Stanford-CIDE Coronavirus Simulation Model (SC-COSMO) – Technical Description Document, Version 1.0

Jeremy Goldhaber-Fiebert, PhD[1], Fernando Alarid-Escudero, PhD[2], and Jason Andrews, MD, MPH[1]

[1]Stanford University, CA, US
[2]Center for Research and Teaching in Economics (CIDE), Aguascalientes, Mexico

April 17, 2020

1 Model description

The epidemiology of COVID19 in the absence of treatment or vaccination can be described as a multi-compartment susceptible-exposed-infected-recovered (MC-SEIR) model [1] where the exposed (E) compartments represent individuals that are infected without symptoms and are not yet infectious, and the infectious (I) compartments represent individuals that are infected with symptoms and are infectious [1], meaning that they can infect other susceptibles if a contact occurs. Notably, by using differential rates of detection and differential rates of transmission (described in sections below) for the various I compartments, the model structure can also capture asymptomatic infectiousness as well as time-varying infectiousness over the course of infection. Figure 1 depicts a diagram of the generalized structure of the non-age-stratified MC-SEIR compartmental model. Each of these compartments represents part of the population characterized by their COVID-19 status as a function of time.

To appropriately model the age-dependent dynamics of COVID-19 and case severity, we expanded the simple MC-SEIR model to include a realistic age structure (RAS) and heterogeneous age-structured mixing. We divided the population into N age groups where each a-th age group has its own set of S_a, E_a^j, $I_{l,a}^k$, $IDX_{l,a}^k$, R_a compartments, for $a = 1 \ldots, N$, $j = 1 \ldots, J$, $k = 1 \ldots, K$, $l = 1 \ldots, L$. The RAS SEIR model for COVID-19 is described by the following system of $[(3 + J + 2(L + K))N]$ ODEs:

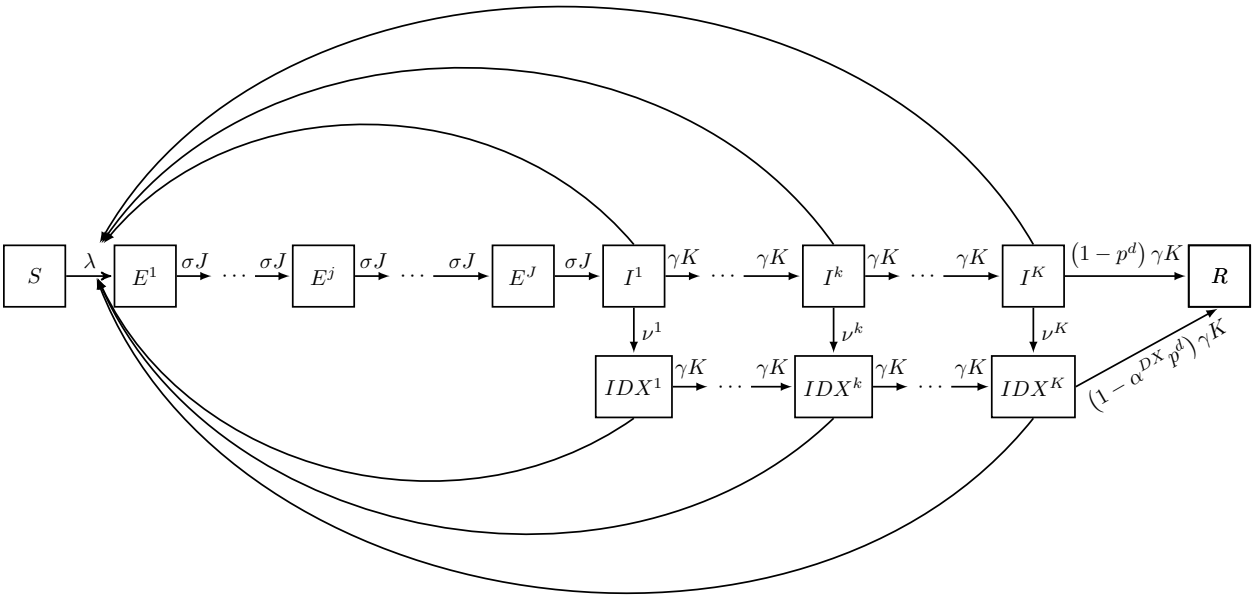

Figure 1: Diagram of the SC-COSMO model.

$$
\begin{aligned}
\frac{dS_a}{dt} &= -\left(\lambda_a + \mu_a\right) S_a, \\
\frac{dE_a^1}{dt} &= \lambda_a S_a - \left(\sigma J + \mu_a\right) E_a^1, \\
\frac{dE_a^j}{dt} &= \sigma J E_a^{j-1} - \left(\sigma J + \mu_a\right) E_a^j \qquad \text{for } j = 2, \ldots, J \\
\frac{dI_{l,a}^1}{dt} &= p_l \sigma J E_a^J - \left(\gamma K + \nu_l^1 + \mu_a\right) I_{l,a}^1, \\
\frac{dI_{l,a}^k}{dt} &= \gamma K I_{l,a}^{k-1} - \left(\gamma K + \nu_l^k + \mu_a\right) I_{l,a}^k, \qquad \text{for } k = 2, \ldots, K;\ l = 1, \ldots, L \\
\frac{dIDX_{l,a}^1}{dt} &= \nu_l^1 I_{l,a}^1 - \left(\gamma K + \mu_a\right) IDX_{l,a}^1, \\
\frac{dIDX_{l,a}^k}{dt} &= \nu_l^k I_{l,a}^{k-1} - \left(\gamma K + \mu_a\right) IDX_{l,a}^k, \qquad \text{for } k = 2, \ldots, K;\ l = 1, \ldots, L \\
\frac{dR_a}{dt} &= \sum_{l=1}^{L} \left[\left(1 - p_l^d\right) \gamma K \left(I_{l,a}^K\right) + \left(1 - \alpha_l^{DX} p_l^d\right) \gamma K \left(IDX_{l,a}^K\right) \right] - \mu_a R_a,
\end{aligned}
\tag{1}
$$

