COVID-19 Case study in South Korea, implementing Bayesian method from 'Towards reduction in bias in epidemic curves due to outcome misclassification through Bayesian analysis of time-series of laboratory test results'

Dongyoung Kim

Background: Using Rapid COVID-19 Antigen Test is controversial due to its sensitivity issue within medical society in South Korea. The purpose of this study is to see the impact of poor sensitivity in epidemic curves.

Methods: The Bayesian method of Dr. Paul Gustafson is used. Time-series data of Ministry of Health and Welfare of South Korea for COVID-19 test status were examined through Bayesian analysis to infer the sensitivity and specificity. Different sensitivity values are simulated to see the impact of misclassification errors through Monte Carlo simulation. Clinically evaluated sensitivity values are used in the study.

Results: Bayesian analysis about the sensitivity and specificity indicated that the value of specificity is almost 100%, which is expected from clinical performance results from other studies. On the other hand, sensitivity is inferred to be approximately 60%, which is a value in-between clinical RT-PCR performance (95%) and estimated Ag test performance (40%) from studies conducted in South Korea. Through Monte Carlo sensitivity analysis, the impact of these sensitivity values is examined. However, in most cases the number of test positives is underestimated.

Conclusions: The impact of misclassification due to poor sensitivity was worse than anticipated. Special cares have to be taken in using Ag test.

Introduction

There has been a study that tried to reduce bias in epidemic curves through Bayesian analysis [1]. Bayesian analysis to infer test sensitivity (Sn) and specificity (Sp), and Monte Carlo uncertainty analysis of epidemic curves were the main two themes. The study concluded that empirical evaluation of misclassification of diagnosis in clinical settings is necessary. Since the publication of this study in 2020, numerous empirical evaluations regarding Sn have been conducted worldwide including South Korea [2] [3] [4] [5]. In South Korea, there has been a controversy in using Rapid STANDARD Q COVID-19 Antigen Test (hereafter abbreviated as Ag Test) because of its Sn credibility. Therefore, several clinical evaluations of Ag Test have been performed [6] [7]. The range of evaluated Sn values varies widely from lowest 17.5% to highest 90%. Nevertheless, Korea Disease Control and Prevention Agency (KDCA) announced that Ag Test is approved by Ministry of Food and Drug Safety (MFDS) with 90% Sn test result [6]. However, Korean Medical Association (KMA) urged not to use Ag test targeting normal citizens who have no symptoms, since the clinical validation is not accurate and has limitations. Furthermore, there was a recent study that validated clinical performance of Ag test in a large-scale setting with 680 specimens by Korean Society for Laboratory Medicine (KSLM). The Sn value was estimated as 41.8% based on the proportion of the observed threshold cycle (Ct) value distribution in the Korean population [8].

Thus, I borrowed the methodology of the previous study [1] to examine two topics. First, Sn and Sp are inferred with the generative model, and the Sn valued estimated by KSLM is used as a prior knowledge [8]. Second, epidemic curve simulation was conducted with several Sn values, which are clinically validated in South Korea [4] [6] [8]. This is to see the impact of misclassification errors in epidemic curves in the worst case scenario.

Only evaluations conducted in South Korea were used because COVID-19 Ag Test kits that are used in Korea are mostly produced by domestic companies.

Methods

Bayesian analysis

Methods used in this study are created by Dr. Paul Gustafson. Computing code in R language is accessible at https://github.com/paulgstf/misclass-covid-19-testing. Here is a basic concept of the statistical model used in the study [1] based on my understanding (Fig. 1, 2).

