

RESEARCH ARTICLE

Open Access



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

Igor Burstyn^{1,2*} , Neal D. Goldstein² and Paul Gustafson³

Abstract

Background: Despite widespread use, the accuracy of the diagnostic test for SARS-CoV-2 infection is poorly understood. The aim of our work was to better quantify misclassification errors in identification of true cases of COVID-19 and to study the impact of these errors in epidemic curves using publicly available surveillance data from Alberta, Canada and Philadelphia, USA.

Methods: We examined time-series data of laboratory tests for SARS-CoV-2 viral infection, the causal agent for COVID-19, to try to explore, using a Bayesian approach, the sensitivity and specificity of the diagnostic test.

Results: Our analysis revealed that the data were compatible with near-perfect specificity, but it was challenging to gain information about sensitivity. We applied these insights to uncertainty/bias analysis of epidemic curves under the assumptions of both improving and degrading sensitivity. If the sensitivity improved from 60 to 95%, the adjusted epidemic curves likely falls within the 95% confidence intervals of the observed counts. However, bias in the shape and peak of the epidemic curves can be pronounced, if sensitivity either degrades or remains poor in the 60–70% range. In the extreme scenario, hundreds of undiagnosed cases, even among the tested, are possible, potentially leading to further unchecked contagion should these cases not self-isolate.

Conclusion: The best way to better understand bias in the epidemic curves of COVID-19 due to errors in testing is to empirically evaluate misclassification of diagnosis in clinical settings and apply this knowledge to adjustment of epidemic curves.

Keywords: Epidemic, Bias analysis, Uncertainty, Simulation, Under-diagnosis, Sensitivity, Specificity

* Correspondence: ib68@drexel.edu

¹Department of Environmental and Occupational Health, Drexel University Dornsife School of Public Health, 3215 Market St, Philadelphia, PA 19104, USA

²Department of Epidemiology & Biostatistics, Drexel University Dornsife School of Public Health, Philadelphia, PA, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

It is well known that outcome misclassification can bias epidemiologic results yet is infrequently quantified and adjusted for in results. In the context of infectious disease outbreaks, such as during the COVID-19 pandemic of 2019–20, false positive diagnoses may lead to a waste of limited resources, such as testing kits, hospital beds, and absence of the healthcare workforce. On the other hand, false negative diagnoses contribute to uncontrolled spread of contagion, should these cases not self-isolate. In an ongoing epidemic, where test sensitivity (S_n) and specificity (S_p) of case ascertainment are fixed, prevalence of the outcome (infection), determines whether false positives or negatives dominate. For COVID-19, Goldstein & Burstyn show that suboptimal test S_n despite excellent S_p results in an overestimation of cases in the early stages of an outbreak, and substantial underestimation of cases as prevalence increases to levels seen at the time of writing [1]. However, understanding the true scope of the pandemic depends on precise insights into accuracy of laboratory tests used for case confirmation. Undiagnosed cases are of particular concern; they arise from untested persons who may or may not have symptoms (under-ascertainment) and from errors in testing among those selected for the test. We focus on misclassified patients only due to errors in tests that were performed as part of applying the World Health Organization's case definition [2]. Presently, the accuracy of testing for SARS-CoV-2 viral infection, the causal agent for COVID-19, is unknown in Canada and the USA, but globally it is reported that S_p exceeds S_n [3–5].

In a typical scenario, clinical and laboratory validation studies are needed to fully quantify the performance of a diagnostic assay (measured through S_n and S_p). However, during a pandemic, limited resources are likely to be allocated to testing and managing patients, rather than performing the validation work. After all, imperfect testing can still shed a crude light on the scope of the public health emergency. Indeed, counts of observed positive and negative tests can be informative about S_n and S_p , because certain combinations of these parameters are far more likely to be compatible with data and reasonable assertions about true positive tests. In general, more severe cases of disease are expected at the onset of an outbreak (and reflected in tested samples as strong clinical suspicion for the test produces higher likelihood of having the disease) but the overall prevalence in the population would remain low. Then, as the outbreak progresses with more public awareness and consequently both symptomatic and asymptomatic people being tested, the overall prevalence of disease is expected to rapidly increase while the severity of the disease at a population level is tempered. It is reasonable to

expect, as was indeed reported anecdotally early in the COVID-19 outbreak, for laboratory tests to be inaccurate, because the virus itself and its unique identifying features exploited in the test are themselves uncertain, and laboratory procedures can contain errors ahead of standardization and regulatory approval. Again, anecdotally, S_n was supposed to be worse than S_p , which is congruent with reports of early diagnostic tests from China [4, 5], with both S_n and S_p improving as the laboratories around the world rushed to perfect testing [6–8] to approach the performance seen in tests for similar viruses [9, 10]. Using publicly available time-series data of laboratory testing results for SARS-CoV-2 and our prior knowledge of infectious disease outbreaks, we may be able to gain insights into the true accuracy of the diagnostic assay.

Thus, we pursued two specific aims: (a) to develop a Bayesian method to attempt to learn from publicly available time-series of COVID-19 testing about S_n and S_p of the laboratory tests and (b) to conduct a Monte Carlo (probabilistic) sensitivity analysis of the impact of the plausible extent of this misclassification on bias in epidemic curves.

Methods

Sharing

Data and methods can be accessed at <https://github.com/paulgstf/misclass-covid-19-testing>; data are also displayed in Appendix A in the Supplemental Material.

