

OCHA-BUCKY: A COVID-19  
MODEL TO INFORM  
HUMANITARIAN OPERATIONS  
MODEL METHODOLOGY

OCTOBER 2020

THE CENTRE FOR HUMANITARIAN DATA  
in collaboration with THE JOHNS HOPKINS  
UNIVERSITY APPLIED PHYSICS LABORATORY



OCHA

centre for humdata



JOHNS HOPKINS  
APPLIED PHYSICS LABORATORY

## Contents

<b>Contents</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
<b>2 Model Description</b>	<b>4</b>
2.1 Model Overview . . . . .	4
2.2 Model Parameterization . . . . .	6
2.2.1 Disease/Healthcare Parameters . . . . .	6
2.2.2 Approximated and Derived Parameters . . . . .	8
2.3 Data Sources . . . . .	9
2.3.1 Population Data . . . . .	9
2.3.2 Vulnerability Parameters . . . . .	9
2.3.3 Mobility Data and Matrix Generation . . . . .	10
2.3.4 COVID-19 Cases and Deaths . . . . .	11
2.3.5 Contact Matrix . . . . .	11
2.3.6 Non-pharmaceutical Interventions (NPIs) . . . . .	11
<b>3 Model Initialization and Output</b>	<b>12</b>
<b>4 Model Open Source Code and Documentation</b>	<b>13</b>
<b>5 References</b>	<b>14</b>

## Acknowledgements

This model methodology overview was produced by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data, in partnership with the Johns Hopkins University Applied Physics Laboratory (JHUAPL). Contributions to the model and corresponding documentation were given by the following members of the JHUAPL Team: Matthew Kinsey, Kate Tallaksen, R.F. Obrecht, Laura Asher, Cash Costello, Michael Kelbaugh, Howard Burkom, David Maxson, Shelby Wilson, Jessica Dymond, and Jason Lee. Contributions were also given by the following members of the OCHA Centre for Humanitarian Data: Leonardo Milano, Josée Poirier, and Tinka Valentijn, and Monica Turner (on secondment from MapAction). This work was made possible with funding from The Rockefeller Foundation and the Governments of the Netherlands, Germany, Belgium, and the United Kingdom. For more information, please contact the Centre for Humanitarian Data at [centrehumdata@un.org](mailto:centrehumdata@un.org).

# 1 Introduction

Severe acute respiratory syndrome-coronavirus 2 (SARS-CoV2) emerged in Wuhan, China, in November 2019. By October 2020, this virus had resulted in more than 41.5 million cases of COVID-19 and 1 million deaths. The majority of COVID-19 cases currently being reported are predominantly in developed countries, such as the United States, Brazil, Russia, and France. Despite the increasing number of models and studies on COVID-19, there is very little information available to inform humanitarian response interventions, the need for which may be unprecedented, particularly where infrastructure is lacking to effectively prevent spread of transmission and treat affected patients.

In late 2019, the United Nations (UN) Office for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data created a new workstream for predictive analytics. This was based on demand from OCHA's leadership to "use data, and especially the tools of predictive analytics to get ahead, to be more anticipatory, to predict what is about to happen and to trigger the response earlier." This ambition aligns with the overall goal of the Centre, which is to increase the use and impact of data in the humanitarian sector. The COVID-19 pandemic has brought into stark focus the need for data and the value of models to inform response strategies. Anticipatory action is no longer an abstract idea but something populations are actively doing by staying home and increasing the number of hospital beds to protect the most vulnerable populations.

Epidemic forecasting is one tool through which we can gain an understanding of the final outbreak size and indicators of when the COVID-19 epidemic peaks in a country. This provides decision-makers with the capability to plan, surge, and manage resources during a pandemic. UN OCHA and the Johns Hopkins University Applied Physics Laboratory have therefore established a partnership to inform COVID-19 strategies for humanitarian interventions by both national authorities and the humanitarian community in selected high-priority countries, resulting in increased technical capacity to predict new and compounded humanitarian needs, and use of data science to arrive at interventions to mitigate them.

This partnership developed a series of adjustments to a novel COVID-19 model (JHUAPL-Bucky) that incorporates different vulnerability factors to provide insights on the scale of the crisis in priority countries at national and sub-national levels, how different response interventions are expected to impact the epidemic curve, and the duration of the crisis in specific locations. The resultant model (OCHA-Bucky) stratifies COVID-19 dynamics by age and population vulnerability. Input to the model consists of geographically distributed COVID-19 cases and deaths, as well as attributes such as inter-regional mobility, population vulnerability, nonpharmaceutical interventions (NPIs), and social contact matrices. Model output consists of future projections of these same quantities, as well as severe cases (defined as a proportion of total cases). The model considers both inter-regional mobility of the population and time-varying NPIs. OCHA-Bucky has been used to provide weekly projections to six OCHA country offices: Afghanistan, the Democratic Republic of Congo, Iraq, Somalia, Sudan, and South Sudan.

The results of this model, when applied in the context of the United States, have been included in the [Centers for Disease Control and Prevention \(CDC\) COVID-19 Mathematical Modeling Forecasting Ensemble \[1\]](#). Here, we detail the modifications of the JHUAPL-Bucky model used to shift results to the context of countries in receipt of humanitarian aid.

