



HCEO WORKING PAPER SERIES

Working Paper



HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Effect of Universal TB Vaccination and Other Policy-Relevant Factors on the Probability of Patient Death from COVID-19

Amos Golan^{abc1}, Tinatin Mumladze^a, Danielle Wilson^a, Elissa Cohen^a, Troy McGuinness^a,
William Mooney^a and Jisung Moon^a

^aDepartment of Economics, American University, Kreeger Hall, 4400 Massachusetts Ave, NW, Washington DC, 20016, ^bSanta Fe Institute, ^cPembroke College, Oxford

The possibility of reoccurring waves of the novel coronavirus that triggered the 2020 pandemic makes it critical to identify underlying policy-relevant factors that could be leveraged to decrease future COVID-19 death rates. We examined variation in a number of underlying, policy-relevant, country-level factors and COVID-19 death rates across countries. We found three such factors that significantly impact the survival probability of patients infected with COVID-19. In order of impact, these are universal TB (BCG) vaccination, air pollution deaths and a health-related expenditure. We quantify each probability change by age and sex. To deal with small sample size and high correlations, we use an information-theoretic inferential method that also allows us to introduce priors constructed from independent SARS data.

BCG vaccine | coronavirus | COVID-19 | health expenditure | inference | information theory
| policy | pollution level

The authors declare no competing interests

¹To whom correspondence may be addressed. Email: agolan@american.edu.

The possibility of reoccurring waves of the novel coronavirus virus that triggered the 2020 pandemic makes it critical to identify underlying policy-relevant factors that could be leveraged to decrease future COVID-19 death rates. In addition to a patient's age and sex, the conditional probability of recovery also depends on a county's underlying social and environmental factors, many of which are influenced by public policy. Using variation in these factors and in death rate from COVID-19, we sought to identify policy-relevant factors that had a significant impact on the probability of death in COVID-19 infected patients.

Investigation into the impact of underlying factors on COVID-19 death rates is complicated by a lack of data. While millions have been infected and thousands have died from COVID-19, the available patient-level data are quite limited. To deal with small sample size and high correlations, we use an information-theoretic inferential method. This approach also allows us to complement existing data with priors constructed from the death frequency (by age and sex) of individuals who were infected with Severe Acute Respiratory Syndrome (SARS).

A number of recent studies also looked into unconditional frequencies of demographics and correlations among certain health policies, vaccinations and mortality rate of COVID-19 (1-4). Our approach goes beyond the correlation analyses presented in these studies to estimate death probability by age and sex conditional on other major factors.

Results

Using a sample of 485 observations from twenty countries we found that three policy-relevant factors have a significant impact on the survival rate of COVID-19. First and most pronounced is a county's past or present universal TB (BCG) vaccination. Second is the air-pollution death rate in the country. Third is health-related expenditure (the World Health Organization and World Bank indicator 'domestic private health expenditure per capita').

Fig. 1 (Top Left) shows the marginal effects from our information-theoretic, logistic regression for model n (no priors) and model p (with priors). The vertical line is the mean death probability of a 52-year-old male from a country with universal BCG vaccination and other sample values at the mean. The circles on the right/left of that line are the marginal effects of each policy. The marginal effects are the impact of a unit change in a certain covariate (say, age) on the probability of dying. For dummy variables, the marginal effect is

the difference in the average probability of dying among the two groups (say male – female). In terms of lives saved, the universal BCG vaccination saves the most (survival rate increases by about 55 percentage points), then the pollution death rate (survival rate increases by 2.6 percentage points for a unit decrease in the pollution death rate) and then the health indicator (survival rate increases by 0.2. percentage points for a unit increase in the indicator). Being a female, increases the survival rate by approximately 5 percentage points.

Fig. 1 (Top Right) shows the predicted death probability, for each patient. The brown line is from model p. The green dots are from model n. The black dots are the observed values. The patients are sorted from lowest (left) to highest predicted probabilities based on model p. These results show that model p's predictions are more accurate.