Data

We digitized data released by the Canadian province of Alberta on 3/28/2020 from their “Fig. 6: People tested for COVID-19 in Alberta by day” under “Laboratory testing” tab [11]. Samples (e.g., nasopharyngeal (NP) swab; bronchial wash) undergo nucleic acid testing (NAT) that use primers/probes targeting the E (envelope protein) (Corman et al. 2020) and RdRp (RNA-dependent RNA polymerase) (qualitative detection method developed at ProvLab of Alberta) genes of the COVID-19 virus. The data were digitized as shown in Table A1 of Appendix A in the Supplemental Material. The relevant data notes are reproduced in full here:

“Data sources: The Provincial Surveillance Information system (PSI) is a laboratory surveillance system which receives positive results for all Notifiable Diseases and diseases under laboratory surveillance from Alberta Precision Labs (APL). The system also receives negative results for a subset of organisms such as COVID-19. ... Disclaimer: The content and format of this report are subject to change. Cases are under investigation and numbers may fluctuate as cases are resolved. Data included in the

interactive data application are up-to-date as of midday of the date of posting.”

Data from the city of Philadelphia were obtained on 03/31/2020 [12]. It was indicated that “test results might take several days to process.” Most testing is PCR-based with samples collected from an NP swab, performed at one of the three labs (State Public Health, Labcorps, Quest). In addition, some hospitals perform this test using ‘in-house’ PCR methods. There is a perception (but no empirical data available to us) that S_n is around 0.7 and there are reports of false negatives based on clinical features of patients consistent with COVID-19 disease. Issues arise from problems with specimen collection and timing of the collection, in addition to test performance characteristics. The data were digitized as shown in Table A2 in Appendix A in the Supplemental Material.

Bayesian method to infer test sensitivity and specificity

A brief description of the modelling strategy follows here, with full details of both the model and its implementation given in Appendix B in the Supplemental Material. Both daily prevalence of infection in the testing pool and daily test sensitivity are modelled as piecewise-linear on a small number of adjacent time intervals (four intervals of equal width, in both examples), with the interval endpoints referred to as “knots” (hence there are five knots, in both examples). The prior distribution for prevalence is constructed by specifying lower and upper bounds for prevalence at each knot, with a uniform distribution in between these bounds. The prior distribution of sensitivity is constructed similarly, but with a modification to encourage more smoothness in the variation over time (see Appendix B in the Supplemental Material for full details). The test specificity is considered constant over time, with a uniform prior distribution between specified lower and upper bounds.

With the above specification, a posterior distribution ensues for all the unknown parameters and latent variables given the observed data, i.e., given the daily counts of negative and positive test results. This distribution describes knowledge of prevalence, sensitivity, specificity, and the time-series of the latent Y_t , the number of truly positive among those tested on the t -th day. Thus, we learn the posterior distribution of the Y_t time-series, giving an adjusted series for the number of true positives in the testing pool, along with an indication of uncertainty.

As discussed at more length in Appendix B in the Supplemental Material, this model formulation neither rules in, nor rules out, learning about test sensitivity and specificity from the reported data. Particularly in a high specificity regime, the problem of separating out infection prevalence and test sensitivity is mathematically

challenging. The data directly inform only the product of prevalence and sensitivity. Trying to separate the two can be regarded statistically as an “unidentified” problem (while mathematicians might speak of an ill-posed inverse problem, or engineers might refer to a “blind source separation” challenge). However, some circumstances might be more amenable to some degree of separation. In particular, with piecewise-linear structure for sensitivity and prevalence, strong quadratic patterns in the observed data, if present, could be particularly helpful in guiding separation. On the other hand, if little or no separation can be achieved, the analysis will naturally revert back to a sensitivity analysis, with the a priori uncertainty about test sensitivity and infection prevalence being acknowledged.

Some of the more reliable PCR-based assays can achieve near-perfect S_p and S_n of around 0.95 [3, 6–8]. We expected S_p to be high and selected a time-invariant uniform prior bounded by 0.95 and 1. However, early in the COVID-19 outbreak problems with the sensitivity of the diagnostic test were widely reported owing to specimen collection and reagent preparation, but not quantified. Based on these reports, we posited a lower bound on prior S_n of 0.6 and an upper bound of 0.9. We cannot justify a higher lower bound on S_n , since obtaining a sample is challenging as the virus may not be detectable in the cultured area based on timing of infection, despite replicating in other parts of the respiratory tract. Also, known variation in testing strategy over time could drive variation in S_n over time. Consequently, we adopted a flexible data-driven approach by allowing sensitivity to change over time, within the specified range (see Appendix B in the Supplemental Material). The prevalence of truly infected among those tested likely changed over time as well -- for example due to prioritization of testing based on age, occupation, and morbidity [13] -- but this is difficult to quantify, as it differs from population prevalence of infected that would be “seen” by a random sample (governed by known population dynamics models). Thus, our model also allows this prevalence to vary over time across a broad range.

Monte Carlo (probabilistic) uncertainty/bias analysis of epidemic curves

We next examined how much more we could have learnt from epidemic curves if we knew sensitivity of laboratory testing. To do so, we applied insights into the plausible extent of sensitivity and specificity to recalculate epidemic curves for COVID-19 in Alberta, Canada. Data on observed counts versus presumed incident dates (“date reported to Alberta Health”) was obtained on 3/28/2020 from their “Fig. 3: COVID-19 cases in Alberta by day” under “Case counts” tab [11]. The count of cases is shown in Table A1 as C_t and they are

matched to dates t (same as dates of laboratory tests). We also repeated these calculations with data available for City of Philadelphia, under a strong assumption that date of tests is the same as date of onset, i.e. $Y_t^* = C_t^*$. We removed March 30–31, 2020, counts because of a reported delay of several days in laboratory tests.

For each observed count of incident cases C_t^* , we estimated true counts $C_t = C_t^* / \widetilde{Sn}$ under the assumption that specificity is indistinguishable from perfect. Here \widetilde{Sn} is the assumed sensitivity for the purpose of uncertainty analysis, to not be confused with the posterior distribution of Sn derived in Bayesian modelling. We considered a situation of no time trend in line with above findings, as well as sensitivity either improving (realistic best case), or degrading (pessimistic worst case). We simulated various values of \widetilde{Sn} using *Beta* distribution ranging in means from 0.60 to 0.95, with a fixed standard deviation of 0.05 (parameters set using <https://www.desmos.com/calculator/kx83qio7yl>). It is apparent that epidemic curves generated in this manner will have higher counts than the observed curves, and our main interest is to illustrate how much the underestimation can bias the depiction. Our uncertainty/bias analysis only reflects systematic errors for illustrative purposes and under the common assumptions (and experience) that they dwarf random errors. Computing code in R (R Foundation for Statistical Computing, Vienna, Austria) for the uncertainty analysis is in Appendix C in the Supplemental Material.