A number of critical complementary components must be modified in order to perform accurate disease modeling within this context. Namely, these include:

- Estimating disease parameters;
- Acquiring data sources that accurately reflect both the current and historical states of the outbreak;
- Estimating mobility within a country;
- Estimating the impact of secondary/tertiary factors on the vulnerability of a given population;
- Estimations of the effects of measures taken by individuals and governments to curb the spread of the disease; and

- Estimations of the effects of the above at both the sub-national and national levels.

The following sections describe the OCHA-Bucky model in detail. Section 2.1 gives an overview of the model and its components. Further details related to parameter estimation, data sourcing, and model initialization are given in Sections 2.2 and 2.3, respectively. Section 3 details the model input and output. Lastly, this model is publicly available; details about how to access the model as well as the corresponding documentation are given in Section 4.

## 2 Model Description

The JHUAPL-Bucky model is a COVID-19 metapopulation compartment model initially designed to estimate medium-term (on the order of weeks) case incidence and healthcare usage at the second administrative (admin-2, ADM2) level (counties in the United States; cities or districts in various countries). These ADM2 regions are all coupled using mobility information to approximate the inter- and intra-regional contacts between the members of the populations. Using the historical case and death data, local demographic data, and a set of parameters derived from empirical studies (Section 2.2.1), the model infers a number of localized features (given in Table 1) that are related to the spread of COVID-19. Projecting forward in time, Bucky then uses an age-stratified compartment model to estimate the case load as well as provide outputs relating to the healthcare burden of each locality. These time forecasts are performed numerous times (Monte Carlo experiments), with each individual simulation using minor modifications to the input parameters at random, scaled to the uncertainty of the estimates. The resulting collection of simulations is then used to obtain probabilistic estimates for all output variables. A fork of JHUAPL-Bucky, the OCHA-Bucky model, additionally includes a set of new features to properly handle the unique aspects of epidemic modeling in countries beyond the United States. To this end, the OCHA-Bucky model includes a series of factors related to the relative vulnerability of local populations.

### 2.1 Model Overview

At its base, the OCHA-Bucky model is a spatially distributed SEIR model. SEIR models are a class of deterministic models used to demonstrate how infectious diseases that are spread by person-to-person transmission in a population. The simplest versions of such models are systems of ordinary differential equations and are analysed mathematically [2].

Within the context of an SEIR model, disease dynamics are modeled over time by moving the population through a series of compartments (otherwise known as "bins" or "states"). Those states are as follows:

- Susceptible (S): the fraction of the population that could be potentially subjected to the infection;
- Exposed (E): the fraction of the population that has been infected but does not show symptoms yet;
- Infectious (I): the fraction of the population that is infective after the latent period;
- Recovered (R): the fraction of the population that has been infected and recovered from the infection.

The total population is represented by the sum of the compartments. Basic assumptions of this type of model:

- Once the model is initialized, no individuals are added to the susceptible group. It follows that births and natural deaths are unaccounted for, migration into and out of the region is frozen for the duration of a simulation, and none of the population has been vaccinated or is immune to the pathogen;
- The population within each strata is uniform, and each pair of individuals within the strata are equally likely to interact;
- The probability of interaction between individuals in the population is not rare;
- Once infected, an individual cannot be reinfected with the virus.

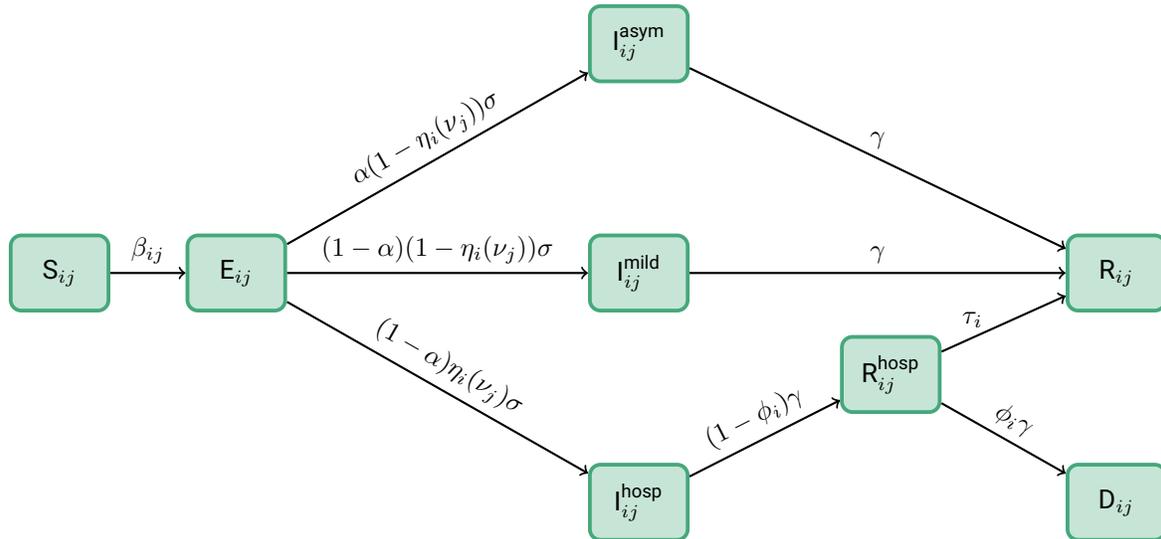


Figure 1: Model Diagram. See Table 1 for a description of the compartments and transition parameters. Note that the compartments  $E$ ,  $I^{\text{asym}}$ ,  $I^{\text{mild}}$ ,  $I^{\text{hosp}}$  and  $R^{\text{hosp}}$  are gamma-distributed with shape parameters specified in the configuration file.

