates based on a naive logistic approach are likely relatively unbiased for the first 3 rounds of Ciao Corona antibody testing, but that after that any seroprevalence estimates that do not adjust for test characteristics are likely quite biased. According to the formulae given above, the largest bias occurs at disease prevalence near 100%, where the bias is approximately 6%. We corrected the seroprevalence estimates using a Bayesian hierarchical model approach in all rounds of testing.

4 Discussion

I have demonstrated that bias in prevalence estimates can be higher than desired when using a naive approach of calculation based on standard logistic regression, even if sensitivity and specificity are 90% or higher. Further, I have derived a range of disease prevalence values for which the naive approach gives reasonably unbiased prevalence estimates. Taken together, the results emphasize the necessity to not simply report raw proportions of positive tests, even if those are adjusted for demographic characteristics using logistic regression. Since disease prevalence is of course not known precisely prior to study conduct, the most straightforward approach is then to plan statistical methods so that sensitivity and specificity are accounted for. Care should also be taken in reading publications reporting (sero)prevalence estimates to insure that suitable

statistical methods have been used.

These results are based on the definitions of sensitivity and specificity only and require no complicated derivations. I have not adjusted for demographic characteristics, such as age and gender, or used weighting to approximate the target population, as is typical in surveys of disease prevalence. However, such adjustment cannot alleviate any general concerns of bias as presented here. The bias demontrated here is also an average bias, and may vary more or less depending on the size of the sample. Additionally, the results do not account for other possible issues with a diagnostic test (13–15), that can often not be corrected with statistical methods.

The question remains as to how best to account for diagnostic test sensitivity and specificity when estimating disease prevalence. A nice outline of some appropriate methods along with implementation in R (16) code is given by (8,17,18). To calculate corrected confidence intervals for prevalence, the R package `bootComb` (19) is available, while Bayesian methods are available in `prevalence` (20). Adjusting for demographic factors, or application of post-stratification weights may unfortunately need to be done without the use of such prepackaged code. Collaboration with experienced statisticians is invaluable in insuring that correct analysis techniques are used so that unbiased prevalence estimates can be reported.

Conflict of Interest statement: No conflicts of interest have been declared.

Funding statement: The Ciao Corona study, used in our example, is part of Corona Immunitas research network, coordinated by the Swiss School of Public Health (SSPH+), and funded by fundraising of SSPH+ that includes funds of the Swiss Federal Office of Public Health and private funders (ethical guidelines for funding stated by SSPH+ will be respected), by funds of the Cantons of Switzerland (Vaud, Zurich, and Basel) and by institutional funds of the Universities. Additional funding, specific to this study is available from the University of Zurich Foundation. No additional funding was acquired for the analysis presented here.

Contributions: SRH initiated the analysis, developed the methodology, performed the statistical analysis, and wrote the manuscript. No others meeting criteria for authorship have been omitted.

Acknowledgements: Thank you to Julia Braun, Thomas Radtke, and Milo Puhan for their critical comments.

References

1. Bergeri MAW I AND Whelan. Global SARS-CoV-2 seroprevalence from January 2020 to April 2022: A systematic review and meta-analysis of standardized population-based studies. PLOS Medicine. 2022 Nov;19(11):1–24.
2. Rogan W, Gladen B. Estimating prevalence from the results of a screening test. American Journal of Epidemiology. 1978;107(41).
3. Lew R, Levy P. Estimation of prevalence on the basis of screening tests. Statistics in Medicine. 1989;8:1225–30.
4. Lang Z, Reiczigel J. Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. Preventive Veterinary Medicine. 2014;113(1):13–22.
5. Joseph L, Gyorkos T, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. American Journal of Epidemiology. 1995;141(3):263–72.
6. Berkvens D, Speybroeck N, Praet N, Adel A, Lesaffre E. Estimating disease prevalence in a Bayesian framework using probabilistic constraints. Epidemiology. 2006;17(2):145–53.
7. Gelman A, Carpenter B. Bayesian analysis of tests with unknown specificity and sensitivity. JRSS Series C: Applied Statistics. 2020;
8. Flor M, Selhorst MWT, Müller-Graf C, Greiner M. Comparison of Bayesian and frequentist methods for prevalence estimation under misclassification. BMC Public Health. 2020;
9. Altman D, Bland J. Statistics notes: Diagnostic tests 1: Sensitivity and specificity. BMJ. 1994;308(1552).
10. Pepe M. The statistical evaluation of medical tests for classification and prediction. OUP; 2003.
11. Lewis M. Bayes' theorem and covid-19 testing [Internet]. Significance Magazine blog; 2020. Available from: https://www.significancemagazine.com/science/660-bayes-theorem-and-covid-19-testing
12. Haile S, Raineri A, Rueegg S, Radtke T, Ulytė A, Puhan M, et al. Heterogeneous evolution of SARS-CoV-2 seroprevalence in school-age children: Results from the school-based cohort study Ciao Corona in November-December 2021 in the canton of Zurich. Swiss Medical Weekly. 2022;in press.
13. Takahashi S, Greenhouse B, Rodriguez-Barraquer I. Are seroprevalence estimates for severe actue respiratory syndrom2 coronavirus 2 biased? Journal of Infectious Diseases. 2020;222:1772–5.
14. Burgess S, Ponsford M, Gill D. Are we underestimating seroprevalence of SARS-CoV-2? BMJ. 2020;370(m3364).

