Title: Face mask wearing rate predicts COVID-19 death rates across countries

Authors: D Miyazawa[1]*, G Kaneko[2]

Affiliations:

[1]Miyazawa Clinic, 1-6-5 Akuraminami Takarazuka, Hyogo 665-0823, Japan

[2]School of Arts & Sciences, University of Houston-Victoria. 3007 Ben Wilson, Victoria, TX 77901, United States.

*Correspondence to: kusami1@ybb.ne.jp

Abstract: Identifying biomedical and socioeconomic predictors of the number of deaths caused by COVID-19 can help the development of effective interventions. In this study, we used the hypothesis-driven regression approach to test the hypothesis that the mask wearing rate, along with age and obesity, can largely predict the cumulative number of deaths across countries. Our regression models explained 69% of the variation in the cumulative number of deaths per million (March to June 2020) among 22 countries, identifying the face mask wearing rate in March as an important predictor. The number of deaths per million predicted by our elastic net regression model showed high correlation (r = 0.86) with observed numbers. These findings emphasize the importance of face masks in preventing the ongoing pandemic of COVID-19.

(125 words/100–125 words)

One Sentence Summary: Face mask wearing rate in March is a strong predictor of the cumulative number of deaths per million caused by COVID-19 among 22 countries.

(116/125 characters)

Main Text:

There have been considerable differences in the number of deaths caused by COVID-19 across countries. Western countries, in particular, have recorded a high number of deaths compared to Eastern countries. Several regression studies have identified predictors for the differences in the number of deaths such as age, obesity, and previous BCG vaccination (*1-3*). However, to our knowledge, there have been no cross-country regression studies that used the face mask wearing rate as a predictor, although face masks have been given an increasing attention as an effective means to prevent transmission of COVID-19 (*4, 5*). This is possibly because the mask wearing rate has been available only for 22 countries. Since the number of predictors should not generally exceed 10% of the sample size in traditional regression analysis, the mask wearing rate may have been excluded from the pool of potential predictors in previous studies that used a large number of samples and predictors to achieve high prediction rates, limiting public awareness of the importance of face masks in protecting their health.

In this study, we employed the hypothesis-driven regression approach to identify the association of mask wearing rate and the cumulative number of deaths caused by COVID-19 across countries, based on the hypothesis that the mask wearing rate can largely predict the cumulative number of deaths along with age and obesity, which are the other relatively independent risk factors for hospitalized COVID-19 patients (*6-8*). This approach uses a limited number of *a priori* determined predictors based on a certain hypothesis and is especially useful in providing biological insights regarding the association with a small number of samples (*9-11*), while the implication is restricted to the hypothesis.

We first calculated the Spearman's correlations between the cumulative number of deaths per million from various dates and predictors related to transmission (mask non-wearing rate and rate to avoid public spaces), age, and obesity in order to determine variables to include the regression (Fig. 1). The correlation coefficients between the cumulative number of deaths per million and mask non-wearing rates (both Mar and Apr-May) showed time-dependent increase, reaching an apparent plateau in early May. The mask non-wearing rate in Mar generally showed higher correlation coefficients compared to that of Apr-May with the highest correlation of 0.79 on June 6, 2020 (P = 6.865e-06). The rate to avoid public spaces showed low correlations with the cumulative number of deaths per million.

Age-related predictors were highly associated with the cumulative number of deaths per million in consistent with a previous report (*7*). The correlation coefficients were lower in younger age groups (age 65-69 years) compared to the higher age groups in both sexes (highest rho = 0.68). In contrast to the mask non-wearing rate, correlations between the cumulative number of deaths and age-related predictors were highest in mid-May and decreased thereafter.

Body mass index (BMI) generally showed lower correlations with the cumulative number of deaths per million with a clear sex-dependent difference. Male BMI was more closely associated with the cumulative number of deaths per million than female BMI, corroborating with the fact that males are more susceptible to severe outcomes from COVID-19 compared to females (*12*, *13*). The highest correlation of 0.59 was observed between male BMI and the cumulative number of deaths per million on May 26.

Correlation analyses using the weekly number of deaths per million showed the similar tendency, in which the mask non-wearing rate in Mar showed the highest association (Fig. S1). Time course profiles for Spearman's correlation coefficients of several highly correlated parameters further confirmed the unique feature of age ≥ 80 (male) among parameters used (Fig. S2). Spearman's correlation coefficients for this parameter remained almost constant throughout the study period, whereas those for other three parameters showed time-dependent increase and reached the apparent plateau around day 50 (mid-May).