The OCHA-Bucky model consists of a collection of coupled and stratified SEIR models. Since COVID-19 exhibits heavily age-dependent properties, wherein a majority of severe cases are in older individuals, SEIR models are stratified via the age demographic structure of a geographic region to obtain accurate estimates of case severity and deaths. Additionally, to model the spatial dynamics of COVID spread, we consider a set of SEIR sub-models at the smallest geographic level for which we have appropriate data. (See Section 2.3 for a discussion of available data.)

The basic structure of the model is displayed in Figure 1. Age is denoted by index  $i$ , and geographic regions are denoted by index  $j$ . Within each strata, OCHA-Bucky models the susceptible and exposed populations, followed by one of three possible infected states: asymptomatic ( $I^{\text{asym}}$ ), mild ( $I^{\text{mild}}$ ), and severe ( $I^{\text{hosp}}$ ). Members of the population who are either asymptomatic or exhibit mild symptoms recover from the virus at a rate  $\gamma$ . Those who exhibit severe symptoms and are in need of healthcare support will either recover after a period of illness at rate  $1/\tau_i$  or expire as a result of the virus at rate  $\phi_i\gamma$ .

A critical component of the OCHA-Bucky model is the parameterization of the model (Section 2.2). A number of parameters must be derived and/or estimated from their original data sources. These include, but are not limited to, those listed in Table 1 as well as local estimates of local case doubling time, case reporting rate, case fatality rate, and case hospitalization rate. Further details of these quantities as well as how they are estimated are given in Section 2.2. All parameter estimation for the model includes the basic assumption that, once estimated and initialized, these parameters remain constant during the simulation period.

Coupling individual age and geographically stratified sub-models occurs across a number of dimensions including disease state and pre-existing vulnerability. Sub-models are coupled together using both the spatial mobility matrix (see Section 2.3.3) and age-based contact matrices (see Section 2.3.5). Modeling of the overall interaction rates between geographic locations and age groups is an important component in accurately modeling non-pharmaceutical Interventions (NPIs). OCHA-Bucky accounts for the implementation of NPIs (e.g., school closures, border closures, face mask wearing) via modifying either the social contact matrices or the basic reproductive number,  $R_0$ . Further details are given in Section 2.3.6.

Together, these components contribute to a model that is adaptable to a number of contexts. OCHA-Bucky is calibrated to the uncertainties in the case data and the disease parameters, leading to a model that is robust to the quality and resolution of available input data.

Variable	Description
$S_{ij}$	Proportion of individuals who are susceptible to the virus
$E_{ij}$	Proportion of individuals who have been exposed to the virus
$I_{ij}^{\text{hosp}}$	Proportion of individuals that are exhibiting severe disease symptoms, and are in need of a healthcare facility
$I_{ij}^{\text{mild}}$	Proportion of individuals that are exhibiting mild disease symptoms
$I_{ij}^{\text{asympt}}$	Proportion of individuals who are infected but asymptomatic
$R_{ij}$	Proportion of individuals who have recovered from the virus and are no longer capable of infecting other individuals
$R_{ij}^{\text{hosp}}$	Proportion of individuals who have recovered from the virus after a period of time in a healthcare facility
$D_{ij}$	Proportion of individuals who have succumbed as a direct result of the virus
Parameter	Description
$\beta_{ij}$	Force of infection on a member of age group $i$ in location $j$
$1/\sigma$	Viral latent period
$\alpha$	Rate of infections that are asymptomatic
$\eta_i(\nu_j)$	Fraction of cases necessitating healthcare facilities for age group $i$ as a function of the local vulnerability index, $\nu_j$
$\phi_i$	Case fatality rate for age group $i$
$1/\gamma$	Infectious period
$\tau_i$	Recovery period from severe infection for age group $i$

Table 1: Description of OCHA-Bucky model variables and parameters corresponding to Figure 1.

## 2.2 Model Parameterization

### 2.2.1 Disease/Healthcare Parameters

The CDC has published pandemic planning scenarios [3] that contain recommended parameters describing biological and epidemiological factors. Of these five planning scenarios, OCHA-Bucky uses scenario five, which contains the CDC's current best estimates for disease severity and transmission. These parameters are described in detail, based on information available from the CDC, and summarized in Table 2.