where σ is the rate at which exposed individuals in class E_a^j progress to class E_a^{j+1} for $j = 1, \ldots, (J-1)$ and from the exposed class E_a^J to the infected class $I_{l,a}^1$ where p_l is the proportion that progresses to each of the severity classes $l = 1, \ldots, L$; ν_l^k is the rate of detection from which $I_{l,a}^k$ go to $IDX_{l,a}^k$ for $k =$

$1, \ldots, (K-1)$; γ is the rate at which infectious individuals in class $I_{l,a}^k$ progress to class $I_{l,a}^{k+1}$ and also from $IDX_{l,a}^k$ progress to class $IDX_{l,a}^{k+1}$ and the recovery rate from the infectious class $I_{l,a}^K$ and from the class $IDX_{l,a}^K$ to the recovered class R_a; p_l^d the proportion of infectious individuals in class $I_{l,a}^K$ that die from COVID-19 where α_l^{DX} is a reduction in the proportion within a severity class who die from COVID-19 due to detection and appropriate health care; and μ_a represents the age-specific background mortality.

As the model is intended to simulate COVID-19 over relatively short time horizons: days, weeks, or months, it does not include rates of birth or immigration into the modeled population. If run over longer time horizons or for populations that receive substantial in-migration, its structure and equations would need to be modified accordingly.

Table 1: Description of variables, subscripts and superscripts

Symbol	Description
Subscripts	
a	Age group $\{1, \ldots, N\}$
l	Severity levels $\{1, \ldots, L\}$
Superscripts	
j	Number of exposed compartment $\{1, \ldots, J\}$
k	Number of infectious compartment $\{1, \ldots, K\}$
Variables	
λ_a	Force of infection at age group a
S_a	Susceptible in age group a
E_a^j	j-th infectious in age group a; for $j = 1, \ldots, J$
$I_{l,a}^k$	k-th infectious in age group a; for $k = 1, \ldots, K$ and $l = 1, \ldots, L$
$IDX_{l,a}^k$	k-th detected infectious in age group a; for $k = 1, \ldots, K$ and $l = 1, \ldots, L$
R_a	Recovered in age group a

1.0.1 Force of infection

The force of infection (FOI), λ, is the key quantity governing the transmission of infection within a given population, defined as the instantaneous per capita rate at which susceptibles acquire infection. FOI reflects both the degree of contact between susceptibles and infected/infectious individuals and the transmissibility of the pathogen per contact.

The age-dependent force of infection for the COVID-19 MC-SEIR model is defined as

$$\begin{aligned}\lambda_a = &\sum_{a'=1}^{N} \beta_a^1 W_{a,a'} I_{l,a'}^1 + \ldots + \sum_{a'=1}^{N} \beta_a^k W_{a,a'} I_{l,a'}^k + \ldots + \sum_{a'=1}^{N} \beta_a^K W_{a,a'} I_{l,a'}^K \\ &\sum_{a'=1}^{N} f\beta_a^1 W_{a,a'} IDX_{l,a'}^1 + \ldots + \sum_{a'=1}^{N} f\beta_a^k W_{a,a'} IDX_{l,a'}^k + \ldots + \sum_{a'=1}^{N} f\beta_a^K W_{a,a'} IDX_{l,a'}^K \quad a = 1, \ldots, N,\end{aligned} \tag{2}$$

where the transmission rate, β_a^k, describes the probability that an infected individual of age a' who is k days into his infectious period will infect a susceptible of age a per unit of time and $W_{a,a'}$ is the {a,a'} entry of the Who-Acquired-Infection-From-Whom (WAIFW) matrix, W, and $f \in [0, 1]$ is a reduction factor in transmission from infectious individuals that are diagnosed. The WAIFW matrix has N^2 elements,

representing mixing between each pair of age groups in the model. The FOI in Equation (2) is therefore a system of N equations.

$$\lambda = \beta WI + f\beta WIDX$$
$$\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix} = \beta \begin{bmatrix}W_{1,1} & W_{1,2} & \cdots & W_{1,N}\\ W_{2,1} & W_{2,2} & \cdots & W_{2,N}\\ \vdots & \vdots & \ddots & \vdots\\ W_{N,1} & W_{N,2} & \cdots & W_{N,N}\end{bmatrix} \begin{bmatrix}\sum_{k=1}^{K} I_{l,1}^{k} & \sum_{k=1}^{K} IDX_{l,1}^{k}\\ \sum_{k=1}^{K} I_{l,2}^{k} & \sum_{k=1}^{K} IDX_{l,2}^{k}\\ \vdots & \\ \sum_{k=1}^{K} I_{l,N}^{k} & \sum_{k=1}^{K} IDX_{l,N}^{k}\end{bmatrix} \begin{bmatrix}1\\ f\end{bmatrix} \quad (3)$$

The FOI $\lambda(a)$ represents the rate of disease transmission from infected/infectious people in all age groups to susceptibles in age group a [2].