	Test positive Y_t^*	Test negative $n_t - Y_t^*$
Actual positive $Y_t \sim Bin(n_t, r_t)$	True positive $Y_{A,t} \sim Bin(Y_t, s_{n_t})$	False negative $Y_t - Y_{A,t}$
Actual negative $n_t - Y_t$	False positive $Y_{B,t} \sim Bin(n_t - Y_t, 1 - s_p)$	True negative $n_t - Y_t - Y_{B,t}$

Fig. 1 Contingency table of variables with its priors

Hierarchical Bayesian model: $\mathrm{f}(Y_t^*, Y_{A,t}, Y_{B,t}, Y_t | r_t,\ Sn_t,\ Sp) = \mathrm{f}(Y_t^* | Y_{A,t}, Y_{B,t})\mathrm{f}(Y_{A,t}, Y_{B,t} | Y_t, Sn_t, Sp)\mathrm{f}(Y_t | r_t)$

(Paul Gustafson, 2020, Appendix B)

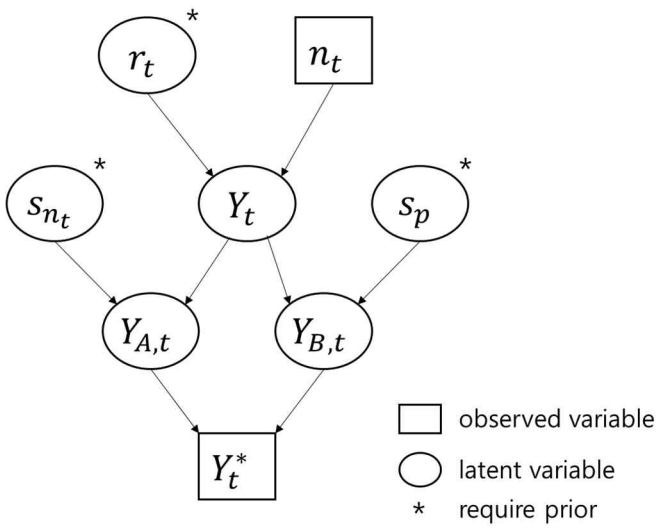

Fig. 2 Diagram of generative model to estimate the number of actual positives

The prior distributions for prevalence and Sp are assumed to follow the same upper and lower bounds written in Dr. Paul Gustafson's coding. However, the prior distribution for Sn was modified. Lower bound 0.4 is assigned to reflect laboratory test results as a prior knowledge [8]

Data

Data access

COVID-19 testing data is collected by Ministry of Health and Welfare. The data is publicly available at Korea Public Data Portal website (https://www.data.go.kr/en) upon permission. Authorization was not granted until making a call to a government officer. Upon receiving the API key, COVID-19 testing data with 14 parameters were accessed (Fig. 3). Raw data of negative and positive test numbers were in the form of cumulative data, so I transformed them into daily format. When compared to the number of tests each day, the sum of negative tests and positive tests are the same, which confirms that the data is correct.

Fig. 3 COVID-19 testing data OpenAPI access on the left, and the data is put into csv file on the right

Limitation in data

Different from Dr. Paul Gustafson's research data, C_t^* values (true count of persons having onset of symptoms at time t and who have tested positive) were not achievable, but the number of Y_t^* (test positives) were available. Similar to Philadelphia's case, Date of tests is assumed to be the same as date of onset, therefore $Y_t^* = C_t^*$.

In the raw data, daily test counts and completed test counts were both available. Completed test counts are reported daily, but not always accurate due to delays in laboratory tests. However, the completed test results were used in this analysis because the number of test positives and negatives are based off of the completed test results.

Results

Proportion of positive tests and Posterior distribution of positive tests

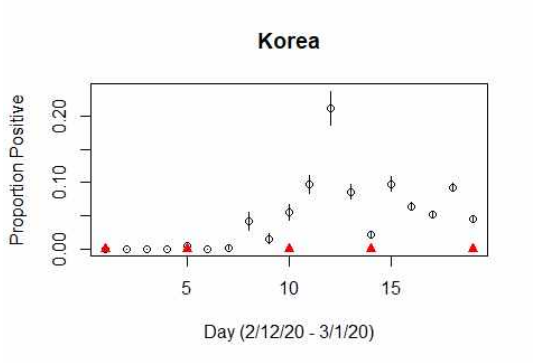

Fig. 4 Proportion of observed positive tests in time with 95% confidence intervals