Parameter Description	Bucky Variable Name	Value (Interquartile Range)
Fraction of infections that are asymptomatic, $\alpha$	asym_frac	0.4
Relative infectiousness of asymptomatic individuals	rel_inf_asym	0.75
Fraction of transmission prior to symptom onset	frac_trans_before_sym	0.5
Mean serial interval ( <i>days</i> )	$T_s$	6 (5.7)
Mean generation interval ( <i>days</i> )	$T_g$	7 (5.5, 8.5)
Case fatality ratio, $\phi_i$	CFR	0-49 years : 0.0005 50-64 years: 0.002 65+ years: 0.013
Case hospitalization ratio, $\eta_i$	CHR	0-49 years : 0.017 50-64 years: 0.045 65+ years: 0.074
Time from symptom onset to hospitalization ( <i>days</i> )	I_TO_H_TIME	0-49 years : 6 50-64 years: 6 65+ years: 4
Duration of hospitalization ( <i>days</i> )	H_TIME	0-49 years : 4.9 50-64 years: 7.6 65+ years: 8.1
Time between death and reporting ( <i>days</i> )	D_REPORT_TIME	0-49 years : 7.1 50-64 years: 7.2 65+ years: 6.6

Table 2: CDC-Recommended Parameters as described in Scenario 5 of [3].

### 2.2.1.1 Disease Transmission

The following parameters describe the transmissibility of the virus. The percentage of infections that are asymptomatic ( $\alpha$ ) refers to the percentage of infections that will never develop symptoms. This is a difficult parameter to estimate due to logistical complications (individuals would need to be tested to ensure they remain asymptomatic while infectious) and because the level of asymptomatic infections varies by age. The best estimate for this parameter is the midpoint between the lower bound of [4], the upper bound of [5], which corresponds to the estimates from [6].

The relative infectiousness of asymptomatic individuals compared to symptomatic individuals (rel\_inf\_asym) is calculated using upper and lower bounds on the difference in viral dynamics between asymptomatic and symptomatic cases. The lower bound is derived from data indicating that more severe cases have higher viral loads [7] and a study that indicates symptomatic cases shed for longer and have higher viral loads than asymptomatic cases [8]. Other studies indicate that both symptomatic and asymptomatic cases have similar duration and viral shedding [9], which is used as the upper bound.

The final parameter relating to disease transmission is the fraction of transmission prior to symptom onset (frac\_trans\_before\_sym), which corresponds to the percentage of new cases caused by transmission from an individual before they become symptomatic. The lower bound is derived from [10], with the upper bound derived from [11].

### 2.2.1.2 Disease Characteristics and Severity

The mean serial interval,  $T_s$ , is the time in days from exposure to onset of symptoms and is taken from [12]. The mean generation interval,  $T_g$ , is the period of time (in days) between symptom onset for one individual and symptom onset for a person they have infected. This value is from [10].

The case fatality ratio (CFR) is the number of individuals who will die of the disease; the case hospitalization-severity ratio (CHR) corresponds to the number of cases that are severe and necessitate hospitalization. Within the context of the United States, this ratio corresponds to the individuals admitted to a hospital. In a context where access to medical care is limited, this ratio corresponds to the ratio of individuals who exhibit severe disease symptoms.

Hospital-related parameters are derived using data from COVID-Net [13] and the CDC's Data Collation and Integration for Public Health Event Response (DCIPHER). All data is taken from the period between March 1, 2020 to July 15, 2020 unless otherwise noted. The time it takes from symptom onset to hospitalization in days is denoted by  $I\_to\_H\_time$ . The number of days an individual will be hospitalized is  $H\_TIME$ . Finally, the number of days between death and reporting is  $D\_REPORT\_TIME$ .

### 2.2.2 Approximated and Derived Parameters

Unless otherwise noted, the model parameters that are derived from input parameters<sup>1</sup> and the historical data are estimated as the average value across the entire population of a region (i.e., they are not age-stratified). This limitation results from having a lack of age-stratified historical case and death data in most regions. Additionally, some parameters are only derived at the national (ADM0) level due to sparse reporting at the sub-national level (see footnotes).

- Local mean Case Fatality Ratio,  $\phi_j$ ;

We must begin by estimating the local overall CFR for each region based on the local demographics. Since the fatality rate for COVID-19 is highly age dependent, this overall CFR will vary dramatically between region with different age structure [15]. By contracting the CFR as a function of age with the local age demographics of each region, we can obtain the region overall estimate for the CFR,

$$\phi_j = \phi^i N_{ij}.$$

- Case Reporting Rate,  $\rho_j$ ;

In order to estimate the local fraction of COVID-19 cases that are reported in the historical data, OCHA-Bucky uses a method similar to Russel et al. [16] based on the deviation from the expected CFR. Using the derived CFR and the historical case and death data, the OCHA-Bucky model first estimates the overall case reporting rate for symptomatic cases at the ADM2 level. This is done using the assumption that the ratio of the expected CFR for each region based on the local demographics,  $\phi_j$  to the CFR calculated from the cumulative historical data,  $\phi_j^{hist}$ <sup>2</sup> i.e.

$$\rho_j = \frac{\phi_j}{\phi_j^{hist}} = \frac{\phi^i N_{ij}}{\phi_j^{hist}}.$$

In calculating  $\phi_j^{data}$ , a lag in the reporting of deaths equal to the given input parameter is assumed ???. The estimated case reporting rate for each of the last X number of days is then calculated; the mean value in each region is used.