Fig. 1 (Bottom Left) shows the priors. Bottom Right shows out-of-sample prediction. Applying both models to a randomly chosen subsample of 200 observations we predict the other 285 observations (using the rule that if the predicted probability of dying is at least 0.5, the predicted outcome is death). The superiority of model p is clear and reaffirms our choice of priors and provides a robustness test of our findings.

The marginal effects and the predictions do not provide the complete story due to the nonlinearity of the inferential models. We therefore conducted a set of comparisons to investigate the impact of the factors across the entire age spectrum, by sex.

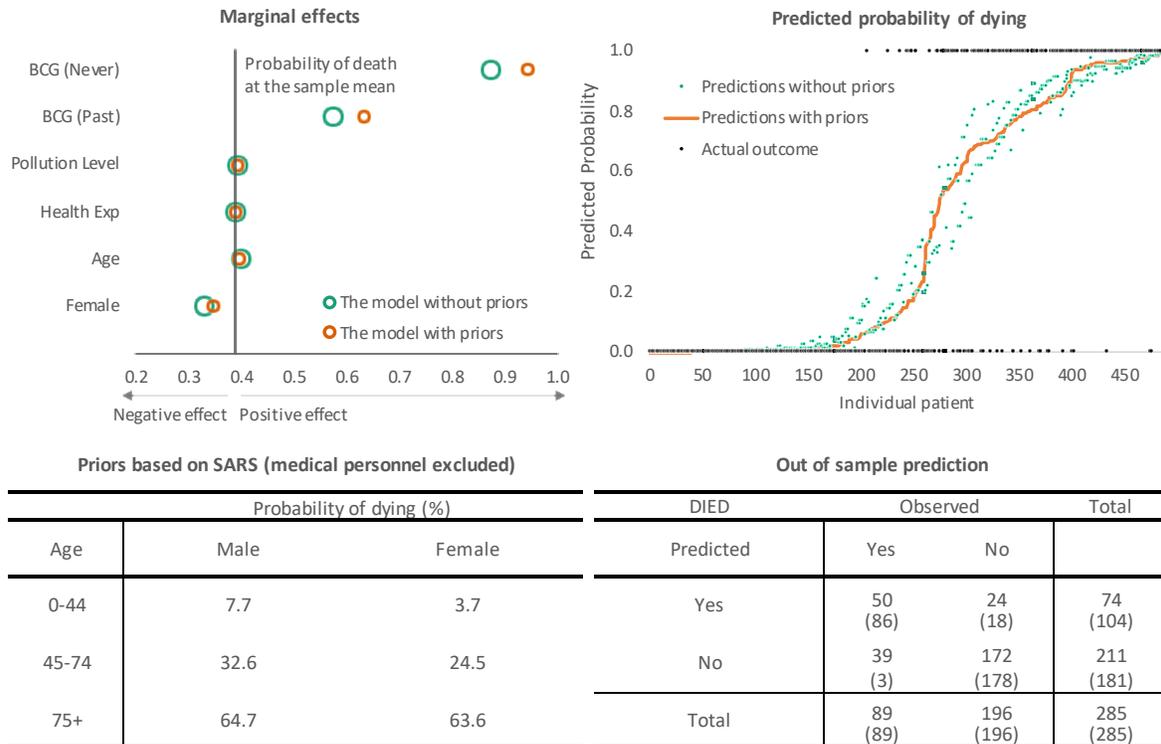


Fig. 1. Marginal effects (Top Left) of model n (green) and model p (brown). The predicted death probabilities for the patients under both models are on the Top Right. The black dots are the observed values of zero (recover) or one (died). The priors are on the Bottom Left. The out-of-sample prediction table is on the Bottom Right. Numbers in parentheses are the predictions from model p, the others are based on model n.

Fig. 2 (Top Left) presents the results from the first comparison, in which we examined death probability conditional on universal BCG vaccination. There are three universal vaccination possibilities: countries that never had it, that currently have it, or that never had it. That panel shows the huge impact on survival rates, across ages, of a universal BCG vaccination.