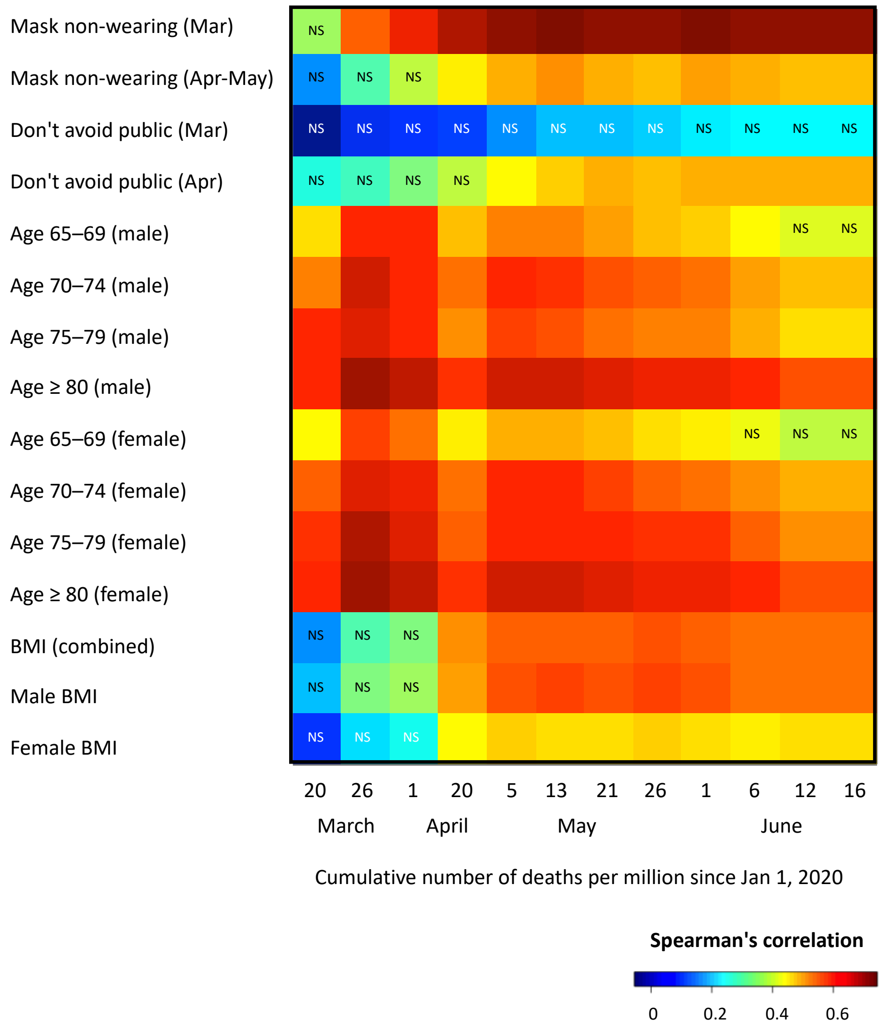


Fig. 1. **Spearman's correlation.** Correlations were calculated between the cumulative number of coronavirus disease-related deaths per million, and potential predictors related transmission, age, and obesity. Mask wearing rates in Mar and Apr-May were calculated from the survey responses during March 9–18, 2020 and April 26 – May 1, 2020, respectively. Age-related predictors indicate the percentage of the population in the specified range. BMI, body mass index; NS, not significant ($P > 0.05$).

Next, we created scatter plots representing the relationship between the cumulative number of deaths per million and the several potential predictors showing high Spearman's correlations (Fig. 2A, top panels). The number of deaths per million on May 13 was used because it showed high correlations with the selected predictors. The cumulative number of deaths per million showed exponential association with most predictors with clear separations between the Western and Eastern countries, especially in the plot with the mask non-wearing rate (Mar). We therefore applied the logarithmic transformation to the cumulative number of deaths per million (Fig. 2A, bottom panels), which resulted in significant linear correlations of the transformed value with the mask non-wearing rate in Mar (r = 0.796, P = 9.356e-06), mask non-wearing rate from Apr-May (r = 0.496, *P* = 0.01897), age ≥80 years (male) (r = 0.658, *P* = 0.0008717), and male BMI (r = 0.682, *P* = 0.0004762).

Interestingly, in the Western countries there was a tendency toward a negative association between the number of deaths per million and mask non-wearing rate in April - May (Spearman's correlation; r = -0.4126551, P = 0.1611). This is attributed to the marked reduction in the mask non-wearing rate in some countries with high number of deaths, possibly caused by the fear of the disease. Likewise, within the United States, the mask non-wearing rate in April across states also showed a negative correlation with the number of deaths (Spearman's correlation; r = -0.477556, P = 0.0003946) (Fig. S3).