15. Accorsi E, Qiu X, Rumpler E, Kennedy-Shaffer L, Kahn R, Joshi K, et al. How to detect and reduce potential sources of biases in studies of SARS-CoV-2 and COVID-19. European Journal of Epidemiology. 2021;36:179–96.

16. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available from: https://www.R-project.org/

17. Lewis F, Torgerson P. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. Emerging Themes in Epidemiology. 2012;9(9).

18. Diggle P. Estimating prevalence using an imperfect test. Epidemiology Research International. 2011;608719.

19. Henrion M. bootComb—an R package to derive confidence intervals for combinations of independent parameter estimates. International Journal of Epidemiology. 2021;50(4):1071–6.

20. Devleesschauwer B, Torgerson P, Charlier J, Levecke B, Praet N, Roelandt S, et al. Prevalence: Tools for prevalence assessment studies. [Internet]. 2022. Available from: https://cran.r-project.org/package=prevalence

Tables

Table 1: Typical example of 2x2 table comparing diagnostic test results and disease status.

	Test neg	Test pos	total
Disease neg	724	80	804
Disease pos	20	176	196
total	744	256	1000

Table 2: Estimated bias (in percentage points) for selected combinations of sensitivity, specificity and disease prevalence

Se	Sp	p2	p10	p30	p50	p90	p98
90	99	0.8%	-0.2%	-2.3%	-4.6%	-9.0%	-9.8%
90	95	4.7%	3.2%	0.5%	-2.3%	-8.5%	-9.7%
90	90	9.8%	8.0%	4.0%	-0.4%	-7.6%	-9.6%
95	99	0.9%	0.4%	-0.8%	-2.0%	-4.4%	-4.9%
95	95	4.8%	3.8%	1.8%	0.0%	-4.0%	-4.8%
95	90	9.7%	8.2%	5.9%	2.5%	-3.3%	-4.7%
99	99	1.0%	0.8%	0.4%	0.0%	-0.8%	-1.0%
99	95	4.8%	4.4%	3.1%	2.1%	-0.3%	-0.9%
99	90	9.8%	8.8%	6.5%	4.9%	0.1%	-0.8%