Fig. 2, Top Right and Bottom Left demonstrate the probability of dying conditional on health-related expenditure or air pollution death. The continuous lines reflect the lower 10th percentile of health and 90th percentile of pollution. The dashed lines reflect the 90th percentile level of health and 10th percentile of pollution. Bottom Right shows a combined policy of simultaneously increasing health expenditure and decreasing pollution level, each by 25% from the mean, for female patients. Such a policy will increase the survival probability of an elderly patient by approximately 50%.

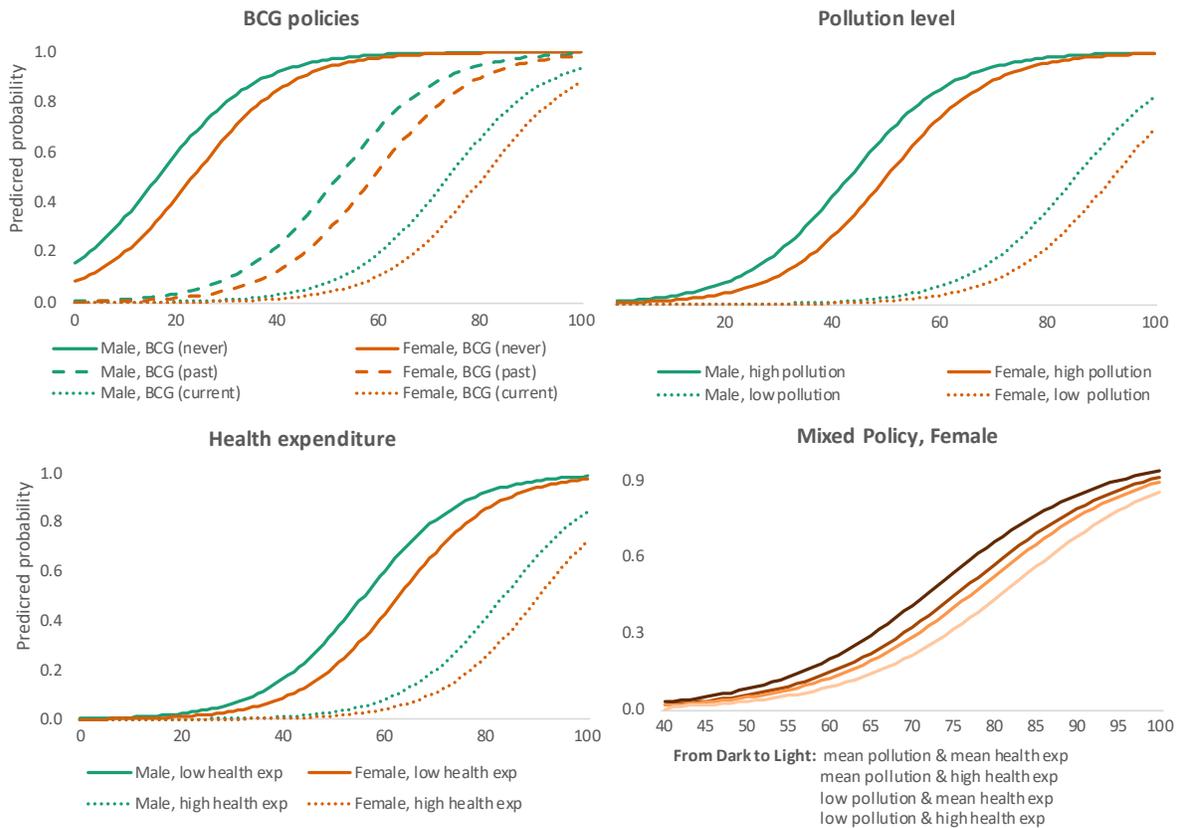


Fig. 2. The probability of dying conditional on different factors. BCG vaccination policies (Top Left). Pollution (Top Right) show the death rate in the 10th percentile (dots) vs. those at the 90th percentile (continuous). Health expenditure (Bottom Left) show the death rate in the 10th percentile (continuous) vs. those at the 90th percentile (dots). Simultaneously (Bottom Right) increasing health expenditure and decreasing pollution level, each by 25% from the mean, for female patients.