Results

Inference about sensitivity and specificity

In both jurisdictions, there is evidence of non-linearity in the observed proportion of positive tests (Fig. 1), justifying our flexible approach to variation of sensitivity and

prevalence that can exhibit a quadratic pattern in observed prevalence between knots. The data in both jurisdictions is consistent with the hypothesis that *the number of truly infected is being under-estimated*, even though observed counts tend to fall within 95% credible intervals of posterior distribution of the counts of true positive tests (Fig. 2). The under-diagnosis is more pronounced when there are both more positive cases and the prevalence of positive tests is higher, i.e. in Philadelphia relative to Alberta. In Philadelphia, the posterior of prevalence was between 5 and 24% (100’s of positive tests a day in late March) but in Alberta, the median of the posterior of prevalence was under 3% (30 to 50 positive tests a day in late March). This is not surprising because the number of false negatives is proportional to observed cases for the same sensitivity. The specificity appears to be high enough for the observed prevalence to produce negligible numbers of false positives, with false negatives dominating. There was clear evidence of shift in posterior distribution of *specificity* from uniform to favouring values > 0.98 (Fig. 3). In Alberta, posterior distribution of Sp was centered on 0.997 (95% credible interval (CrI): 0.993, 0.99995), and in Philadelphia it had a posterior median of 0.984 (95%CrI: 0.954, 0.999). Our analysis indicates that under our models there is little evidence in time-series of laboratory tests about either the time trend or magnitude of *sensitivity* of laboratory tests in either jurisdiction (Fig. 4). Posterior distributions are indistinguishable from the priors, such that we are still left with an impression that sensitivity of COVID-19 tests can be anywhere between 0.6 and 0.9, centered around 0.75. One can speculate on the departure of the posterior distribution from uniform prior given that the posterior appears concentrated somewhat around the prior mean of 0.75 (more lines in Fig. 4 near the mean than the dotted edges that bound the prior). However,

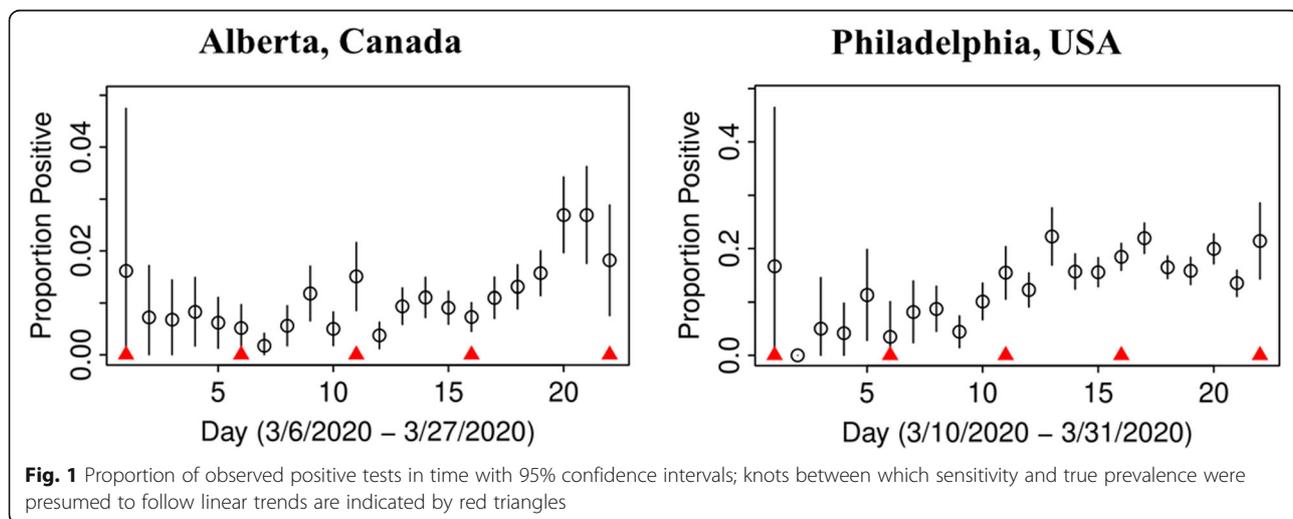
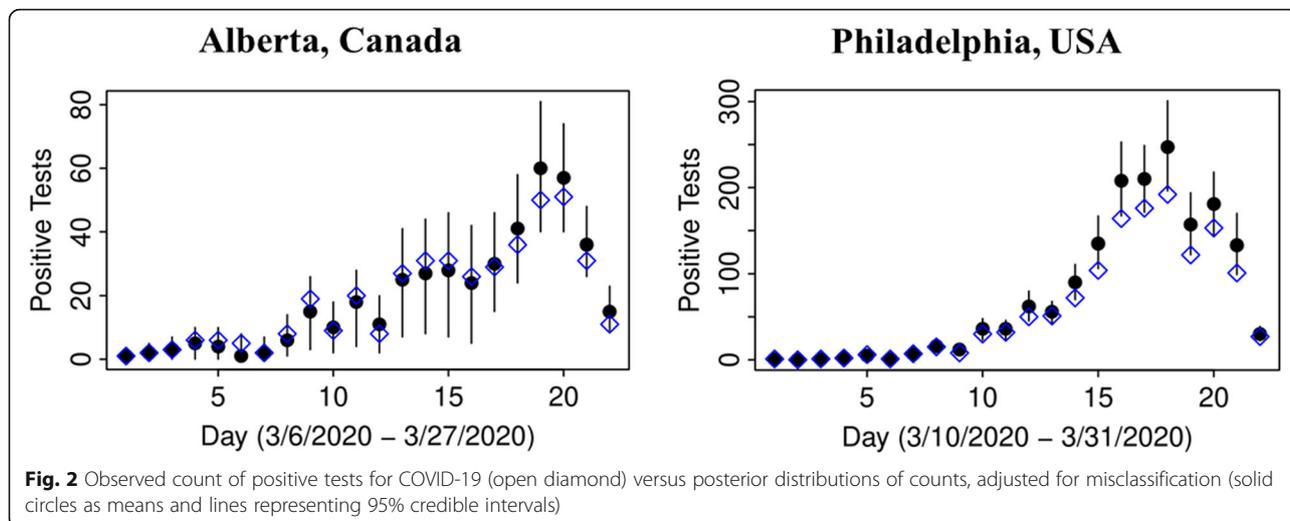