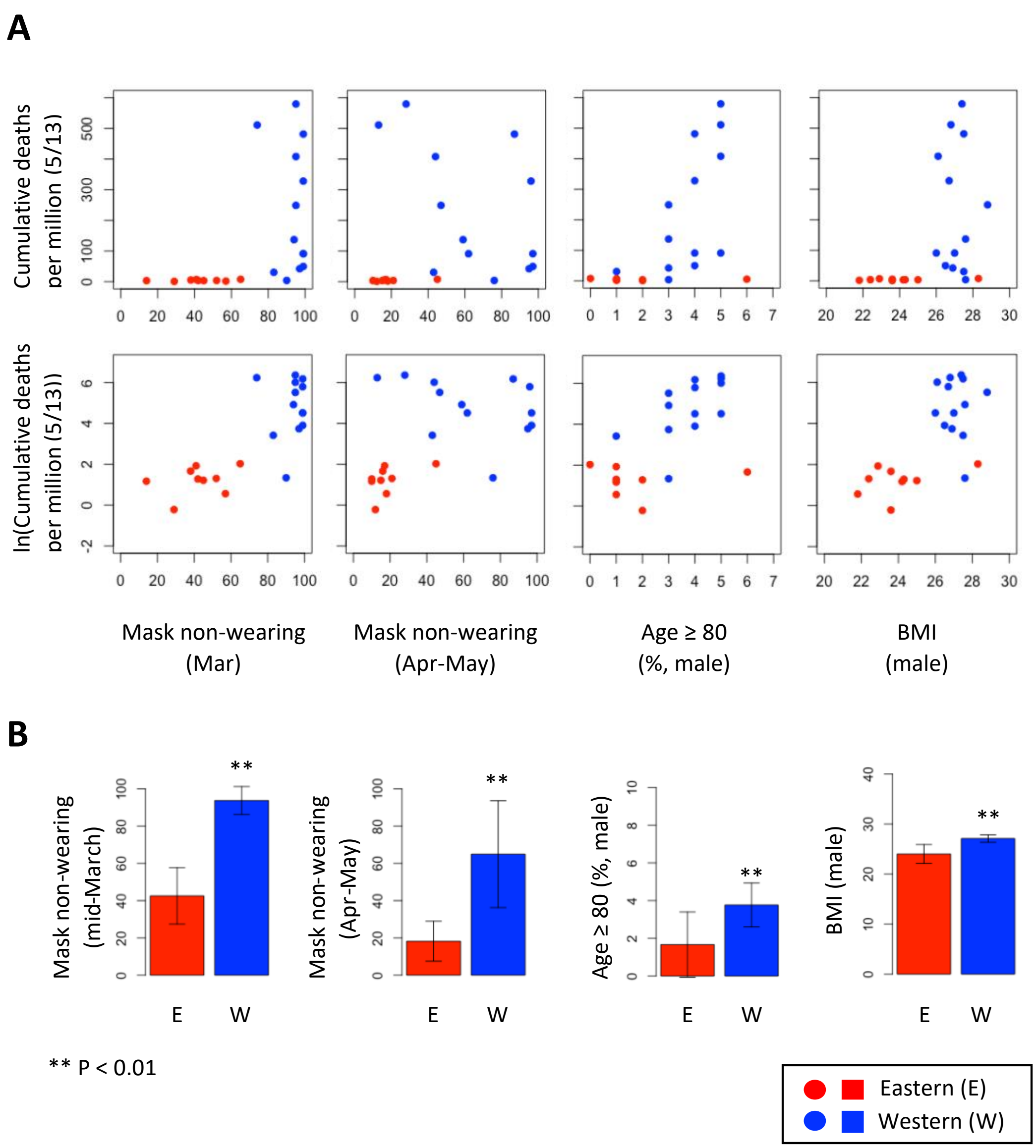


Fig. 2. Logarithmic transformation of the cumulative number of deaths and West-East difference. (**A**) Scatter plots showing the correlations among the parameters related to the number of cumulative deaths per million, mask non-wearing rate, age, and obesity. The number of deaths per million on May 13 was used with (bottom panels) and without (top panels) logarithmic

transformation. BMI, body mass index. (**B**) Western countries have higher mask non-wearing rates, percentage of individuals ≥80 years, and male BMI compared to the Eastern countries. ***P* < 0.01 in Wilcoxon rank-sum test. Western countries include UK, France, Italy, USA, Spain, Mexico, Germany, Canada, Sweden, Norway, Finland, Denmark, and Australia, whereas Eastern countries include Malaysia, China, Saudi Arabia, India, Indonesia, Philippines, Japan, Singapore, and Thailand.

We subsequently attempted to predict the log-transformed cumulative number of deaths per million using the multiple linear regression approach. The mask non-wearing rate (Mar) along with age ≥80 (male) explained 68.6% of the variation of the response variable (model 1). Male BMI was a relatively weak predictor compared to age ≥80 (male) but was significantly associated with the response variable when the mask non-wearing rate (Mar) was excluded from the model (models 2 and 3). On the other hand, mask non-wearing rate in Apr-May was not significantly associated with the response variable (models 4 and 5), suggesting the importance of face masks in the early phase of the pandemic. The mask non-wearing rate (Mar) by itself predicted 61.6% of the variation in the cumulative number of deaths per million in a single regression analysis (beta = 0.06146, *P* = 9.36e-06).

It is noted that we found a weak but significant correlation (Pearson's correlation ~ 0.5) between the mask non-wearing rate (Mar), age ≥80 (male), and male BMI (Fig. S4), although the degree of correlation was too low to introduce multicollinearity in our regression (Table 1). This correlation may be attributed to the breathing difficulties because both obese (*14*) and aged (*15*) individuals show impaired lung function.