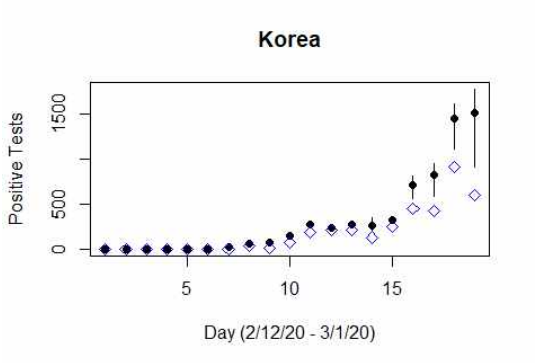

Fig. 5 Observed count of positive tests for COVID-19 versus posterior distributions of counts, adjusted for misclassification

Proportion of positive tests in total tests is not constant, showing non-linear pattern (Fig. 4). Non-linearity of proportion of positive tests proved the assumption of time variant Sn and prevalence. The comparison of the posterior distribution and observed count of positive tests shows that there have been many misclassification cases in early COVID-19 outbreak (Fig. 5). Observed counts do not even fall under 95% credible intervals so that the number of actual positive persons are severely under counted. The extent of underestimation worsens as the number of cases arise.

Sensitivity and specificity inferences

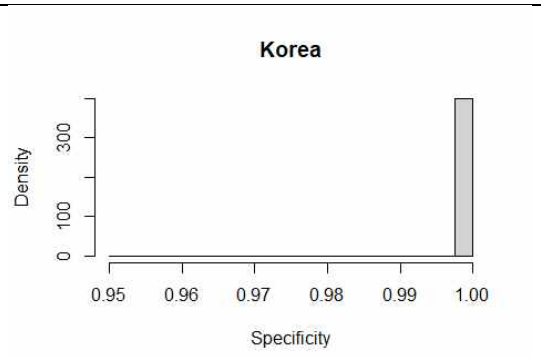

Fig. 6 Posterior distribution of specificity

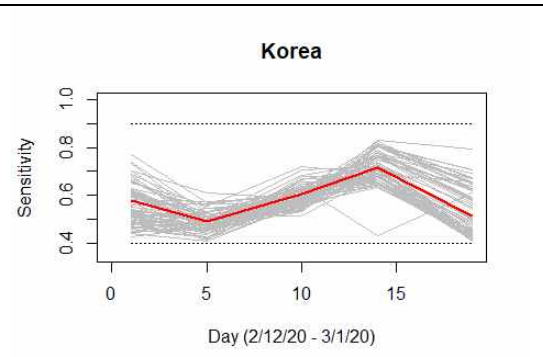

Fig. 7 Posterior distributions of sensitivity of laboratory test for COVID-19 in time (grey), posterior mean (red), and boundaries of priors (dotted)

The posterior distribution of Sp shows that COVID-19 tests are conducted with almost perfect Sp in South Korea (Fig. 6). Which is not surprising because most of the clinical evaluations show that Sp is close to perfection and even rapid Ag test showed 100% Sp [6] [7] [8]. The posterior distribution of Sn fluctuates within the prior

distribution range of 0.4 to 0.9. The average Sn value is 0.58, but since there no clear pattern, it is hard to estimate Sn value.