- Doubling Time,  $T_D$ ;

<sup>1</sup>Throughout our discussion in this section we make use of Einstein notation[14] to describe the mathematics. This is owed to the fact that most of the variables are tensors by virtue of the fact that they are stratified in multiple ways (e.g., age and locality)

Most importantly, we use the notation of a tensor contraction to represent the summed multiplication along an axis. i.e.  $c_i x^i = \sum_i c_i x^i$  for all indices that appear as both upper and lower indices in a single term but are otherwise undefined.

<sup>2</sup>In OCHA-Bucky, the value is the CFR derived from the historical data,  $\phi_j^{hist}$  is estimated at the the country level from the WHO country level case/death history and used uniformly throughout all sub-regions. This is done to circumvent potential data issues present in the local case data in various countries.

Rather than explicitly include a parameter controlling the transmission rate, Bucky uses the recent historical data to estimate transmission at the local level<sup>3</sup>. The local doubling time  $T_{D,j}$  is estimated by first considering the historical data prior to the date in which the simulation will begin. Historical data is used to calculate the doubling time for the seven days prior to simulation. These historical doubling times are then averaged to yield the value of  $T_{D,j}$  for a particular simulation. Once these values are calculated, methods similar to those described in [17] can be used to estimate the latent period,  $1/\sigma$ , infectious period,  $1/\gamma$ , as well as an estimate of the effective  $R_0$  value. These methods also yield the shape parameters of the gamma-distributions corresponding to E,  $I^{\text{asym}}$ ,  $I^{\text{mild}}$ , and  $I^{\text{hosp}}$  compartments.

- Case Hospitalization Ratio with Vulnerability,  $\eta_i(\nu_j)$ ;

To account for the inclusion of the vulnerability index in OCHA-Bucky, we rescale the population fraction requiring healthcare support (the CHR,  $\eta_i$ ) based on the local fraction of vulnerable population,  $\nu_j$ :

$$\eta_i(\nu_j) = (1 - \nu_j)\eta_i + V\nu_j\eta_i$$

where  $V$  is the relative severity among the vulnerable population (See 2.3.2 for how this is estimated).

- Contact rate between susceptible and infected individuals,  $\beta_{ij}$ ;

The quantity  $\beta_{ij}$  represents the average number of contacts per person per time, multiplied by the probability of disease transmission in a contact between a susceptible and an infectious subject. Within Bucky, this consists of the rate that an individual in age group  $i$  in location  $j$  gets infected by the virus. This depends on three quantities:

1. The infectivity rate of the virus;
2. The average contact between someone from age group  $i$  and individuals across all other age groups;
3. The average rate of mobility into location  $j$  from across all other locations in the simulation.

## 2.3 Data Sources

### 2.3.1 Population Data

In order to estimate the age- and sex-disaggregated population for each ADM2 region two different datasets are combined:

- Administrative boundaries shapefiles at the ADM2 level from the [Humanitarian Data Exchange \(HDX\)](#);
- United Nations -adjusted [WorldPop](#) age and sex disaggregated population raster data.

Input consists of raster data with the administrative region polygons for the province/state (ADM1) and district/county (ADM2) to get total populations for each region, stratified by age and sex. Then, values are scaled so that the total population is consistent with the regional UN-adjusted population.

### 2.3.2 Vulnerability Parameters

The fraction of the population in an ADM2 region that is particularly vulnerable to COVID-19 is estimated by calculating the proportion of the population effected by those factors described in Sections 2.3.2.1 – 2.3.2.4. To avoid confounding factors contributing to a population's increased vulnerability, we select the maximum of these proportions and consider this to be the proportion of the population that is at increased vulnerability to the effects of COVID-19. Those in the population that are considered vulnerable have a higher risk of a severe case or death (and hence a higher CHR and CFR) than non-vulnerable members of a region. Given literature estimates, the vulnerable population is given a CHR that is 1.5 times greater than the average CHR of a region (calculated via age and gender demographic information).<sup>4</sup>

<sup>3</sup>In OCHA-Bucky, this estimation is performed at the nation level using WHO case data, similar to the estimation of reported CFR.

<sup>4</sup>The CFR carries the CHR scalar multiplication forward since CFR is estimated as a proportion of those with severe cases.

### 2.3.2.1 Urban/Rural disaggregation

Although not a risk factor on its own, there are several vulnerabilities that scale differently with urban and rural populations. The COVID-19 crisis has impacted urban areas significantly more than rural areas, due primarily to mobility and proximity to other ADM2 regions, which increase transmission rates. Urban/rural stratification is available in raster format from WorldPop. To calculate the urban-rural fraction for each ADM2 region, data is taken from [Global Human Settlement Layer](#). Based on the classifications available in [18] anything denser than suburban (class 21 or above) is considered to be urban, and the rest to be rural.

### 2.3.2.2 Food insecurity

There is evidence that weight impacts the severity of influenza and other respiratory viruses [19]. We have used food security as an indicator of population weight and other factors that effect the populations baseline ability to respond to respiratory illness. Hence, we scale our vulnerability factor,  $\nu_j$  by levels of food security within the region. Food insecurity data is obtained from the Integrated [Food Security Phase Classification \(IPC\) Global Platform](#). The proportion of individuals living in regions at IPC level 3 or greater are considered at-risk. The data tend to have a mix of information at the ADM1 and ADM2 levels; the most granular data is utilized for this calculation.