Table 1. Regression

Model	Predictors	Beta (SE)	P value	Adjusted R2 (P value) VIF
1	Mask non-wearing rate (Mar)	0.048 (0.011)	0.000347**	0.686 (6.468e-06) 1.369
	Age ≥80 (male)	0.415 (0.177)	0.030394*	
	Intercept	-1.367 (0.740)	0.080565	
2	Mask non-wearing rate (Mar)	0.050 (0.016)	0.00488**	0.614 (4.539e-05) 2.261
	Male BMI	0.210 (0.217)	0.34587	
	Intercept	-5.733 (4.823)	0.24919	
3	Age ≥80 (male)	0.609 (0.169)	0.00185**	0.649 (1.832e-05) 1.110
	Male BMI	0.561 (0.145)	0.00104**	
	Intercept	-12.93 (3.633)	0.00209**	
4	Mask non-wearing rate (Apr-May)	0.020 (0.011)	0.08514	0.466 (0.0009962) 1.130
	Age ≥80 (male)	0.685 (0.210)	0.00407**	
	Intercept	0.409 (0.732)	0.58320	
5	Mask non-wearing rate (Apr-May)	0.010 (0.013)	0.45933	0.4256

	Male BMI	0.632 (0.216)	0.00862	(0.001993)
	Intercept	-13.46 (5.257)	0.01915*	1.501

*P < 0.05, **P < 0.01. VIF, variance inflation factor; SE, standard error. VIF < 2.5 is a strict threshold of VIF in evaluating the presence of multicollinearity.

Lastly, we predicted the log-transformed number of deaths per million by the mask non-wearing rate (Mar), age ≥80 (male), and male BMI using the elastic net regression. The elastic net regression is a machine learning method that estimates the regression coefficients with penalty terms, enabling us to include a larger number of predictors compared to the traditional multiple regression. Observed cumulative number of deaths per million were significantly correlated with those predicted by the elastic net regression model (Fig. 3).

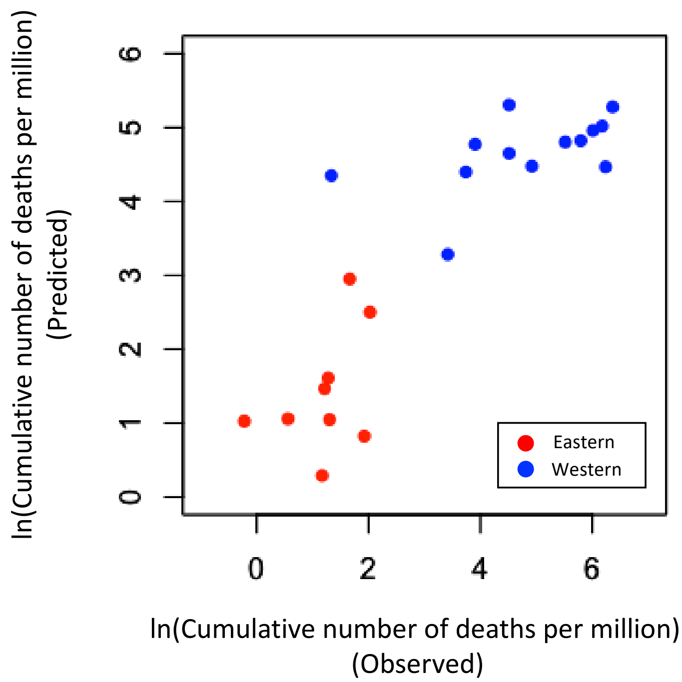


Fig. 3. Elastic net regression. The observed and predicted cumulative number of deaths per million were highly correlated with a Pearson's correlation coefficient of 0.863 (P = 2.417e-07). Lambda = 0.1, alpha = 0.8, intercept = -6.519. Coefficients for age ≥80, male BMI, and mask non-wearing rate in Mar were 0.410, 0.246, and 0.032, respectively.

Face masks are considered effective in preventing the transmission of COVID-19 (*16*, *17*). In line with these previous studies, the mask non-wearing rate in March was found to be the strongest predictor of the cumulative number of deaths per million in this study, where even the single regression explained 61.6% of the variation of the response variable. We also observed that there is a considerable difference in face mask wearing rates between the Western and Eastern countries. A country's policy regarding wearing face masks alone cannot explain this big difference because, for example, face masks were never been mandated in Japan, despite the high face mask wearing rate observed in the country (*18*). We speculate that cultural factors could be the major reason for the difference in mask wearing rate as Jack et al. (*19*) described that "whereas Western Caucasian internal representations predominantly featured the eyebrows and mouth, East Asian internal representations showed a preference for expressive information in the eye region." This tendency could explain why it is considered rude to wear sunglasses among East Asians and suspicious to wear face masks in Western countries (*20*).

