

Supplement for “Population-Level Mortality Rates from Novel Coronavirus (COVID-19) in South Korea”

Samir Soneji* JaeWon Yang[†] Caroline Mann[‡] Hiram Beltrán-Sánchez[§]

*Department of Health Behavior; University of North Carolina, Chapel Hill. Email: sonejis@email.unc.edu

[†]Brown University

[‡]University of North Carolina, Chapel Hill

[§]Department of Community Health Sciences and California Center for Population Research, University of California, Los Angeles. Email: beltrans@ucla.edu

A Methodology

Let *Cases* and *Deaths* represent the total number of reported cases of and deaths, respectively, from the novel coronavirus (COVID-19) in a given country (e.g., South Korea) observed over a specified exposure period (*Exposure Period*). Let \mathcal{A} be a set of the starting ages for the age intervals and ω represent the starting age of the oldest age interval. Let n_a represent the width (years) of an age interval starting at age $a \in \mathcal{A}$. Typically, the width of age intervals is the same for all but the oldest age interval $[\omega, \infty)$, that is $n_a = n$ for all $a \in \mathcal{A} \setminus \{\omega\}$ and $n_\omega = \infty$. Let $cases(a)$ represent the number of reported cases in age interval $a \in \mathcal{A}$.

$$Cases = \sum_{a \in \mathcal{A}} cases(a)$$

Similarly, let $deaths(a)$ represent the number of deaths in age interval $a \in \mathcal{A}$.

$$Deaths = \sum_{a \in \mathcal{A}} deaths(a)$$

Let *Case Distribution* represent the proportion of cases in age interval $a \in \mathcal{A}$ among all cases.

$$Case\ Distribution(a) = \frac{cases(a)}{\sum_{a \in \mathcal{A}} cases(a)}$$

Let *Case Fatality Rate* represent the proportion of cases in age interval $a \in \mathcal{A}$ who died among all cases in age interval $a \in \mathcal{A}$.

$$Case\ Fatality\ Rate(a) = \frac{deaths(a)}{\sum_{a \in \mathcal{A}} cases(a)}$$

Let $N(a)$ represent the exposure measured in the number of person-years lived in the calendar year among individuals aged $[a, a + n_a)$. We approximate $N(a)$ by the mid-year population of age interval $a \in \mathcal{A}$. Finally, let *Prevalence*(a) represent the population-level prevalence of COVID-19 in age interval $a \in \mathcal{A}$.

$$Prevalence(a) = \frac{cases(a)}{N(a)}$$

The population-level mortality rate (m) for age interval $a \in \mathcal{A}$ over the exposure period is equal to:

$$m(a) = \frac{deaths(a)}{N(a)}. \tag{1}$$

Thus, deaths can be expressed as a function of the mortality rate and the exposure (approximated by the mid-year population count).

$$deaths(a) = m(a) \times N(a) \tag{2}$$

We can express $m(a)$ in equation 1 as a function of the *Case Fatality Rate* and *Prevalence(a)* as follows.

$$\begin{aligned}
m(a) &= \frac{deaths(a)}{N(a)} \\
&= \frac{deaths(a)}{N(a)} \times \frac{cases(a)}{cases(a)} \\
&= \frac{deaths(a)}{cases(a)} \times \frac{cases(a)}{N(a)} \\
&= Case\ Fatality\ Rate(a) \times Prevalence(a)
\end{aligned} \tag{3}$$

Taking the natural logarithm of equation 3 and the identity in equation 1,

$$\begin{aligned}
\log(m(a)) &= \log(Case\ Fatality\ Rate(a)) + \log(Prevalence(a)) \\
\log\left(\frac{deaths(a)}{N(a)}\right) &= \log(Case\ Fatality\ Rate(a)) + \log(Prevalence(a)) \\
\log(deaths(a)) &= \log(Case\ Fatality\ Rate(a)) + \log(Prevalence(a)) + \log(N(a))
\end{aligned} \tag{4}$$

Equation 4 may be estimated by fitting a negative binomial regression on the number of deaths, $deaths(a)$ with $\log(N(a))$ as an offset. Let $\hat{deaths}(a)$ represent the predicted number of deaths in age interval $a \in \mathcal{A}$ based on the negative binomial regression model. The predicted mortality rate over the exposure period, $\hat{m}(a)$, equals $\frac{\hat{deaths}(a)}{N(a)}$. Finally, the predicted mortality rate for age interval $a \in \mathcal{A}$ over an entire year, $\hat{M}(a)$, equals:

$$\hat{M}(a) = \frac{Number\ Days\ in\ Year}{Exposure\ Period} \times \frac{\hat{deaths}(a)}{N(a)}$$

B Estimated Population-Level Mortality Rates, South Korea

We applied the methodology described in Section A to South Korean data current as of 22 March 2020 (8897 confirmed cases of and 104 deaths from COVID-19).

Supplemental Table 1: Population-Level Mortality Rates (Deaths per 100,000 Person-Years)

Age Interval (Years)	Pt. Est.	95% CI, Lower Bound	95%, Upper Bound
0-9	0.0	0.0	0.0
10-19	0.0	0.0	0.0
20-29	0.0	0.0	0.0
30-39	0.1	0.0	0.2
40-49	0.1	0.0	0.1
50-59	0.5	0.3	0.8
60-69	1.5	1.2	2.0
70-79	6.0	4.5	8.1
≥ 80	12.8	9.6	17.2