1.0.2 Epidemiologic parameters

The epidemiology of COVID-19 is being elucidated at a rapid rate. Two key epidemiological parameters are the mean time spent as exposed but not yet infectious ($1/\sigma$) and the mean time that one is infectious prior to recovery ($1/\gamma$). Based on [3, 4], we assume that the mean time spent exposed is 5 days, using 3 E compartments such that the distribution of time spent exposed is realistic and consistent with the empirical data (not exponentially distributed but rather gamma distributed) [1]. Based on [4], we assume that the mean time spent infectious is 6.1 days, using 4 I compartments such that the distribution of time spent infectious is realistic. We continue to actively review the literature to further update these parameter values as well as their uncertainties.

1.0.3 Demographic parameters

To model a given population, the main groups of demographic parameters required include: 1) the total size of the population; 2) the age structure of the population (fraction of the total population in each age group); 3) the age-specific background mortality rates (i.e., mortality due to all causes other than COVID-19); and 4) the ingredients to compute adjusted population density (total land area, urban land area, and fraction of total population that is urban).

Typically, all groups of demographic parameters are obtainable from public sources. However, adjusted population density deserves specific attention here because of its role in the model inputs and therefore how we specifically compute it for this purpose. Population density is sometimes defined as total population divided by total area of a given jurisdiction in which the population resides and other times as the total population divided by the land area in which the population resides. However, in our model, adjusted population densities purpose is to help the model to reflect the intensity of mixing (contacts per unit time) that might be expected – more contacts all else equal with higher density (details described below in estimation of contact matrices).

As such, we recommend obtaining the raw ingredients to compute adjusted population density. These include land area and the fraction of the population living in urban areas. Land area (as opposed to total area) is important because we want to know how many people are living per unit area in potentially livable places. Fraction urban is important because urban areas typically occupy a small fraction of the total land area but can contain a high fraction of the population leading to very high densities (e.g., New York City has a population density in excess of 50,000 people per mile^2). Typically we would assume a small number of square miles houses the fraction the population that is urban multiplied by the total population yielding a high density and that the remaining rural population's average density is its size divided by the remaining land area. We then either model urban and rural separately or else take the average of the densities weighted

by the fraction of the population that is urban/rural which will generally be higher than just dividing total population by land area. To reflect relatively local differences, we try obtain such parameters at the finest geographic level possible which is typically county-level or state-level and likewise model at the finest possible geographic level as well.

While multiplying total population by percent urban yields the urban population, a challenge of this method for constructing the population-weighted density is knowing how much land area is urban land area. It is often possible to obtain lists of cities (e.g., for Brazil or Mexico) or census tracts (e.g., for the U.S.) and their populations and land areas by state or county. One can then compute the density of the urban population in each of the geographic areas and likewise the rural density by subtracting the urban land area from the total land area and the urban population from the total population. With these, one can then compute a population weighted average based on the relative sizes of the urban and rural populations.

We typically obtain demographic parameters from publicly available sources: For example, for counties and states in the United States, we use: county population size and age structure from 2018 [5]; state-specific life tables [6]; and county urban/rural status, land area, and population density [7]. Other specific countries on which we are implementing the model have their own sub-national sources. For other countries at a national level we use [8, 9].

1.0.4 Estimation of contact matrices

To model potential transmission between subgroups within the model, we use a contact (WAIFW) matrix approach. Entries in our contact matrix rethe number of daily sufficient contacts a person in a given age group has with people of each age group in the model. A sufficient contact is defined as one that is close, long, and/or intense enough so that transmission could occur.

A data challenge for many populations is that there are no contact matrices estimated based on empirical data sampled from them; hence we use the following approximation method based on available data to estimate contact matrices. The method involves three main inputs: 1) demographic inputs to the model (as described above); 2) contact matrices collected in 8 European countries using comparable study designs and surveys [10]; 3) Information on the functional form of how average contacts scale with population density estimated from several other human and animal diseases [11, 12].

The first step of our method is to estimate a linear regression of weighted average contact rates against the log of population density, consistent with Dalziel et al.'s and Hu et al.'s findings that as density increases from a very low number, average contact rate increases rapidly before plateauing and increasing slowly for similar increments at larger densities.

We compute the outcome of this regression as the weighted average contact rates from the each country's contact matrix W extracted from Mossong et al, whose structure is defined in 3. The weighted average contact rate

$$rate_{contacts,weighted} = \sum_{a=1}^{N} p_a \left(\sum_{a'=1}^{N} W_{a,a'} \right) \tag{4}$$

Each entry in the matrix $W_{a,a'}$ represents the number of contacts in k days by a person of age group a with people of age a'. We sum the number of contacts each age group a has over all the age groups of their contacts and then take the sum of these sums, weighting them by p_a, the proportion of the population in age group a.