A limitation of this study is that we were able to obtain the mask wearing rate from only 22 countries. While we demonstrated the strong association of the predictor and the cumulative number of deaths caused by COVID-19 by taking advantage of the hypothesis-driven regression, this finding should be verified by building a model with a sufficient number of samples and predictors. Another limitation is that the regression models used in the present study do not directly test the causal relationship of observed variables. Appropriate study design such as cohort sampling or introduction of interventions are needed to address the causality.

Taken together, the present study demonstrated the close association between face mask wearing rate and the number of deaths caused by COVID-19, identifying the age and obesity as relatively weak predictors. These findings have an implication for introducing mandatory face mask usage as the precautionary principle.

References and Notes:

Acknowledgments: The authors express sincere thanks to the journalist Mr. Junya Iwai for his help in data collection and constructive discussion. **Author contributions: Daisuke Miyazawa**: Conceptualization, Data curation, Project administration, Resources, Writing - Original Draft, **Gen Kaneko**: Methodology, Software, Formal analysis, Validation, Visualization, Writing - Review and Editing. **Competing interests:** None. **Data and materials availability:** All data and code are available in the main text or the supplementary materials.

Supplementary Materials:

Materials and Methods

Figures S1-S4

1. N. Y. Leung, M. A. Bulterys, P. L. Bulterys, Predictors of COVID-19 incidence, mortality, and epidemic growth rate at the country level. *medRxiv*. 2020.05.15.20101097 [Preprint] 19 May 2020. https://doi.org/10.1101/2020.05.15.20101097.
2. J. Squalli, Evaluating the determinants of COVID-19 mortality: A cross-country study. *medRxiv*. 2020.05.12.20099093 [Preprint] May 17 2020. https://doi.org/10.1101/2020.05.12.20099093.
3. C. R. Knittel, B. Ozaltun, What does and does not correlate with COVID-19 death rates. *medRxiv*. 2020.06.09.20126805 [Preprint] Jun 11 2020. https://doi.org/10.1101/2020.06.09.20126805.
4. C. T. Leffler *et al.*, Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. Update June 2, 2020.

medRxiv. 2020.05.22.20109231 [Preprint] Jul 7 2020. https://doi.org/10.1101/2020.05.22.20109231.

5. D. K. Chu *et al.*, Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*, Published online (2020).
6. D. Miyazawa, Why obesity, hypertension, diabetes, and ethnicities are common risk factors for COVID-19 and H1N1 influenza infections. *J. Med. Virol.* **Published online**, (2020).
7. A. L. Mueller, M. S. McNamara, D. A. Sinclair, Why does COVID-19 disproportionately affect older people? *Aging* **12**, (2020).
8. K. K.-W. To *et al.*, Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *The Lancet Infectious Diseases*, (2020).
9. N. J. Klein, A. J. Beck, D. A. Hutchins, S. A. Sanudo-Wilhelmy, Regression modeling of the North East Atlantic Spring Bloom suggests previously unrecognized biological roles for V and Mo. *Front. Microbiol.* **4**, 45 (2013).
10. M. R. Sabuncu *et al.*, Morphometricity as a measure of the neuroanatomical signature of a trait. *Proc. Natl. Acad. Sci. USA* **113**, E5749-E5756 (2016).
11. A. Carriedo, J. A. Cecchini, C. González, Soccer spectators' moral functioning and aggressive tendencies in life and when watching soccer matches. *Ethics & Behavior*, 1-15 (2020).
12. M. Montopoli *et al.*, Androgen-deprivation therapies for prostate cancer and risk of infection by SARS-CoV-2: a population-based study (n= 4532). *Ann. Oncol.*, (2020).
13. G. Grandi, F. Facchinetti, J. Bitzer, The gendered impact of coronavirus disease (COVID-19): do estrogens play a role? *The European Journal of Contraception & Reproductive Health Care*, 1-2 (2020).
14. L. C. Melo, M. A. M. d. Silva, A. C. d. N. Calles, Obesity and lung function: a systematic review. *Einstein (Sao Paulo)* **12**, 120-125 (2014).
15. R. Dhar, M. Gupta, V. Khajuria, N. Singh, Evaluation of pulmonary function tests in the elderly population. *Int. J. Med. Sci. Public Health* **6**, 1-5 (2017).
16. D. Kai, G.-P. Goldstein, A. Morgunov, V. Nangalia, A. Rotkirch, Universal masking is urgent in the covid-19 pandemic: Seir and agent based models, empirical validation, policy recommendations. *arXiv preprint arXiv:2004.13553*, (2020).
17. V. Chernozhukov, H. Kasahara, S. P, Causal Impact of Masks, Policies, Behavior on Early Covid-19 Pandemic in the U.S. *medRxiv*. 2020.05.27.20115139 [Preprint] Jul 7 2020. https://doi.org/10.1101/2020.05.27.20115139.
18. D. Miyazawa, G. Kaneko, Japan's key "X-factor" for low COVID-19 mortality may be its culture of wearing face masks. *Authorea*. [Preprint] July 02, 2020. 10.22541/au.159373225.52467275.
19. R. E. Jack, R. Caldara, P. G. Schyns, Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General* **141**, 19 (2012).
20. R. R. Gesteland, *Cross-cultural business behavior: A guide for global management*. (Copenhagen Business School Press DK, 2012).