To put this in perspective, consider the current cumulated number of infected individuals of approximately 6.1 million. Assuming 5% of these are female (male) of at least 65 years old (their current percent of world population), means that an increase of 10% in health-related expenditures would reduce the 350,000 deaths from the virus by at least 9,215 elderly women and 11,276 elderly men. A 10% decrease of pollution will lower the deaths of elderly women (men) by about 13,490 (16,700). The combined policy of a 10% increase in health-related expenditure, together with a 10% decrease in pollution, will decrease elderly women (men) deaths by about 21,726 (25,740).

The most pronounced result in this analysis is the significant impact of the universal BCG vaccination on the survival rate. If such a policy existed everywhere, the number of elderly lives saved would have been at least 65,000 (assuming 25% of current deaths are from countries without current or past universal vaccination).

Model Prediction versus Current Observed Death Rate

Using data from the WorldOMeter website we compared our conditional probabilities with the up to date (unconditional) frequencies of death (out of infected individuals) by county. Our measure of comparison is the ratio of death rate in the United States to that of the other country for a 52-year old patient. Neither the United States or Canada ever had a universal BCG vaccination, and both have similar levels of air pollution. They differ in their health-related expenditures, which is 43% higher in the United States. The ratio based on our model is 0.7. The real data shows a ratio of 0.8 (or 0.7 in per 1 million normalization). The United States and Australia are similar in their air pollution and health-related expenditure levels. They differ in their universal BCG vaccination policies. Australia had a universal BCG vaccination until the mid-1980's (BCG World Atlas, 5). The United States never had a universal BCG vaccination policy. Our model's predicted ratio is 12.9. The real data show a ratio of 4.2 (based on the totals). Given our very small and partial data these results are incredibly strong and support the validity of our model.

Materials and Methods

Datasets

The patient-level data are from the Open COVID-19 Data Curation Group (6), and (7-8). Unfortunately, the majority of the patients in these data have missing information and cannot be used. Country-level data are from the World Bank (9) and WHO (10). The SARS data for priors are from (11). Most of the global variables are from most recent data between 2015-2018. The binary data for BCG vaccination are from the BCG World Atlas (2017).

Variables Used

In addition to age and sex, we use the following variables. *BCG vaccination* (dummy variables for 'the country never had a universal BCG vaccination program,' the country had such a policy in the past but currently does not,' and the country currently has a universal BCG

vaccination program). *Health-related expenditures* is an indicator taken from the World Bank and created by WHO called ‘domestic private health expenditure per capita.’ It is an indicator that (i) varies across countries, (ii) does not capture the direct expenditures on immunizations, and (iii) can be easily translated into a policy. It has a very low correlation with the universal BCG vaccination. *Pollution Death*, captures the number of deaths attributable to the joint effects of household and ambient air pollution in a year per 100,000 population. The rates are age standardized. We also used the country’s immunization rates of Measles and Hepatitis B as control variables. *Dead* is a dummy variable with the value 1 for died, and zero for a full recovery. For the multinomial outcome we have three categories: zero for full recovery, 1 for died, and 2 for ‘other.’

Information-Theoretic Model

We develop and use two information-theoretic (IT), discrete choice, binary models, one without priors (model n) and one with priors (model p) to estimate the association between the underlying policy-relevant factors and death. Both models control for age, sex, and whether the country had universal vaccination for measles and Hepatitis B. The models’ coefficients and marginal effects are all highly significant using χ^2 statistics, the Pseudo-R² is 0.61 and 0.54 for models n and p respectively. All predictions and comparisons are done using the models’ coefficients. All in, and out, of sample predictions are approximately 90% correct.