We determine the population density of each country based on its urban population, urban land area, rural population, and rural land area, using these to compute the urban and rural population densities and the population-weighted average density [13]. We then take the natural logarithm of these densities for the

regression. Importantly, we do not believe that contact rates are 0 even in very sparsely populated areas as almost all people live in households and at least occasionally travel to obtain needed staples. Hence, we arbitrarily assume that at a density of 0.0183 people per mile (log density = -4), people have 1.6 contacts based on average household sizes of approximately 2.6 and subtracting the individual him/herself from this count.

The results of this regression provide us with two coefficients (the intercept and the slope for the log-density) which enable us to predict expected weighted average contacts for a population with any density.

For sub-national modeling of contact matrices for countries included in the regression analysis (i.e., those from Mossong's original study), we need to perform the following set of steps. First, we obtain need data: 1) the population age structure of the country as a whole and for each sub-national area we are modeling; 2) Determine the population density for the country as a whole and for each sub-national area.

Next, we need to normalize the contact matrix for the country as a whole. As described above, we compute the weighted average contact rate for the country as a whole from its contact matrix. We also predict its weighted average contact rate from our regression. We then compute the ratio of the predicted contact rate to the actual contact rate. We multiply each entry in the country's contact matrix by the ratio, essentially scaling the matrix to a corresponding point on the regression curve at the country' population density.

Next, to be able to map contact matrices between populations, we require a representation of the country's scaled contact matrix that is independent of its population age structure (a homogeneously mixing population whose subgroups are not all of equal size with have more people mixing with the more prevalent subgroups even without any assortative preference). To remove the age structure, we divide each age group's vector of age-specific contact rates (the $W_{a,a'}$s for a each a and for all a's) by the proportion of the population in each age group (p'_a).

To construct contact matrices for each sub-national area that reflect their population densities and age structures, we scale the normalized national contact matrix for each sub-national area separately. Specifically, for each sub-national area, we predict its weighted average contact rate and compute the ratio of this predicted rate to the predicted rate for the country as a whole. We then re-scale our normalized national matrix (the one that is scaled and no longer has its original age structure) by this ratio. Finally, we multiply the population proportions from the sub-national area (its p'_as) by the entries in the matrix to compute a matrix of appropriate contact frequency with the same underlying assortative preferences for between group contacts in a population of the sub-national area's age structure.

The final part of our method relates to extrapolating contact matrices to populations in countries and sub-national regions outside our database of contact matrices from Mossong's study (or for other regions of the world from other related studies, e.g., [14]). We first perform an additional step to determine which of the contact matrices is most appropriate to use as the "national" contact matrix in the steps outlined above. Once we have done this, we follow the rest of the steps to determine contact matrices for our populations of interest, treating them as the "sub-national" areas in our method steps described above. There are two possibilities with this approach. The first is to subjectively determine which contact matrix in our database is most appropriate (e.g., based on additional information about social and cultural patterns in a country like Brazil assume that Italy's contact patterns will be more appropriate than the contact matrices for other European countries we have available). The second is when we have some other data on contact patterns like amount of contact time per day that individuals in each age group have with all other age groups as we do for the U.S. [15], which does not differentiate how many distinct people a person has contact with in each age group but does give some notion of relative frequency/assortativity. In this latter case, we normalize all matrices whether expressed in contact frequencies or durations to place all entries on a unitless $[0, 1]$ scale. The normalization simply involves subtracting the element-wise minimum entry from all entries and then dividing by the element-wise maximum of the results of this subtraction. To determine the similarity between the country-specific normalized matrix $A_{normalized}$ and the normalized matrices $B_{normalized}$ in

our database of contact frequency matrices, we compute Frobenius norms for all of the normalized matrices $\|M_{normalized}\|_F$ and select the matrix from our database whose Frobenius norm is closest to that of the county/sub-national region we are interested in modeling:

$$B_{closest} = \underset{B \in database}{\arg\min} \; |\|A_{normalized}\|_F - \|B_{normalized}\|_F| \tag{5}$$

Once we have determined $B_{closest}$, we use this to generate the sub-national contact matrices based on the process described above.

1.0.5 Case/Infection Fatality Rates

Excess mortality risk due to COVID-19 infection (the Infection Fatality Rate [IFR]) is difficult to determine with currently available data because both the population at risk (i.e., denominator) and the number of observed COVID-19 deaths are limited to those individuals who are diagnosed with COVID-19 prior to death. What can be computed directly from these observed quantities is the Case Fatality Rate (CFR) and its extension, the age-specific CFR. However, the CFR and its age-specific versions are almost certainly overestimates of the corresponding IFRs because they do not include undiagnosed COVID-19 infections and deaths and diagnosis without active surveillance selects for more severe cases which are more likely to die. Below we describe our current approach for quantifying deaths from COVID-19.

The overall CFR (or IFR) estimated in one population is also likely to be a biased estimate for other populations if their population age structures – specifically the age structures of COVID-19 cases/infections – differ substantially. This is because the overall CFR (or IFR) is a weighted mean where the weights are number of people in each age group at risk. Hence, we prefer the age-specific versions of these quantities. An additional source of potential bias is that supportive medical care may modify the age-specific probabilities of death from COVID-19, and in some settings (e.g., low resource settings vs. wealthier settings) such differences could also be substantial. While we currently do not have a formal method for correcting for this sort of potential bias, we imagine an approach that incorporates measures/proxies of healthcare system efficacy and perhaps differences in mortality for other conditions (e.g., hospitalized pneumonia in general).