Fig. 1 Proportion of observed positive tests in time with 95% confidence intervals; knots between which sensitivity and true prevalence were presumed to follow linear trends are indicated by red triangles



any such signal is weak and there is no evidence of a time-trend that was favoured by the model.

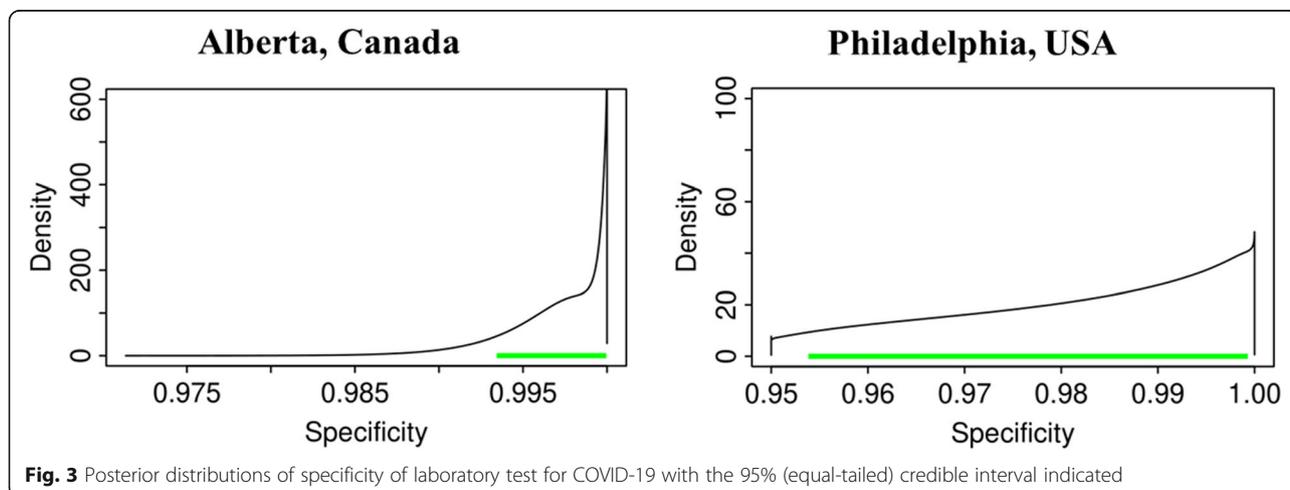
Uncertainty in epidemic curves due to imperfect testing: Alberta, Canada

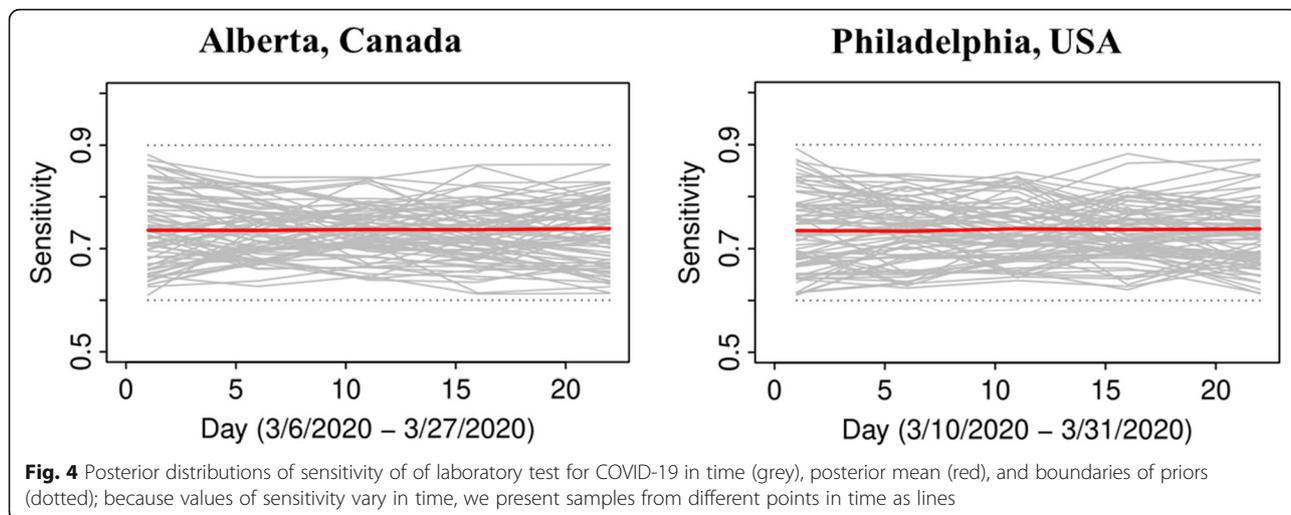
Figure 5 presents the impact on the epidemic curve of degrading sensitivity that is constant in time. As expected, when misclassification errors increase, uncertainty about epidemic curves also increases. There is an under-estimation of incident cases that is more apparent later in the epidemic when the numbers rise. Figure 6 indicates how, as expected, if sensitivity improves over time (green lines), then the true epidemic curve is expected to be flatter than the observed. It also appears that observed and true curves may well fall within the range of 95% confidence intervals around the observed counts (blue lines). If sensitivity decreases over time (brown lines), then the true epidemic curve is expected to be steeper than the observed. In either scenario, there

can be an under-counting of cases by nearly a factor of two, most apparent as the incidence grows, such that on day March 24, 2020 ($t = 19$), there may have been almost 120 cases vs. 62 observed. This is alarming, because misdiagnosed patients can spread infection if they have not self-isolated (perhaps a negative test results provided a false sense of security) and it is impossible to know who they are among thousands of symptomatic persons tested around that time per day (Table A1).