### 2.3.2.3 Indoor air pollution - indoor cooking fuels

The use of indoor cooking fuels is a risk factor for COVID-19 whose prevalence depends on an urban or rural setting. Data on the use of solid fuels is obtained from the [World Health Organization](#). The total fraction of the population using solid fuels is calculated by re-scaling at each location the urban/rural estimates from WHO by the corresponding urban/rural population fraction estimated according to 2.3.2.1. There is evidence linking indoor air pollution and respiratory illness [20]. Based on the available literature, we estimated that those that used solid cooking fuels are 1.8 times more likely to develop a severe infection.

### 2.3.2.4 Medical comorbidities

Medical co-morbidities, such as diabetes, cardiovascular disease, and tuberculosis, were considered as possible additional vulnerability parameters for cases and deaths. However, the OCHA-Bucky model does not use these parameters, as they are highly correlated with age and gender, and have little additional impact on vulnerability and death rate estimates.

## 2.3.3 Mobility Data and Matrix Generation

A required input of the model is the amount of interaction that occurs between different regions at the sub-national level. Mobility matrix data is used to simulate disease spread across different administrative units of a country. The JHUAPL-Bucky model assumes contact between various regions through using a combination of various sources of cell-phone based mobility estimates. In the OCHA-Bucky model, due to a lack of mobile phone based estimates, contact between ADM2 regions is approximated through an analysis of road network data.

Road density data from [Humanitarian OpenStreetMap \(HOTOSM\)](#) is used to estimate the strength of the connection between all ADM2 regions considered. To create the mobility matrix using roads data, a roads shapefile is read, and roads that intersect with at least one ADM2 border are kept. For each road, a list is then created of which ADM2 regions the road passes through. Completing this gives a list of region to region pairs, for which geographic distance (between ADM2 geographic centroids) and road connectivity are calculated. To calculate road connectivity, road network lines are obtained from [HOTOSM on HDX](#). The road classification system from HOTOSM is used to apply weights for each road type, calculated by multiplying the estimated typical speed and number of lanes for each road class; connectivity is defined as the sum of all roads connecting the region pair.

Road connectivity, the motor vehicle ownership fraction, and the reciprocal geographic distance between regions are all then multiplied together to give estimates of connectivity between ADM2 pairs. Adjacent regions with many large road connections necessarily have higher connectivity values than those which have few small or even no road connections. Long-range or non-adjacent interactions between regions are also considered, especially in cases where a road crosses more than two regional boundaries. However, the diminishing weakness of non-adjacent ADM2 pairs is captured through the reciprocal of the geographic

centroids; higher values indicate closer regions.

Within the model, the probability of the disease to spread from one ADM2 region to another is proportional to the estimated weights in the mobility matrix.

### 2.3.4 COVID-19 Cases and Deaths

Sub-national historical daily number of COVID-19 cases are obtained from HDX. In most countries of interest, the source for COVID case counts is the nation's Ministry of Public Health. Additional federal (ADM0) historical data is also obtained from the WHO. Data is usually available at the ADM1 level. To obtain ADM2-level case counts, ADM1 case counts are scaled using the ADM2 population.

### 2.3.5 Contact Matrix

Contact matrices quantify how much people from different age groups interact, and are extracted from Prem et al. [21]. The contact matrices used are taken from the home, school, work, and other locations. For some countries, the contact matrix is not directly available. In these instances, a country in the same region and with similar socioeconomic indicators is used as a proxy.<sup>5</sup>

### 2.3.6 Non-pharmaceutical Interventions (NPIs)

NPIs are mitigations, apart from getting vaccinated and taking medicine, that people and communities can take to help slow the spread of communicable diseases. As a vaccine for COVID-19 has yet to be deployed, NPIs are among the the best strategies for controlling the spread of the current COVID-19 virus. The structure of the OCHA-Bucky model allows for the incorporation of NPIs via the modification of a combination of the following : the basic reproduction number, local contact matrices, and inter-regional mobility matrices.

For each country an initial list of NPIs was obtained from the ACAPS [COVID-19 Government Measures Dataset](#). This dataset is complemented with additional qualitative information from in-country stakeholders. The estimated compliance level are tailored to specific countries.

#### 2.3.6.1 Implementation of Nonpharmaceutical Interventions

NPIs are categorized and implemented in OCHA-Bucky based on their classification into three categories:

1. Contact-Matrix Based NPIs;

These NPIs are those that effect only certain age groups within the total population. These NPI effect the ratios relating the components of the contact matrices. The NPI that fall under this category are:

- School closure
- Shielding elderly

2. Mobility Based NPI;

This classification is for those NPI that lead to changes in mobility/movement between administrative districts (as opposed to movement within an administrative district). The NPI that fall under this category are :

- Closing of borders, ports, and/or international flights
- Restricting inter-regional movement

3. Reproduction Number Based NPI;

This classification is for those NPI that have an effect on the overall scaling of transmissibility. It encompasses both intra-regional measures to reduce transmission as well as national level initiatives designed to reduce transmission throughout the country. The NPI that fall under this category are :

- Social distancing;

---

<sup>5</sup>The model from [The London School of Hygiene and Tropical Medicine](#) also utilizes this strategy.