We use estimates of the age-specific IFR and CFR [16]. We also use estimates of age-specific CFR. Assuming that detected cases generally represent a more severe mix of COVID-19 infections (i..e, $CFR > IFR$), we have the following relationship for each age group a:

$$IFR_a = DR_a * CFR_a + (1 - DR_a) * UnDXFR_a, \tag{6}$$

$$UnDXFR_a = \frac{IFR_a - DR_a * CFR_a}{(1 - DR_a)} \tag{7}$$

In the equations above, DR_a refers to the age-specific detection rate (whose derivation is discussed in a different section) and $UnDXFR_a$ refers to the mortality rate from COVID-19 for people who are undiagnosed (not known cases). We solve for $UnDXFR_a$ conditional on empirically estimated values of CFR_a and IFR_a and derived values for DR_a.

Note that this approach is generalizable to more categories of disease severity within age groups provided either that there are stratified estimates of CFR and IFR by severity along with information on case detection by severity or else under assumptions made about how stratification relates to the sizes of these estimates.

1.0.6 Case detection rate

Just like fatality rates, detection rates are challenging to estimate because we lack an important component of the denominator, the total number of prevalent infections that could be detected if perfectly sensitive and specific testing were applied to all of them. The metric that is more commonly reported is the time-series

of the fraction of people tested who are positive for COVID-19 along with the time-series of the number of people tested. But the relationship between the commonly reported metrics and the detection rate is likely time-varying and confounded by a number of other factors.

To begin to get a sense of the possible confounding and complexity, one can imagine the process (ignoring age groups) as the following. At a given point in time, people who are infected with COVID-19 have a range of symptoms/severity. All else equal, we would expect people with more severe symptoms to seek testing. However, we would also expect people with other Influenza-like Illnesses (ILIs) or who may believe they were exposed to COVID-19 to also be more likely to seek testing. For both groups, those who seek testing/care will interact with the healthcare system. At a given point in time, the system has specific criteria for testing (which may or may not be strictly followed) as well as a supply constraint on the number of tests they can perform. So the number of potentially true and false positive individuals who are tested depends and hence the test positive rate is likely time-varying and challenging to determine.

We have a separate document describing our approach to putting bounds on the detection rate based upon logical constraints and available data on the testing capacity time series and other modeling.

Currently, using our bounding approach for counties in the State of California, we establish three scenarios where the cumulative detection rate (in actuality, a cumulative fraction) is either 5, 10, or 20 percent. Further we believe it likely that the detection rate was lower earlier on in the epidemic when less testing was done and supply constraints and criteria for testing were in general tighter. Hence, we assume that over the first 10 days after the first case is detected in a given county the cumulative detection rises from half its scenario level to its scenario level (e.g., from 2.5 percent to 5 percent) and then remains level from thereon. We conduct all of our calibrations and analyses using these three scenarios. In other jurisdictions where testing may be much more constrained, we have and will use lower detection rate assumptions.

We translate cumulative detection fractions for each scenario into daily detection rates using a simple exponential transformation:

$$DR_{a,t} = -\frac{\ln(1 - CumulativeDetection_{a,t})}{\frac{1}{\gamma}} \tag{8}$$

While the transformation does not incorporate competing risks (e.g., background mortality), we verify that it produces cumulative detection fractions in the model simulations that are very close to the input values for each scenario (the ratios of prevalent detected infections [cases] to total prevalent cases lagged by the average delay in detection [approximately 3 days] for each simulated day).

1.0.7 Risk of hospitalization and conditional risk of requiring ICU given hospitalization

Risks of COVID hospitalization by age are obtained from Verity et al. 2020 [16], which was modified to adjust for the age groups included in our model. For specific geographic areas/populations whose healthcare system characteristics may differ, we try to incorporate local data whenever they are available and of sufficiently high quality.

The age-specific conditional risks of ICU given hospitalization and severity were obtained from Wuan et al. 2020 [17]. In brief, Wuan et al. report the number of hospitalized patients within each of their age categories who are severe or non-severe. From these we computed the age-specific likelihood that a hospitalized patient was severe. Wuan et al. also report the overall risk of ICU among hospitalized COVID patients and the risk of ICU among hospitalized COVID patients conditional on severity, but they do not report these risks stratified by age. We assumed that these ICU risks conditional on severity were the same across all of the age groups and then validated this assumption by showing that the overall risk of ICU was the same when we weighted by the size of the age groups in the Wuan et al. study. For specific geographic areas/populations whose healthcare system characteristics may differ, we try to incorporate local data whenever they are available and of sufficiently high quality.

1.0.8 Interventions

The model currently includes a single category of non-pharmaceutical intervention (NPI) which represent a social distancing intervention. The intervention reduces contact frequency thereby reducing the force of infection. The size of the intervention's effect can vary across age group. Currently, the intervention has a start and end time. The start time is typically when shelter-in-place orders went into effect for a given population, and the end time is the end of the projection (i.e., the intervention does not end over the period being simulated). Though not currently implemented, it is easy to extend this intervention's implementation to allow its effect to vary by both age group and time since its initiation.