Supplementary Materials:

Materials and Methods

All data were collected from publicly available secondary sources. Analyses in the present study included 13 Western countries (UK, France, Italy, USA, Spain, Mexico, Germany, Canada, Sweden, Norway, Finland, Denmark, and Australia) and 9 Asian countries (Malaysia, China, Saudi Arabia, India, Indonesia, Philippines, Japan, Singapore, and Thailand). These countries were chosen because of the availability of March mask-wearing data. The face mask wearing rates in March (March 9 to March 18) 2020 and late April to early May (April 26 to May 1) 2020 across countries were derived from "percentage of people in each country who answered that they are wearing a face mask when in public spaces" from the YouGov database (https://yougov.co.uk/topics/international/articles-reports/2020/05/01/international-covid-19-tracker-update-2-may). This database has collaborated with the Institute of Global Health Innovation at Imperial College London, and it summarizes interviews conducted with nationally representative sample sizes (150-2000/week depending on countries). Total COVID-19 deaths per million were obtained from https://ourworldindata.org/coronavirus-source-data. The BMI data were obtained from the Global Status Report on Non-communicable Diseases 2014 (October 5, 2015; https://www.who.int/nmh/publications/ncd-status-report-2014/en/). Population percent by age data was obtained from https://data.worldbank.org on June 15, 2020.

Data analysis

Statistical analyses were conducted using R version 3.6.2. The R packages used in this study include fields, ggplot2, glmnet, diverse, broom, car, and ggcorrplot. Pearson's and Spearman's correlations were calculated using the "cor" and "cor.test" functions. Multiple and single linear regression models were built using the "lm" function.

Figures S1-S4

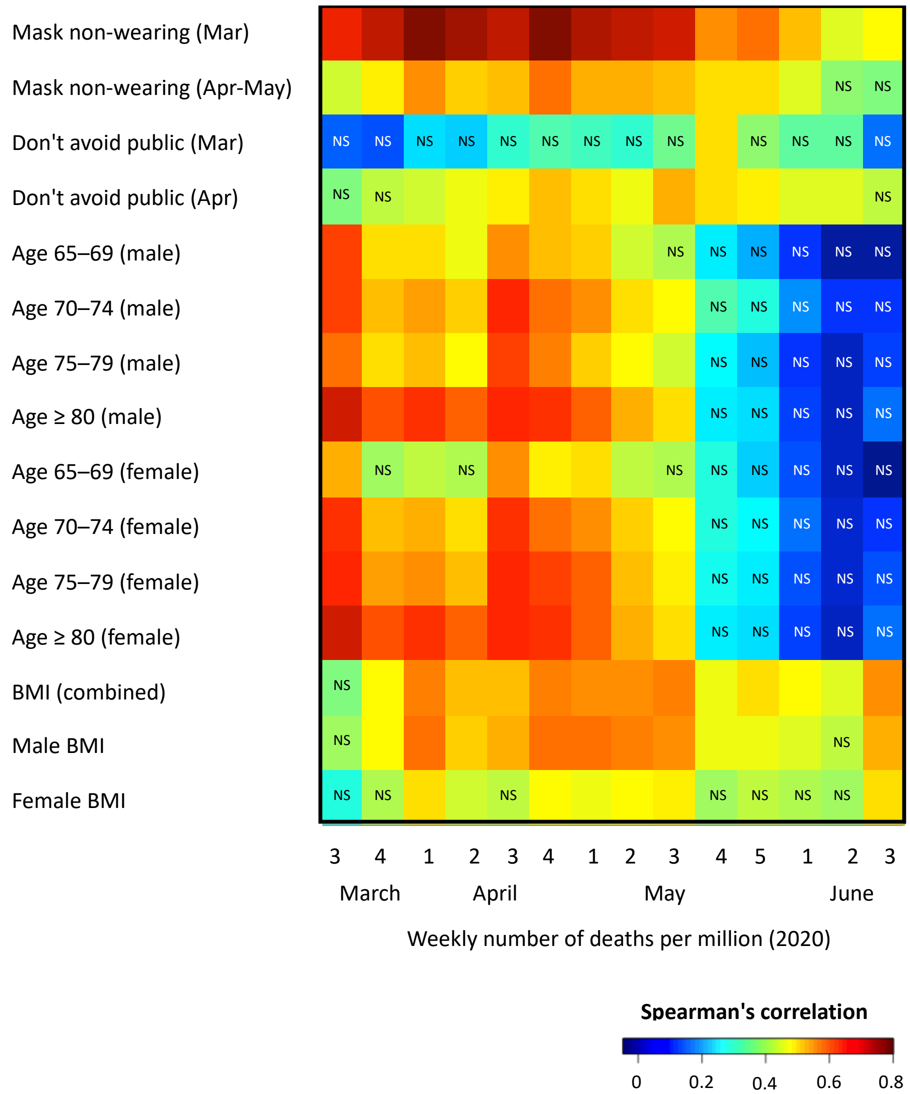
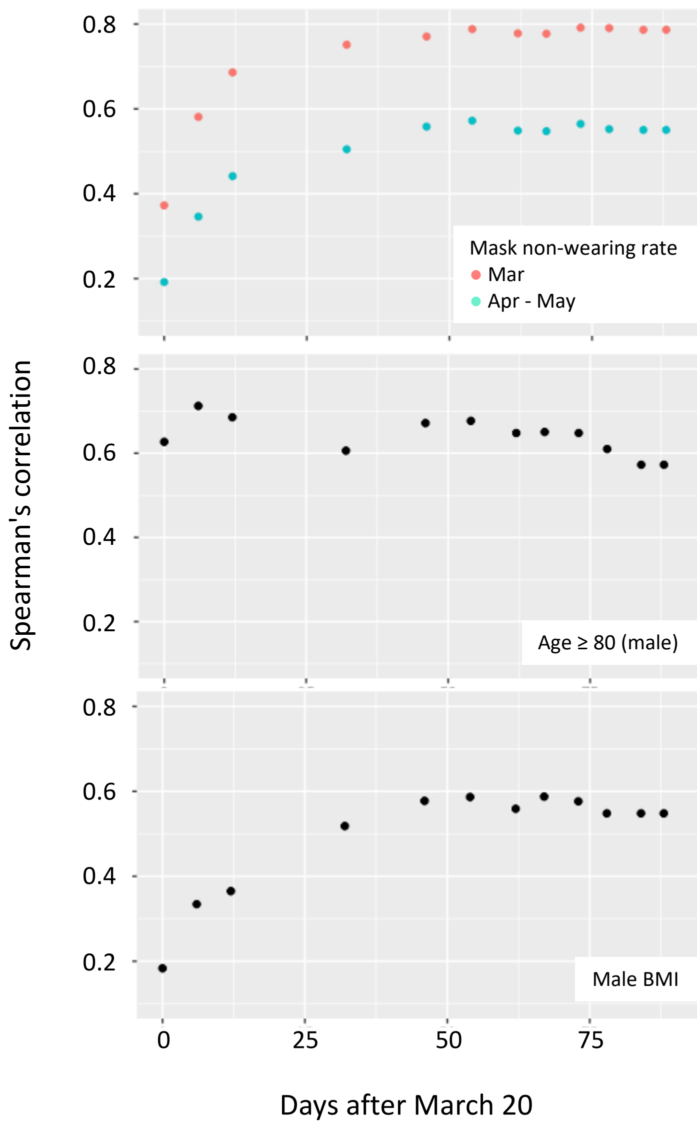