- Face mask wearing;
- Installation of hand washing stations;
- Reduction of size of public gatherings;
- Closing businesses;
- Partial lockdown;
- Awareness campaigns (e.g., vaccination programs).

A summary of the NPIs that are currently implemented in OCHA-Bucky are given in Table 3. This table includes the classification, effects, and sources that are currently being used to approximate the effects of various NPI.

With the current implementation, we have the ability to distinguish between the effects of NPI within the categories mentioned above. For the case in which multiple NPI within category III are implemented, we have implemented a value-added approach to calculating their effectiveness in reducing the basic reproduction number. In this case, we calculate the reduction in  $R_0$  based on the number of NPIs in place. If 1 NPI is in place,  $R_0$  is reduced by 40%. If 2 NPI are in place,  $R_0$  is reduced by 60%. If 3 or more NPI are in place, then  $R_0$  is reduced by 70%.

NPI Classification	Effect in Model	Mean Reduction (SD)	Source
Contact-based: School Closure	Reduce contact between school aged groups and increase the contacts in the home environment	~44% reduction in overall community transmission	[22]
Mobility-based	Reduction in mobility between regions	60% (10)	[23][22]
Reproduction Number-based	60-85% reduction in overall community transmission	72.5% (6.25)	[24][25]

Table 3: Summary of Effects of NPI on model components

### 3 Model Initialization and Output

Using the data sources and methods described in Sections 2.2 and 2.3, we are able to both estimate values of parameters initialize the values of all the compartments in the model on the last date of the historical data. This will serve as the initial state of our simulation.

As output, OCHA-Bucky generates one data file per Monte Carlo run. The files contain data at the ADM2 level and are indexed by data and run ID. This data is post-processed to combine data across all the dates and simulations. It can then be aggregated at desired geographic levels. A separate file is created for each administrative level, with each row indexed by date, admin ID, and quantile. The columns of this output file are described in Table 4.

Index name	Description
adm0, adm1, or adm2	The adm ID corresponding to the geographic level
date	The date
quantile	Quantile value
Column name	Description
case_reporting_rate	Case reporting rate
active_asymptomatic_cases	Current number of actively infectious but asymptomatic cases
cumulative_cases	Cumulative number of cumulative cases (including unreported)
cumulative_deaths	Cumulative number of deaths
cumulative_deaths_per_100k	Cumulative number of deaths per 100,000 people
cumulative_reported_cases	Cumulative number of reported cases
cumulative_reported_cases_per_100k	Number of reported cumulative cases per 100,000 people
current_hospitalizations	Number of active severe cases
current_hospitalizations_per_100k	Number of active cases per 100,000 people
current_icu_usage	ICU bed usage*
current_vent_usage	Current ventilator usage*
total_population	Population
daily_cases	Number of daily new cases (including unreported)
daily_deaths	Number of daily new deaths
daily_hospitalizations	Number of daily new severe cases
daily_reported_cases	Number of reported daily new cases
doubling_t	Local doubling time as estimated from the historical data
R_eff	Local effective reproductive number

Table 4: Model output column descriptions.

\*indicates model outputs that are not calibrated/validated for the context of low-middle income countries.

## 4 Model Open Source Code and Documentation

Further documentation and resources can be found in the following locations.

- For documentation related to the United States version of the model (JHUAPL-Bucky) :
  - The JHUAPL-Bucky model is available on Github at <https://github.com/mattkinsey/bucky>.
  - Documentation for JHUAPL-Bucky is available at <https://docs.buckymodel.com/en/stable/>.
- For the latest resources on the international version of the model (OCHA-Bucky) :
  - The OCHA-Bucky model is available on Github at <https://github.com/OCHA-DAP/pa-ocha-bucky>.
  - Documentation for OCHA-Bucky is available at <https://ocha-bucky.readthedocs.io/en/latest/index.html>