There are currently no direct measures on the effect of interventions on contact frequencies, duration, or intensity. A variety of sources are attempting to measure parts or proxies of this. The simplest are surveys of people asking them about how they have changed a variety of their activities, how much time they spend in their homes, and for what reasons they go out. Simple indirect measures include quantities like the amount of air population is measured by area which shows declines in commuter traffic and other production that yields emissions but does not show which types of trips are being curtailed. More sophisticated measures focus on contacts (e.g., the collocation of cellphones based on triangulation from towers or the location information of devices running various apps or the temporal proximity of credit card purchases by different purchasers at the same store) or on mobility with or without collocation (e.g., the fraction of time that someone's device is not in the location it is during typical sleeping hours or the number of devices on public transport, or how far devices travel away from home).

Average changes in these measures compared to pre-COVID levels either by individual or by geography all form estimates of components of reductions in contacts sufficient to transmit COVID. For many of these measures, at least in California counties, there are average reductions of 10-70 percent reductions or more which differ by county. What has not been reported to date are measures of individual (or small area) variance of such reductions within county, which are likely important for considering potential transmission among subgroups who must go out of their home (e.g., essential services providers) or who do not comply and then secondary transmission from these groups.

To capture geography-specific intervention effects on reducing transmission, our current approach is as follows. First, we calibrate our β_a^k terms for the force of infection to the cases time-series (accounting for the detection rate) in the period prior to the implementation of NPIs for that specific geography (as described in detail in the Calibration section). Next, we must calibrate an absolute intervention effect which will be multiplied with the β_a^k terms to reduce transmission so that the cases time-series (accounting for the detection rate) in the period after the NPI implementation is matched by the model. For this second step, we are often considering a group of multiple specific geographic areas within a state or country (e.g., the counties of California or the states of Mexico). In such a case we also often have estimates of aspects of the effectiveness of the NPIs described above for each specific geography area. As they are only estimates of one aspect of intervention effectiveness, we treat them as relative measures corresponding to the absolute intervention effect size in each geographic area. When we calibrate the absolute effect on transmission, we do so for the group of geographic areas as a whole. Essentially we simultaneously calibrate the smallest and largest absolute effects to the areas with the smallest and largest relative effects and one additional absolute effect for the 80th percentile of the relative ranking to allow for a non-linear mapping of relative effect to absolute effect. The goal of the simultaneously calibration is to find absolute effect values (min, max, and 80th percentile) that minimize the error in the predicted case time-series across all of the geographic areas for the period after the NPI interventions were implemented.

2 Calibration

Parameters of mathematical models could be either unobserved or unobservable due to different reasons (e.g., financial, practical or ethical). Model calibration is the process of estimating values for unknown or uncertain parameters of a mathematical model by matching model outputs to observed clinical or epidemiological data (known as calibration targets). The goal is to identify parameter values that maximize the fit between model outputs and the calibration targets.[18, 19]

To calibrate the transmission rate, β_a^k, described in section 1.0.1, we adopted a Bayesian approach that allowed us to obtain a joint posterior distribution that characterizes the uncertainty of both the calibration targets and previous knowledge of the parameters of interest in the form of prior distributions [20]. Prior distributions can reflect expert opinion or when little knowledge is available, these could be specified as uniform distributions or other type of distributions that reflect the domain of the intended parameters. For example, for parameters that are bounded between two values, we can use a logit distribution and parameters that are strictly positive can have log-normal distributions. We constructed the likelihood function by assuming that calibration targets, y_t, are normally distributed with mean ϕ_{i_t} and standard deviation σ_{i_t}, That is,

$$y_{i_t} \sim \text{Normal}\left(\phi_{i_t}, \sigma_{i_t}\right),$$

where ϕ_{i_t} is the model-predicted output for each type of target i at time t. To compute an aggregated likelihood measure, we added the log-likelihoods across all targets.

To conduct the Bayesian calibration, we used the incremental mixture importance sampling (IMIS) algorithm [21, 22], which has been previously used to calibrate deterministic health policy models [23, 24, 25]. An advantage of IMIS over other Monte Carlo methods, such as Markov chain Monte Carlo, is that with IMIS the evaluation of the likelihood for different sampled parameter sets could be parallelized, which makes its implementation perfectly suitable for an high-performance computing (HPC) environment.

When only interested on identifying a best-case parameter set without running the sample algorithm, we adopted a Laplace approximation where we computed the posterior mode often called the maximum a posteriori (MAP) point, by maximizing the logarithm of the posterior, and use the MAP point (instead of the mean) as an approximation of the parameter set θ. The inverse of the negative Hessian of the logarithm of the posterior can be used to measure the uncertainty of this approximation.[26, 27, 28, 29, 30]

2.1 Calibration targets

We use time-series of confirmed COVID cases as calibration targets and derived their standard errors assuming a Poisson distribution on the cumulative incidence rate over time. Times-series are typically obtained from publicly available sources or directly from governmental agencies.

3 Outcomes of interest

All the compartments in our models are variables that depend on time, t; however, we omit this index to simplify the notation. We also omit the index, l, referring to severity levels of infectious and diagnosed infectious compartments.