Fig. S1. **Spearman's correlation.** Correlations were calculated between the weekly number of coronavirus disease-related deaths per million, and potential predictors related transmission, age, and obesity. Numbers on x axis refer the 1st to 5th week of each month. Mask wearing rates in Mar and Apr-May were calculated from the survey responses during March 9–18, 2020 and April 26 – May 1, 2020, respectively. Age-related predictors indicate the percentage of the population in the specified range. BMI, body mass index; NS, not significant ($P > 0.05$).



5 **Fig. S2**. **Time-dependent changes in Spearman's correlations.** Spearman's rho values were calculated between the cumulative number of deaths per million on May 13 and mask non-wearing rates (mid-March [Mar] and late April to early May), age ≥80 years, and male body mass index (BMI).

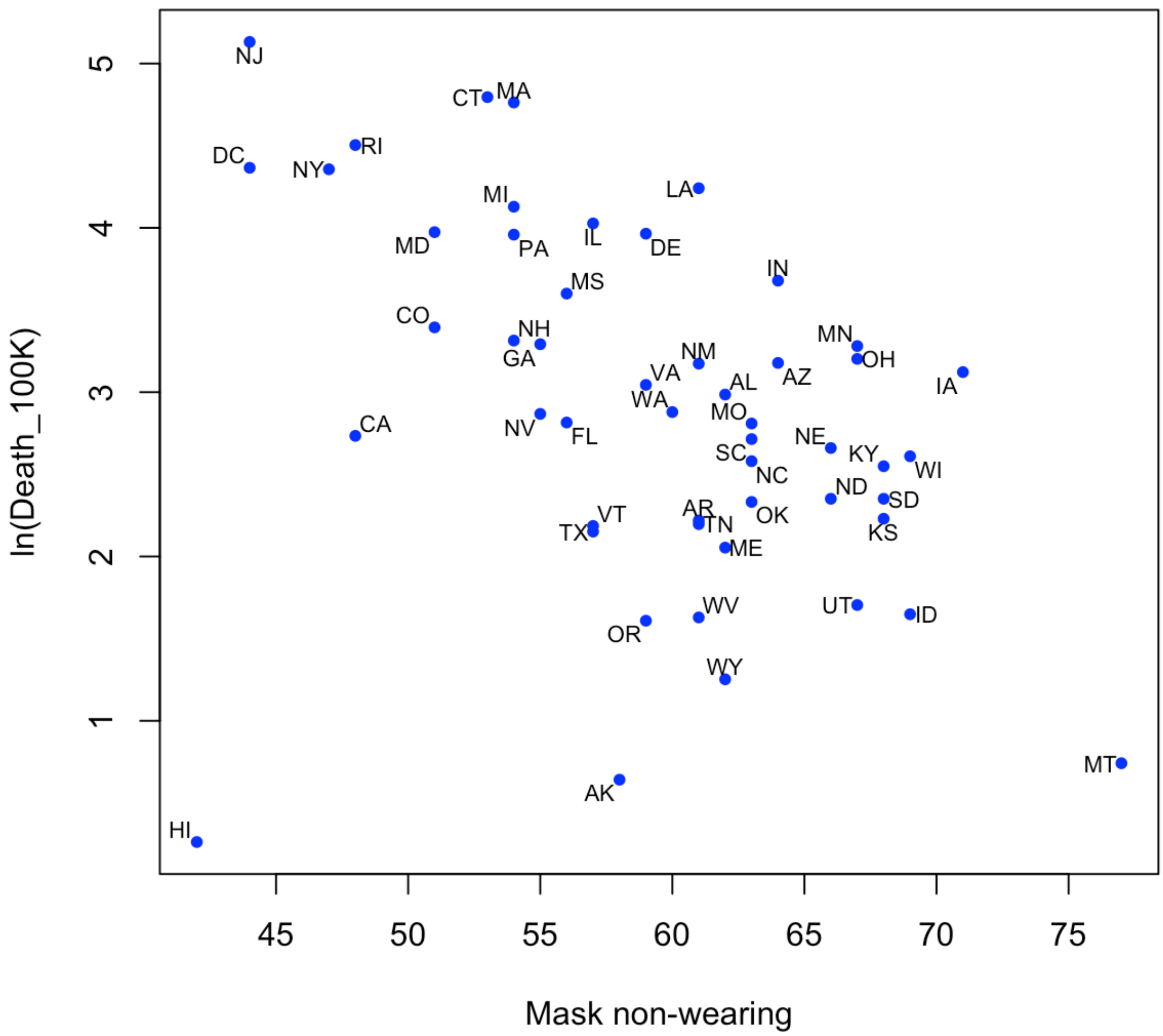


Fig. S3. **Apparent negative correlation between the number of deaths and mask non-wearing rate in the United States**. Scatter plot showing cross-state correlation between ln(total number of deaths) and mask non-wearing rate (April) in the United States.

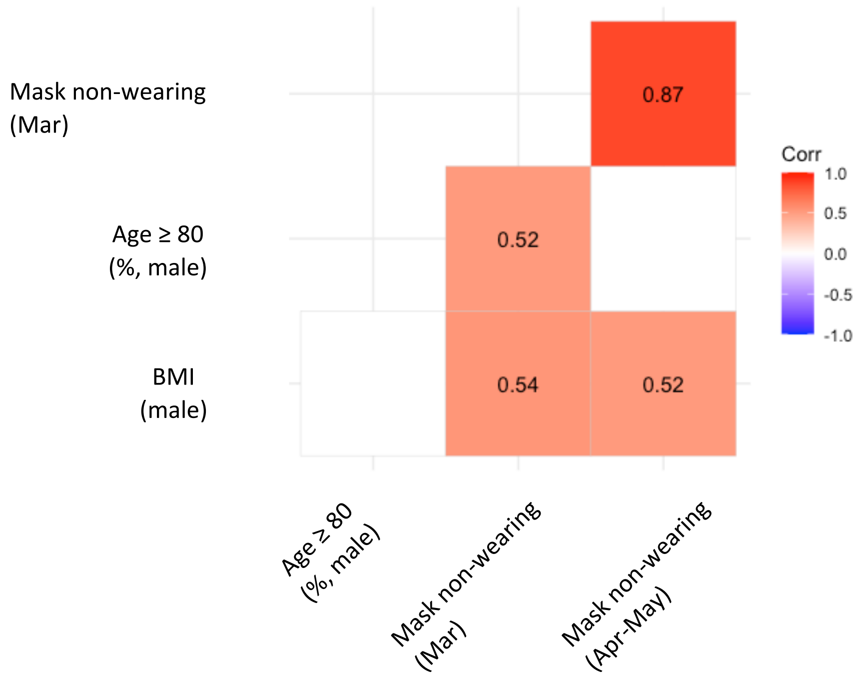


Fig. S4. Spearman's correlation matrix between predictors. Only significant correlations are shown with rho values (P < 0.05). Mask non-wearing rates in Mar and Apr-May showed high correlations, whereas those also showed relatively low correlations with age ≥80 years (male), and male body mass index (BMI).