Uncertainty in epidemic curves due to imperfect testing: Philadelphia, USA

Figure 7 presents the impact on the epidemic curve of degrading S_n that is constant in time. As in Alberta, when misclassification errors increase, uncertainty about epidemic curves also increases. It is also apparent that the shape of the epidemic curve, especially when counts are high, can be far steeper than that inferred assuming perfect testing. Figure 8 indicates that if sensitivity

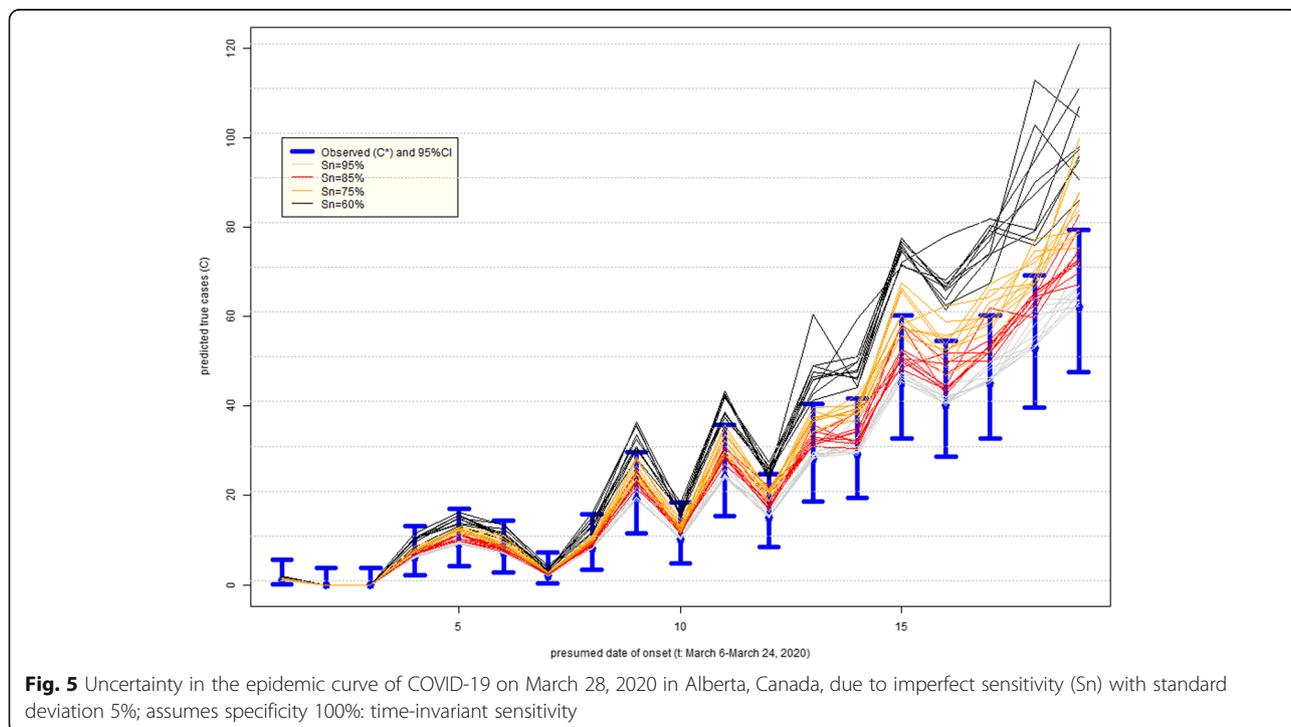




improves over time (green lines), then the true epidemic curve is expected to be practically indistinguishable from the observed one in Philadelphia: e.g. it is within random variation of observed counts represented by 95% confidence intervals (blue lines). This is comforting, because this seems to be the most plausible scenario of improvement in time in quality of testing (identification of truly infected). However, if sensitivity decreases over time (brown lines), then the under-counting of cases by the hundreds in late March 2020 cannot be ruled out. We again have the same concern as for Alberta:

misdiagnosed patients can spread infection unimpeded and it is impossible to know who they are among the hundreds of symptomatic persons tested in late March 2020 (Table A2).

In all examined scenarios, in both Alberta and Philadelphia, the lack of sensitivity in testing seems to matter far less when the observed counts are low early in the epidemic. The gap between observed and adjusted counts grows as the number of observed cases increases. This reinforces the importance of early testing, at least with respect to describing the time-course of the