3.1 Demographic outcomes

3.1.1 Population

The age-group-specific population, Pop_a, is given by

$$Pop_a = S_a + \sum_{j=1}^{J} E_a^j + \sum_{k=1}^{K} I_a^k + \sum_{k=1}^{K} IDX_a^k + R_a, \tag{9}$$

and the total population across all age groups is given by

$$Pop = \sum_{a=1}^{N} Pop_a. \tag{10}$$

3.2 Epidemiological outcomes

3.2.1 Cumulative infections

The age-group-specific cumulative infections, CI_a, are given by

$$CI_a = \int_0^T \sigma J E_a^J dt, \tag{11}$$

and the total cumulative infections across all age groups is given by

$$CI = \int_0^T \sigma J \left(\sum_{a=1}^{N} E_a^J \right) dt, \tag{12}$$

where T is the analytic horizon.

3.2.2 Infections

The age-group-specific total infections at any time t, $TotI_a$, are given by

$$TotI_a = \sum_{k=1}^{K} \left(I_a^k + IDX_a^k \right), \tag{13}$$

and the total infections across all age groups is given by

$$TotI = \sum_{a=1}^{N} TotI_a, \tag{14}$$

3.2.3 Incident infections

The age-group-specific incident infections, $IncI_a$, are given by

$$IncI_a = \sigma J E_a^J, \tag{15}$$

and the total incident infections across all age groups is given by

$$IncI = \sum_{a=1}^{N} IncI_a. \tag{16}$$

3.2.4 Cumulative diagnosed infections

The age-group-specific cumulative diagnosed infections, $CIDX_a$, are given by

$$CIDX_a = \int_0^T \sum_{k=1}^{K} \left(\nu^k I_a^k\right) dt, \tag{17}$$

and the total cumulative diagnosed infections across all age groups is given by

$$CIDX = \int_0^T \sum_{a=1}^{N} \left(\sum_{k=1}^{K} \left(\nu^k I_a^k\right) \right) dt, \tag{18}$$

where T is the analytic horizon.

3.2.5 Diagnosed Infections

The age-group-specific total diagnosed infections at any time t, IDX_a, are given by

$$IDX_a = \sum_{k=1}^{K} IDX_a^k, \tag{19}$$

and the total infections across all age groups is given by

$$IDX = \sum_{a=1}^{N} IDX_a. \tag{20}$$

3.2.6 Incident diagnosed infections

The age-group-specific incident diagnosed infections, $IncIDX_a$, are given by

$$IncIDX_a = \sum_{k=1}^{K} \left(\nu^k I_a^k\right) dt, \tag{21}$$

and the total incident diagnosed infections across all age groups is given by

$$IncIDX = \sum_{a=1}^{N} IncIDX_a. \tag{22}$$

3.2.7 Total COVID19 deaths

The age-group-specific total COVID19 deaths, $DCOVID_a$, are given by

$$TotDCOVID_a = \int_0^T p_l^d \gamma K \left(I_{l,a}^K\right) + \alpha_l^{DX} p_l^d \gamma K \left(IDX_{l,a}^K\right) dt, \tag{23}$$

and the total cumulative infections across all age groups is given by

$$TotDCOVID = \int_0^T p^d \gamma K \left(\sum_{a=1}^{N} I_a^K \right) dt, \tag{24}$$

where T is the analytic horizon.

3.2.8 Known COVID19 deaths

The known COVID19 deaths are those observed from diagnosed infected cases. The age-group-specific known COVID19 deaths, $KnownDCOVID_a$, are given by

$$KnownDCOVID_a = \int_0^T \alpha_l^{DX} p_l^d \gamma K \left(IDX_{l,a}^K\right) dt, \tag{25}$$

and the total known COVID19 deaths across all age groups is given by

$$KnownDCOVID = \int_0^T p_l^d \gamma K \left(\sum_{a=1}^N I_{l,a}^K\right) dt, \tag{26}$$

where T is the analytic horizon.

References

[1] Matthew James Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, N.J., 2008.

[2] S Y Del Valle, J M Hyman, H W Hethcote, and S G Eubank. Mixing patterns between age groups in social networks. *Social Networks*, 29:539–554, 2007.

[3] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 03 2020.

[4] Xi He, Eric H. Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yujuan Guan, Xinghua Tan, and et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature Medicine*, Apr 2020.

[5] US Census Bureau. County population by characteristics: 2010-2018, Feb 2020.

[6] University of California Berkeley. United states mortality database, Feb 2020.

[7] Economic Research Service United States Department of Agriculture. Rural-urban commuting area codes, Oct 2019.

[8] United Nations Department of Economic and Population Dynamics Social Affairs. World population prospect 2019, quinquennial population by five-year age groups - both sexes. de facto population as of 1 july of the year indicated classified by five-year age groups (0-4, 5-9, 10-14, ..., 95-99, 100+). data are presented in thousands., Oct 2019.

[9] United Nations Department of Economic and Population Dynamics Social Affairs. World population prospect 2019, number of survivors by age for a hypothetical cohort of 100,000 newborns who would be subject during all their lives to the mortality rates of a given period., Oct 2019.

[10] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):1–1, 03 2008.

[11] Benjamin D. Dalziel, Stephen Kissler, Julia R. Gog, Cecile Viboud, Ottar N. Bjørnstad, C. Jessica E. Metcalf, and Bryan T. Grenfell. Urbanization and humidity shape the intensity of influenza epidemics in u.s. cities. *Science*, 362(6410):75–79, 2018.