Figures

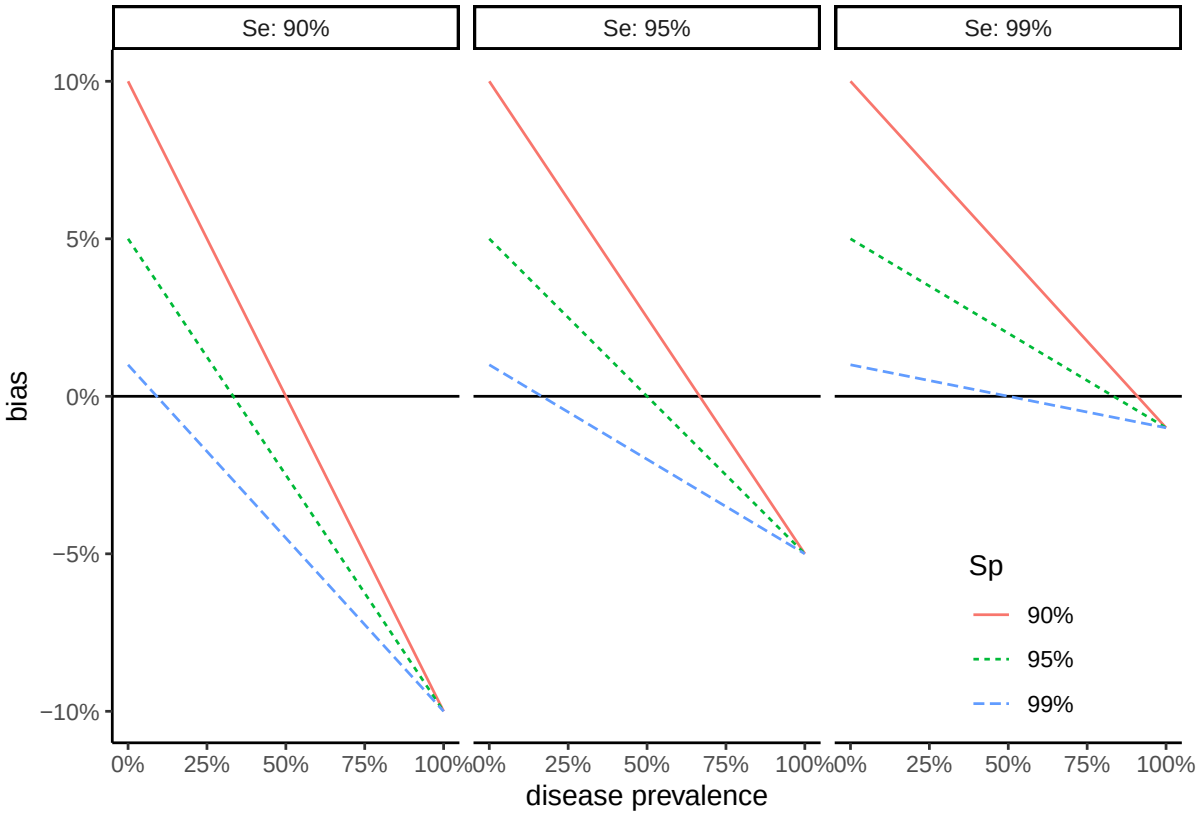

Figure 1: Estimated bias in prevalence estimate for selected combinations of sensitivity, specificity and true disease prevalence

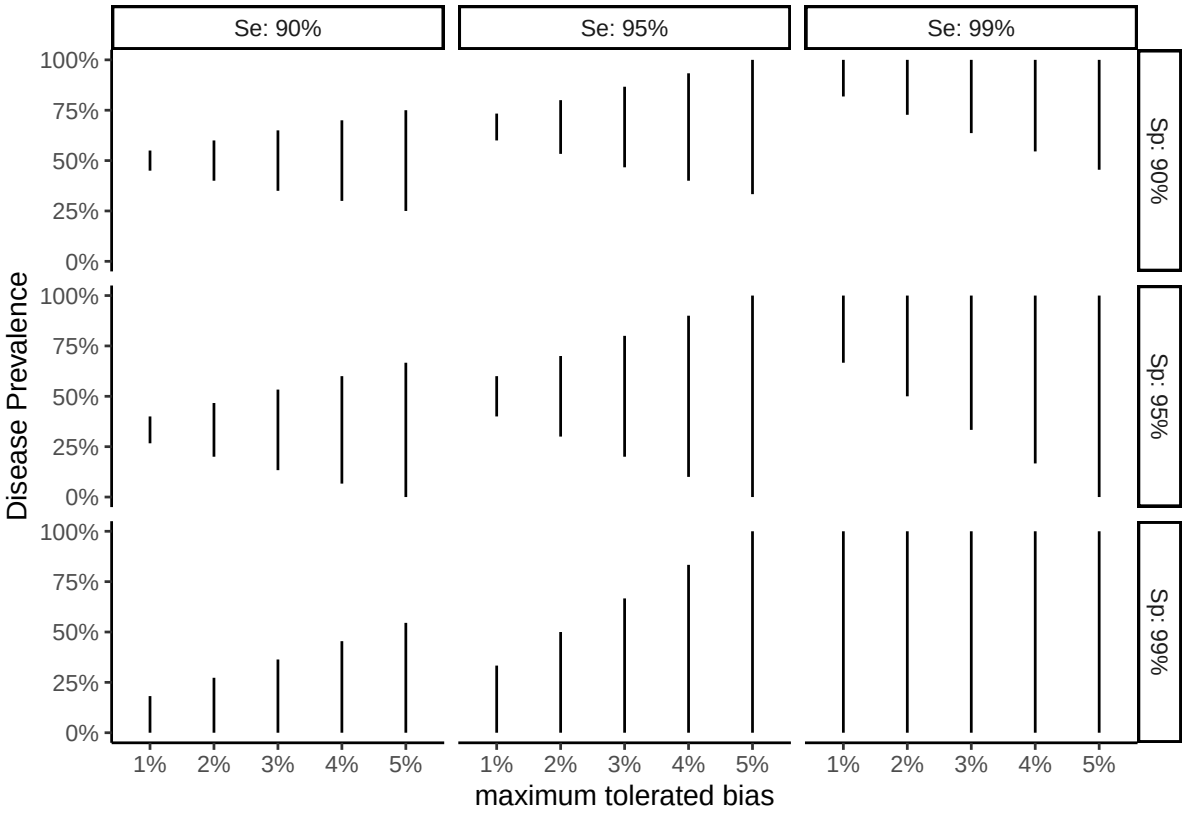

Figure 2: Range of true disease prevalence where the rate of positive tests is a close approximation of disease prevalence, to within maximum absolute tolerated bias