## 5 References

- [1] C. for Disease Control, Prevention, et al., "Covid-19 mathematical modeling forecasts," URL: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html> Accessed October, vol. 18, 2020.
- [2] H. W. Hethcote, *Three Basic Epidemiological Models*, pp. 119–144. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989.
- [3] C. for Disease Control, Prevention, et al., "Covid-19 pandemic planning scenarios," URL: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html> Accessed May, vol. 29, 2020.
- [4] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M.-L. McLaws, and P. Glasziou, "Estimating the extent of true asymptomatic covid-19 and its potential for community transmission: systematic review and meta-analysis," Available at SSRN 3586675, 2020.
- [5] P. Poletti, M. Tirani, D. Cereda, F. Trentini, G. Guzzetta, G. Sabatino, V. Marziano, A. Castrofino, F. Grosso, G. Del Castillo, et al., "Probability of symptoms and critical disease after sars-cov-2 infection," *arXiv preprint arXiv:2006.08471*, 2020.
- [6] D. P. Oran and E. J. Topol, "Prevalence of asymptomatic sars-cov-2 infection: A narrative review," *Annals of Internal Medicine*, 2020.
- [7] Y. Liu, L.-M. Yan, L. Wan, T.-X. Xiang, A. Le, J.-M. Liu, M. Peiris, L. L. Poon, and W. Zhang, "Viral dynamics in mild and severe cases of covid-19," *The Lancet Infectious Diseases*, 2020.
- [8] J. Y. Noh, J. G. Yoon, H. Seong, W. S. Choi, J. W. Sohn, H. J. Cheong, W. J. Kim, and J. Y. Song, "Asymptomatic infection and atypical manifestations of covid-19: comparison of viral shedding duration," *The Journal of Infection*, 2020.
- [9] S. Lee, T. Kim, E. Lee, C. Lee, H. Kim, H. Rhee, S. Y. Park, H.-J. Son, S. Yu, J. W. Park, et al., "Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with sars-cov-2 infection in a community treatment center in the republic of korea," *JAMA internal medicine*, 2020.
- [10] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al., "Temporal dynamics in viral shedding and transmissibility of covid-19," *Nature medicine*, vol. 26, no. 5, pp. 672–675, 2020.
- [11] M. Casey, J. Griffin, C. G. McAloon, A. W. Byrne, J. M. Madden, D. McEvoy, A. B. Collins, K. Hunt, A. Barber, F. Butler, et al., "Estimating pre-symptomatic transmission of covid-19: a secondary analysis using published data," *medRxiv*, 2020.
- [12] C. McAloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, et al., "Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research," *BMJ open*, vol. 10, no. 8, p. e039652, 2020.
- [13] CDC, "The coronavirus disease 2019 (covid-19)-associated hospitalization surveillance network (covid-net)."
- [14] Wikipedia contributors, "Einstein notation," 2020. [Online; accessed 14-October-2020].
- [15] A. Hauser, M. J. Counotte, C. C. Margossian, G. Konstantinoudis, N. Low, C. L. Althaus, and J. Riou, "Estimation of sars-cov-2 mortality during the early stages of an epidemic: A modeling study in hubei, china, and six regions in europe," *PLoS medicine*, vol. 17, pp. e1003189–e1003189, Jul 2020. 32722715[pmid].
- [16] T. W. Russell, N. Golding, J. Hellewell, S. Abbott, L. Wright, C. A. Pearson, K. van Zandvoort, C. I. Jarvis, H. Gibbs, Y. Liu, et al., "Reconstructing the early global dynamics of under-ascertained covid-19 cases and infections," *medRxiv*, 2020.
- [17] M. Roberts and J. Heesterbeek, "Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection," *Journal of mathematical biology*, vol. 55, no. 5-6, p. 803, 2007.

- [18] A. Florczyk, C. Corbane, D. Ehrlich, S. Freire, T. Kemper, L. Maffenini, M. Melchiorri, M. Pesaresi, P. Politis, M. Schiavina, F. Sabo, and L. Zanchetta, "Ghsl data package 2019," 07 2019.
- [19] J.-A. S. Moser, A. Galindo-Fraga, A. A. Ortiz-Hernández, W. Gu, S. Hunsberger, J.-F. Galán-Herrera, M. L. Guerrero, G. M. Ruiz-Palacios, J. H. Beigel, L. R. I. . S. Group, *et al.*, "Underweight, overweight, and obesity as independent risk factors for hospitalization in adults and children from influenza and other respiratory viruses," *Influenza and other respiratory viruses*, vol. 13, no. 1, pp. 3–9, 2019.
- [20] R. Pérez-Padilla, A. Schilman, and H. Riojas-Rodriguez, "Respiratory health effects of indoor air pollution," *The International Journal of Tuberculosis and Lung Disease*, vol. 14, no. 9, pp. 1079–1086, 2010.
- [21] K. Prem, A. R. Cook, and M. Jit, "Projecting social contact matrices in 152 countries using contact surveys and demographic data," *PLoS computational biology*, vol. 13, no. 9, p. e1005697, 2017.
- [22] C. R. Wells, P. Sah, S. M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L. A. Meyers, B. H. Singer, and A. P. Galvani, "Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak," *Proceedings of the National Academy of Sciences*, vol. 117, no. 13, pp. 7504–7509, 2020.
- [23] B. J. Cowling, S. T. Ali, T. W. Y. Ng, T. K. Tsang, J. C. M. Li, M. W. Fong, Q. Liao, M. Y. Kwan, S. L. Lee, S. S. Chiu, J. T. Wu, P. Wu, and G. M. Leung, "Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in hong kong: an observational study," *The Lancet Public Health*, vol. 5, pp. e279–e288, May 2020.
- [24] C. I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, P. Klepac, G. J. Rubin, and W. J. Edmunds, "Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk," *BMC medicine*, vol. 18, pp. 1–10, 2020.
- [25] L. C. Joseph, P.-S. Javier, A. S. Andrew, R. Andrea, and F. Jacques, "Assessing the impact of non-pharmaceutical interventions on sars-cov-2 transmission in switzerland," *Swiss Medical Weekly*, vol. 150, no. ARTICLE, p. w20295, 2020.