An IT inferential approach is particularly useful in this case due to the lack of data and high correlations. IT inferential methods combine the tools of information theory with those of statistical inference, within a constrained optimization setup, and build on the principle of maximum entropy (12-13). All information enters as constraints, and the priors are included in the entropy criterion function. To avoid distributional assumptions, and to handle the COVID-19 ill-behaved data, we use the more generalized, and flexible, special IT model of (14) and (15). As a robustness test we used a sample of 611 observation with an outcome variable of Recover/Died/Other, where ‘Other’ stands for ‘still in recovery’ or ‘unknown yet.’ The results of this model are consistent with those of our binary models for the Recovery/Dead outcomes.

References

1. Beam Dowd, J., et. Al, Demographic science aids in understanding the spread and fatality rates of COVID-19, PNAS May 5, 2020 117 (18) 9696-9698; first published April 16, 2020
2. Miller, A., Reandelar, J. M., Fasciglione, K., Roumenova, V., Li, Y., & Otazu, G. H. (n.d.). *Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: An epidemiological study* | medRxiv. Retrieved May 19, 2020, from <https://www.medrxiv.org/content/10.1101/2020.03.24.20042937v1>
3. Root-Bernstein, R. (2020). *Why Infants Rarely Die of COVID-19 and Morbidity and Mortality Rates Vary by Location: Pneumococcal and Hib Vaccinations as Possible Means to Mitigate Future Pandemics*. <https://doi.org/10.20944/preprints202004.0233.v1>
4. K. Clay, J. Lewis, E. Severnini, Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic. *The Journal of Economic History* **78**, 1179–1209 (2018).
5. Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M (2011) The BCG World Atlas: A Database of Global BCG Vaccination Policies and Practices. PLoS Med 8(3): e1001012. <https://doi.org/10.1371/journal.pmed.1001012>
6. *Latest Data* (April 24, 2020): *Open access epidemiological data from the COVID-19 outbreak*. (n.d.). GitHub. <https://github.com/beoutbreakprepared/nCoV2019>
7. Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., Cohn, E. L., Hswen, Y., Hill, S. C., Cobo, M. M., Zarebski, A. E., Li, S., Wu, C.-H., Hulland, E., Morgan, J. D., Wang, L., O'Brien, K., Scarpino, S. V., Brownstein, J. S., ... Kraemer, M. U. G. (2020b). Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific Data*, 7(1), 106. <https://doi.org/10.1038/s41597-020-0448-0>
8. Xu, B., Kraemer, M. U. G., Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Loskill, A., Wang, L., Cohn, E., Hill, S., Zarebski, A., Li, S., Wu, C.-H., Hulland, E., Morgan, J., Scarpino, S., Brownstein, J., Pybus, O., Pigott, D., & Kraemer, M. (2020). *The Lancet Infectious Diseases*, 20(5), 534. [https://doi.org/10.1016/S1473-3099\(20\)30119-5](https://doi.org/10.1016/S1473-3099(20)30119-5)
9. *World Bank Open Data* | *Data*. (n.d.). Retrieved May 19, 2020, from <https://data.worldbank.org/>
10. *Global Health Organization*. (n.d.). Global Health Observatory Data Repository: BCG - Immunization Coverage Estimates by Country; World Health Organization. <https://apps.who.int/gho/data/node.main.A830?lang=en>
11. Karlberg, J., Chong, D., & Lai, W. (2004). Do Men Have a Higher Case Fatality Rate of Severe Acute Respiratory Syndrome than Women Do? *American Journal of Epidemiology*, 159(3), 229–231. <https://doi.org/10.1093/aje/kwh056>
12. Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620–630. <https://doi.org/10.1103/PhysRev.106.620>
13. Levine, R. D. (1980). An information theoretical approach to inversion problems. *Journal of Physics A: Mathematical and General*, 13(1), 91–108. <https://doi.org/10.1088/0305-4470/13/1/011>
14. Golan, A., Judge, G., & Perloff, J. M. (1996). A Maximum Entropy Approach to Recovering Information from Multinomial Response Data. *Journal of the American Statistical Association*, 91(434), 841–853. <https://doi.org/10.1080/01621459.1996.10476952>
15. Golan, A. (2017). *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information*. Oxford: Oxford University Press USA - OSO.