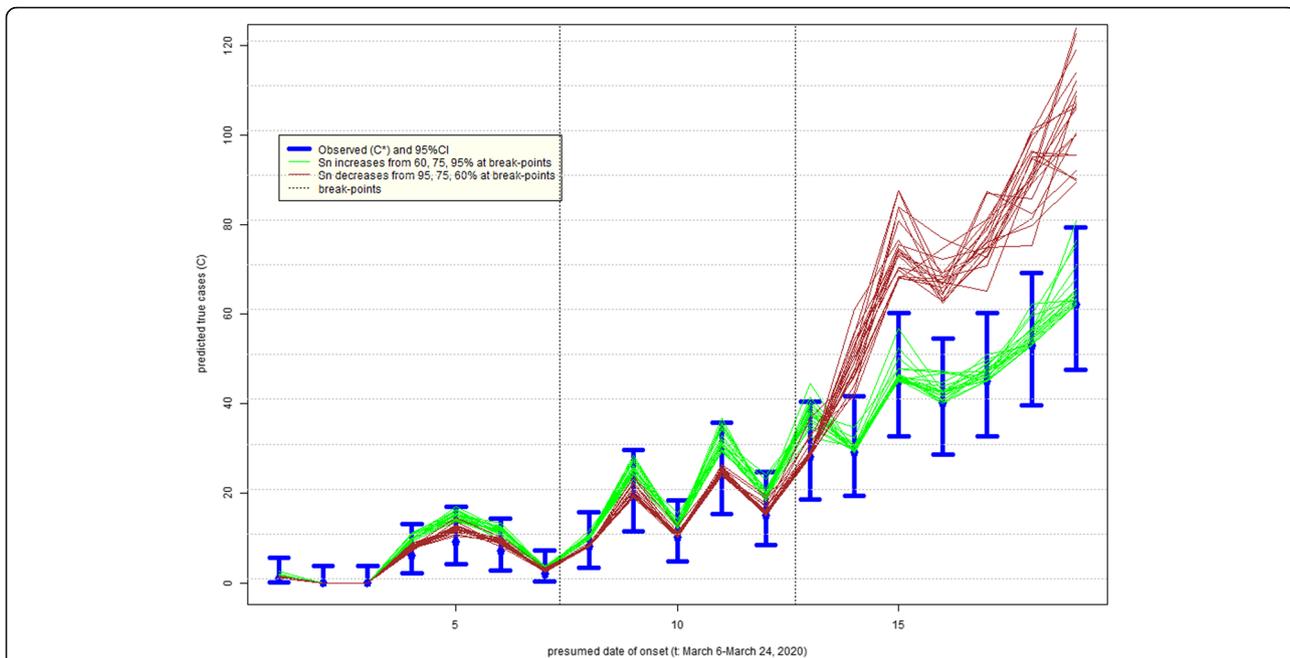


Fig. 6 Uncertainty in the epidemic curve of COVID-19 on March 28, 2020 in Alberta, Canada, due to imperfect sensitivity (S_n) with standard deviation 5%; assumes specificity 100%: increasing or decreasing sensitivity in time

epidemic, even with imperfect technology (Goldstein & Burstyn, 2020).

Discussion

Given the current uncertainty in the accuracy of the SARS-CoV-2 diagnostic assays, we tried to learn about

sensitivity and specificity using the time-series of laboratory tests and time trends in time test results. Although we are confident that *typical specificity exceeds 0.98*, there is *very little learning about sensitivity* from prior to posterior. However, it is important to not generalize this lack of learning about sensitivity, because it can occur

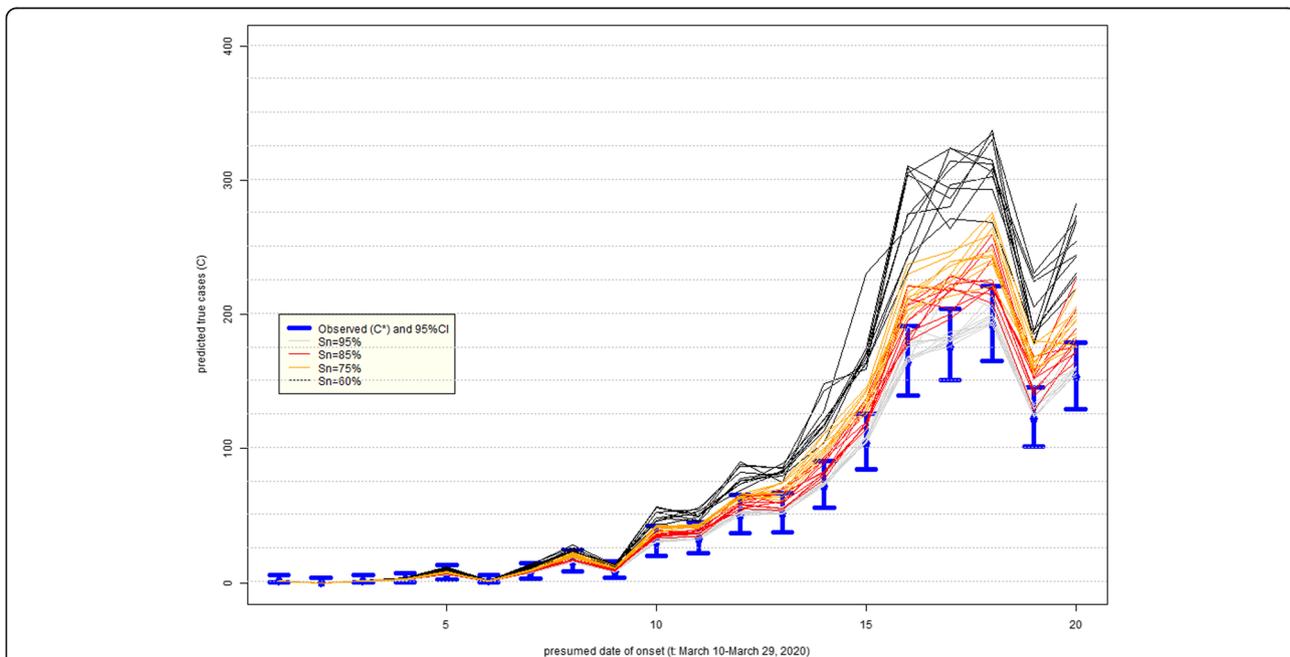
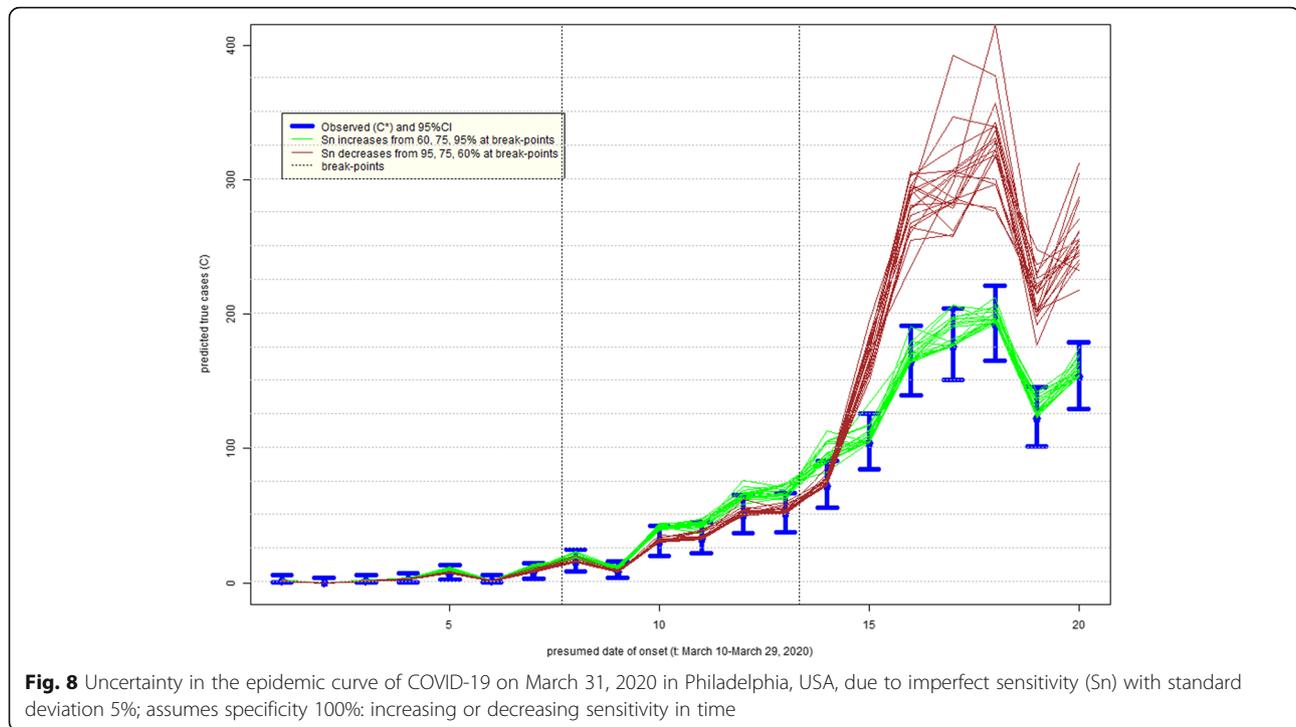