[12] Hao Hu, Karima Nigmatulina, and Philip Eckhoff. The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*, 244(2):125 – 134, 2013.

[13] Center for International Earth Science Information Network CIESIN Columbia University. NASA Socioeconomic Data and Applications Center (SEDAC). Low elevation coastal zone (lecz) urban-rural population and land area estimates, version 2, 2013.

[14] Peter Horby, Pham Quang Thai, Niel Hens, Nguyen Thi Thu Yen, Le Quynh Mai, Dang Dinh Thoang, Nguyen Manh Linh, Nguyen Thu Huong, Neal Alexander, W. John Edmunds, Tran Nhu Duong, Annette Fox, and Nguyen Tran Hien. Social contact patterns in vietnam and implications for the control of infectious diseases. *PLOS ONE*, 6(2):1–7, 02 2011.

[15] Emilio Zagheni, Francesco C. Billari, Piero Manfredi, Alessia Melegaro, Joel Mossong, and W. John Edmunds. Using Time-Use Data to Parameterize Models for the Spread of Close-Contact Infectious Diseases. *American Journal of Epidemiology*, 168(9):1082–1090, 09 2008.

[16] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-dannenburg, Hayley Thompson, Patrick G T Walker, Han Fu, Amy Dighe, Jamie T Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich Fitzjohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-gilani, Steven Riley, Sabine Van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infectious Diseases*, 3099(20):1–9, 2020.

[17] Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu, Hong Shan, Chun-liang Lei, David S.C. Hui, Bin Du, Lan-juan Li, Guang Zeng, Kwok-Yung Yuen, Ruchong Chen, Chun-li Tang, Tao Wang, Ping-yan Chen, Jie Xiang, Shi-yue Li, Jin-lin Wang, Zi-jing Liang, Yi-xiang Peng, Li Wei, Yong Liu, Ya-hua Hu, Peng Peng, Jian-ming Wang, Ji-yang Liu, Zhong Chen, Gang Li, Zhi-jian Zheng, Shao-qin Qiu, Jie Luo, Chang-jiang Ye, Shao-yong Zhu, and Nanshan Zhong. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, pages 1–13, 2020.

[18] Tazio Vanni, Jonathan Karnon, Jason Madan, Richard G White, W John Edmunds, Anna M Foss, and Rosa Legood. Calibrating Models in Economic Evaluation: A Seven-Step Approach. *PharmacoEconomics*, 29(1):35–49, 2011.

[19] Fernando Alarid-Escudero, Richard F. MacLehose, Yadira Peralta, Karen M Kuntz, and Eva A Enns. Nonidentifiability in Model Calibration and Implications for Medical Decision Making. *Medical Decision Making*, 38(7):810–821, oct 2018.

[20] Sanjay Basu and Jeremy D. Goldhaber-Fiebert. Quantifying demographic and socioeconomic transitions for computational epidemiology: an open-source modeling approach applied to India. *Population Health Metrics*, 13(1):19, 2015.

[21] Russell J Steele, Adrian E Raftery, and Mary J Emond. Computing Normalizing Constants for Finite Mixture Models via Incremental Mixture Importance Sampling (IMIS). *Journal of Computational and Graphical Statistics*, 15(3):712–734, 2006.

[22] Adrian E. Raftery and Le Bao. Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling. *Biometrics*, 66(4):1162–1173, 2010.

[23] Nicolas A. Menzies, Djøra I. Soeteman, Ankur Pandya, and Jane J. Kim. Bayesian Methods for Calibrating Health Policy Models: A Tutorial. *PharmacoEconomics*, pages 1–12, 2017.

[24] Caleb A. Easterly, Fernando Alarid-Escudero, Eva A. Enns, and Shalini Kulasingam. Revisiting assumptions about age-based mixing representations in mathematical models of sexually transmitted infections. Supplement. *Vaccine*, 36(37), 2018.

[25] George F Sawaya, Erinn Sanstead, Fernando Alarid-Escudero, Karen Smith-mccune, Steven E Gregorich, Michael J Silverberg, Wendy Leyden, Megan J Huchko, Miriam Kuppermann, and Shalini Kulasingam. Estimated Quality of Life and Economic Outcomes Associated With 12 Cervical Cancer Screening Strategies: A Cost-effectiveness Analysis. *JAMA Internal Medicine*, 179(7):867–878, 2019.

[26] Luke Tierney and Joseph B Kadane. Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

[27] Bradley P. Carlin and Thomas A. Louis. *Bayesian Methods for Data Analysis*. CRC Press, third edition, 2009.

[28] Tan Bui-Thanh, Omar Ghattas, James Martin, and Georg Stadler. A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems Part I: The Linearized Case, with Application to Global Seismic Inversion. *SIAM Journal on Scientific Computing*, 35(6):2494–2523, 2013.

[29] Marco A. Iglesias, Kody J H Law, and Andrew M. Stuart. Evaluation of Gaussian approximations for data assimilation in reservoir models. *Computational Geosciences*, 17(5):851–885, 2013.

[30] Matthias Morzfeld, Xuemin Tu, Jon Wilkening, and Alexandre J. Chorin. Parameter estimation by implicit sampling. *Communications in Applied Mathematics and Computational Science*, 10(2):205–225, 2015.