Uncertainty in epidemic curves due to imperfect testing in South Korea

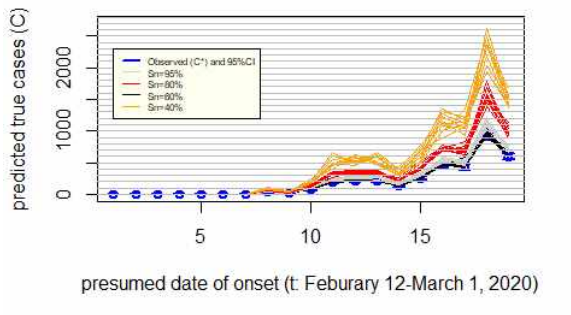

Fig. 8 Uncertainty in the epidemic curve of COVID-19 due to imperfect sensitivity with standard deviation 5%; assumes specificity 100%: time-invariant sensitivity

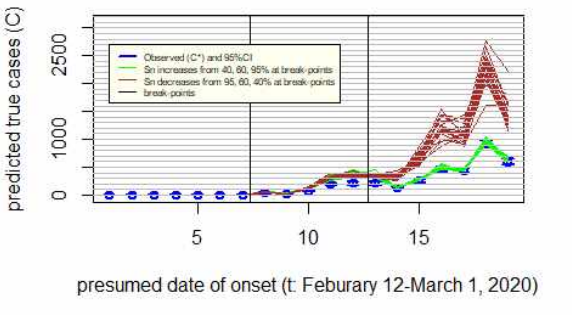

Fig. 9 Uncertainty in the epidemic curve of COVID-19, due to imperfect sensitivity with standard deviation 5%; assumes specificity 100%: increasing or decreasing sensitivity in time

The uncertainty depending on different Sn values is examined in two different ways, constant Sn values and increasing and decreasing Sn values. With constant Sn ranging from 0.4 to 0.95, predicted true cases could reach up to 2,620 cases, which is 2.8 times more than the observed counts (Fig. 8). One of the Sn clinical results was lower than 0.2 [5], which was not used in this study, but the impact of misclassification errors can be hugely excessive. Uncertainty of curves amplifies, while observed cases increase, and thus the number of cases should be more carefully examined as cases surge. When Sn increases from 0.4 to 0.95, the curve is slightly above the 95% confidence intervals of observed cases, however, with decreasing Sn, the curve surges dramatically up to 2,771 cases (observed counts on that day was 909). Understandably, COVID-19 Ag tests are not the majority of tests performed in South Korea. Therefore, the simulation results may seem overemphasized. However, it showcases the danger of selling Ag test kits without proper warning on misclassification errors. Thus, it is important to understand how the epidemic curves can be affected by Sn values with uncertainty.

Conclusions

Sensitivity inference results was approximately 60% during early COVID-19 outbreak using empirically evaluated information. The impact of misclassification due to poor sensitivity was worse than anticipated because the predicted true cases were more than double the number of test positives. Ag tests can be harmful without proper identification on misclassification risks.

Acknowledgements

The aim of this paper is to practice and execute Bayesian analysis created by Dr. Paul Gustafson using South Korea's data. I want to express my gratitude to Dr. Paul Gustafson for sharing his coding publicly because it was a great experience to carry out this Bayesian analysis.

References

[1] 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, US".

[2] "European Centre for Disease Prevention and Control. Options for the use of rapid antigen tests for COVID-19 in the EU/EEA and the UK. Technical Report. ECDC. Stockholm, 2020".

[3] Bianca C, Antonio M, Francesca B "Sensitivity of SARS-CoV-2 Detection With Nasopharyngeal Swabs".

[4] Han Y, Dong J, Seock L, "Usefulness of Sputum and Pharyngeal Swab Specimens for Diagnosis of Coronavirus Disease 2019".

[5] Lauren M, Stephen A, Oliver L, "Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure".

[6] Hyung K, Mi P, Jong L, "Clinical Evaluation of the Rapid STANDARD Q COVID-19 Ag Test for the Screening of Severe Acute Respiratory Syndrome Coronavirus 2".

[7] Sang O, Hyeong J, Eui C, "Clinical Application of the Standard Q COVID-19 Ag Test for the Detection of SARS-CoV-2 Infection".

[8] Jae L, So K, Hee H, "Clinical Performance of the Standard Q COVID-19 Rapid Antigen Test and Simulation of its Real-World Application in Korea".