Fig. 7 Uncertainty in the epidemic curve of COVID-19 on March 31, 2020 in Philadelphia, USA, due to imperfect sensitivity (S_n) with standard deviation 5%; assumes specificity 100%: time-invariant sensitivity



when stronger priors on prevalence are justified and/or when there are more pronounced trends in prevalence of positive tests. We therefore encourage every jurisdiction with suitable data to attempt to gain insights into accuracy of tests using our method: now that the method to do so exists, it is simpler and cheaper than laboratory and clinical validation studies. However, validation studies, with approaches like the one illustrated in Burstyn et al., are still the most reliable means of determining accuracy of a diagnostic test [14].

Knowing sensitivity and specificity is important as demonstrated in uncertainty/bias analysis of impact on epidemic curves under some optimistic assumptions of near-perfect specificity and reasonable range of sensitivity. The observed epidemic curves may bias estimates of the effective reproduction number (R_e) and magnitude of the epidemic (peak) in unpredictable directions. This may also have implications for understanding the proportion of the population non-susceptible to COVID-19. As researchers attempt to develop pharmaceutical prophylaxis (i.e., a vaccine) combined with a greater number of people recovering from SARS-CoV-2 infection, having insight into the herd threshold will be important for resolving current and future outbreaks. Calculations such as the basic and effective reproductive number, and the herd threshold depend upon the accuracy of surveillance data described in the epidemic curves.

As the title suggests, we view the lab time-series and the epidemic curve as two distinct entities: Figs. 1

through 4 are based on the former. This distinction is important to stress, because a lot of the public-facing dashboards etc. are plotting new cases by report date instead of, or in addition to, by symptom onset date; both are commonly labelled as “epidemic curves” while strictly only the latter should be referred to as such. We emphasise this distinction by adopting different notation to count positive test results on t^{th} day t as Y_t vs. incident cases on t^{th} day C_t . Future work is envisioned which links Y_t and C_t , so that joint inference could be undertaken for data from jurisdictions which report both series. In situations where only the lab-testing series is available, external prior knowledge could be used to describe implications for the epidemic curve. As an example, while the modelling in Dehning et al. [15] is in a very different direction, one component of their model is an informative prior distribution on the reporting delay between infection date and lab report date.

Limitations of our approach include the dynamic nature of data that changes daily and may not be perfectly aligned in time due to batch testing. There are some discrepancies in the data that should be resolved in time, like fewer cases tested positive than there are in epidemic curve in Alberta, *but the urgency of the current situation justifies doing our best with what we have now*. We also make some strong ad hoc assumptions about breakpoints in segmented regression of time-trends in sensitivity and prevalence, further assuming that the same breakpoints are suitable for trends in both parameters. Although not as much of an issue based on our

analysis, we do need to consider imperfect specificity, creating false positives, albeit nowhere near the magnitude of false negatives in the middle of an outbreak. This results in wasted resources. In ideal circumstances we employ a two-stage test: a highly sensitive serological assay that if positive triggers a PCR-based assay. Two-stage tests would resolve a lot of uncertainty and speculation over a single PCR test combined with signs and symptoms. Indeed, this is the model used for diagnosis of other infectious diseases, such as HIV and Hepatitis C. Our work also only focuses on validity of laboratory tests, not sensitivity and specificity of the entire process of identification of cases that involves selection for testing via a procedure that is designed to induce systemic biases relative to the population.

Conclusions

We conclude that it is of paramount importance to validate laboratory tests and to share this knowledge, especially as the epidemic matures into its full force. Insights into ascertainment bias by which people are selected for tests and are then used to estimate epidemic curves are likewise important to obtain and quantify. Quantification of these sources of misclassification and bias can lead to adjusted analyses of epidemic curves that can help make more appropriate public health policies.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12874-020-01037-4>.

Additional file 1.

Abbreviations

Y_t^+ : Count of persons who tested positive at time t ; n_t : Count of persons tested at time t (observed from surveillance data); Y_t : True count of persons who tested positive at time t (latent); C_t^+ : Count of persons having onset of symptoms at time t and who have tested positive; C_t : True count of persons having onset of symptoms at time t and who have tested positive (latent); Sn_t : Sensitivity of test $P(Y_t^+ = 1 | Y_t = 1)$ at time t (subscript t is suppressed for simplicity in text); Sp_t : Specificity of test $P(Y_t^+ = 0 | Y_t = 0)$ at time t (subscript t is suppressed for simplicity in text); \widehat{Sn} : Sensitivity of ascertainment of incident case, $P(C_t^+ = 1 | C_t = 1)$; \widehat{Sp} : Specificity of ascertainment of incident case, $P(C_t^+ = 0 | C_t = 0)$

Acknowledgements

The authors thank Isaac R. Burstyn and Marguerite R. Burstyn for their diligence in perfect extraction of Alberta data from the on-line charts. IB also thanks James L.B. Burstyn for not teething every night during preparation of this manuscript. Nicola Chery and Esther Chernak earned our gratitude by assisting with insights on testing in Alberta and Philadelphia, respectively.

Authors' contributions

All authors contributed to conceptualization and writing of the manuscript. PG led development of statistical methodology, with input from IB and NDG; IB led the effort to acquire data. The author(s) read and approved the final manuscript".

Funding

Research reported in this publication was partly supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number K01AI143356 (to NDG).

Availability of data and materials

Data and methods can be accessed at <https://github.com/paulgstf/misclass-covid-19-testing>; data are also displayed in Appendix A in the Supplemental Material.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Environmental and Occupational Health, Drexel University Dornsife School of Public Health, 3215 Market St, Philadelphia, PA 19104, USA. ²Department of Epidemiology & Biostatistics, Drexel University Dornsife School of Public Health, Philadelphia, PA, USA. ³Department of Statistics, University of British Columbia, Vancouver, BC, Canada.

Received: 24 April 2020 Accepted: 27 May 2020

Published online: 06 June 2020

References

- Goldstein ND, Burstyn I. On the importance of early testing even when imperfect in a pandemic such as COVID-19: OSF Preprints; 2020. <https://doi.org/10.31219/osf.io/9pz4d>.
- World Health Organization. Global surveillance for COVID-19 caused by human infection with COVID-19 virus Interim guidance 20 March 2020. Available at: <https://apps.who.int/iris/bitstream/handle/10665/331506/WHO-2019-nCoV-SurveillanceGuidance-2020.6-eng.pdf>.
- Saw Swee Hock School of Public Health COVID-19 Science Report. Available at: <https://sph.nus.edu.sg/wp-content/uploads/2020/03/COVID-19-Science-Report-Diagnostics-13-Mar.pdf>.
- Fang Y, Zhang H1, Xie J. sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 2020;19:200432.
- Ai T, Yang Z1, Hou H. correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020;26:200642.
- Konrad R, Eberle U, Dangel A, et al. Rapid establishment of laboratory diagnostics for the novel coronavirus SARS-CoV-2 in Bavaria, Germany, February 2020. *Euro Surveill*. 2020;25(9).
- Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*. 2020;25(3).
- US Food and Drug Administration. Accelerated emergency use authorization (EUA) summary COVID-19 RT-PCR test (Laboratory Corporation of America). Available at: <https://www.fda.gov/media/136151/download>.
- Binsaeed AA, Al-Khedhairi AA, Mandil AM, et al. A validation study comparing the sensitivity and specificity of the new Dr. KSU H1N1 RT-PCR kit with real-time RT-PCR for diagnosing influenza A (H1N1). *Ann Saudi Med*. 2011;31(4):351–5.
- Merckx J, Wali R1, Schiller I, et al. diagnostic accuracy of novel and traditional rapid tests for influenza infection compared with reverse transcriptase polymerase chain reaction: a systematic review and meta-analysis. *Ann Intern Med*. 2017;167(6):394–409.
- Government of Alberta. COVID-19 in Alberta. Available at: <https://covid19stats.alberta.ca/>. Accessed 28 Mar 2020.
- City of Philadelphia. Coronavirus Disease 2019 (COVID-19). Available at: <https://www.phila.gov/programs/coronavirus-disease-2019-covid-19/the-citys-response/monitoring-and-testing/>. Accessed 31 Mar 2020.
- Centers for Disease Control and Prevention. Evaluating and Testing Persons for Coronavirus Disease 2019 (COVID-19). Available at: <https://www.cdc.gov/coronavirus/2019-nCoV/hcp/clinical-criteria.html>. Accessed 7 Apr 2020.
- Burstyn I, Kapur N, Shalabay C, Bamforth F, Wild TC, Liu J, LeGatt D. Evaluation of the accuracy of self-reported smoking in pregnancy when the

biomarker level in an active smoker is uncertain. *Nicotine Tob Res.* 2009; 11(6):670–8.

15. Dehning J, Zierenberg J, Spitzner PF, Wibral M, Neto JP, Wilczek M, Priesemann V. Inferring change points in the COVID-19 spreading reveals the effectiveness of interventions. *arXiv.* 2004:01105v3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

