# Center for Policing Equity COVID-19 Modeling Project

Model authors (alphabetical):
Phillip Atiba Goff, PhD
Amelia M. Haviland, PhD
Tracey Lloyd, PhD
Mikaela Meyer
Rachel Warren

Documentation prepared October 22, 2020

## Acknowledgements

Lauren Hagani provided skilled research assistance. All errors are our own.

## Table of Contents

**Overview** .......... 1

**Model Description** .......... 3

**Model Inputs and Source Data** .......... 7

**Modeling Uncertainty** .......... 10

## Overview

**We sought to simulate COVID-19 spread in a large, synthetic U.S. city**, focusing on contributions of three domains in driving spread of COVID-19 and related racial disparities:

a) low-wage, essential work,
b) police-public contact, and
c) jail and prison churn (specifically, the movement of people from local jails and state prisons to their home communities).

Our focus is on community spread. We do not focus on spread *within* incarceration facilities, which has been examined in detail by Recidiviz and others.

**We used a Susceptible-Infected-Recovered (SIR) model to illustrate the high level dynamics of the disease spread** (rather than projecting precise case numbers in an actual city). The code and input files to replicate our main analyses are available in this **GitHub repository**.

**Our model is stylized** in several important ways**:**

1. We focus on spread among White and Black residents only, with the **racial/ethnic distribution of these residents** based on their relative share of the U.S. national distribution estimated using

U.S. Census/American Community Survey data: 60.4% White alone, non-Hispanic, and 13.4% Black alone (includes Black-alone-Hispanic).

2. We rely on **2 distinct infection rates**, derived from a March 2020 study of spread in King County, WA, USA (Thakkar et al. 2020): a social distancing rate and a non social distancing rate.
3. We assume a constant **transmission rate** of .015.
4. We assume equal **initial infection rates** in each population subgroup we examine.
    a. We begin with 10 positive cases in the smallest group (Black essential labor).
    b. Other groups' initial case counts are set to match the rate in the smallest group.
5. The **key stylized concepts** captured in our models are:
    a. **Essential labor** is roughly defined as: relatively low-wage, essential workers.
        i. The **labor market** in our synthetic city is set at national averages for both:
            1. distribution of workers across CPS occupational categories, and
            2. racial distribution of workers within each occupational category
        ii. We set the **race-specific shares of essential laborers** within our synthetic city using Bureau of Labor Statistics (BLS) Current Population Survey (CPS) data. We classify each CPS occupational category as (a) essential and low-wage, (b) non-essential and/or higher-wage, or (c) in between. Then we sum counts of Black and White workers (in thousands) across all occupational categories.
            1. Our procedure is conservative with respect to assigning essential worker designations because we exclude workers in occupational categories where the total count of workers is too low for BLS to provide reliable estimates by race/ethnicity.
            2. We also aimed to exclude **higher-wage workers** from our essential labor counts based on the assumption that they will have greater access to PPE and private transportation (i.e., be at lower risk for contracting COVID through their employment) than will lower-wage workers.
        iii. In addition, classification of *essential* workers varies across U.S. states and localities; we consulted lists from Kentucky and New York State.
    b. **Police officer presence** is set at the average rate for U.S. cities with populations above 500K (i.e., at 24.3 per 10K residents) and with the assumption that 70% of all sworn officers are patrol officers.
        i. Daily **police-public contact** is notoriously difficult to estimate; we generate our estimates using public calls for service data and key stakeholder interviews.
    c. **Jail and prison populations** are combined and set to the U.S. national averages for:
        i. per capita incarceration rates (average daily populations),
        ii. per capita daily release/reentry rates, and
        iii. distribution of (i) and (ii) across racial groups.
6. We **do not separately model** spread due to health care workers, correctional officers, parole/probation officers, in-jail and in-prison visitation, nursing homes, food processing plants, schools, daycares, churches, super-spreader events, etc. The contributions of these domains are captured within the "non-essential labor" population subgroups.

Our **source data** are publicly available. We used:

- Census and American Community Survey data on U.S. population and racial distribution,
- Bureau of Labor Statistics summaries of counts of U.S. workers and racial distribution of workers across detailed occupational categories from the Current Population Survey,
- Public calls for service data from LAPD and Baltimore PD, and
- Bureau of Justice Statistics data on jail and prison populations and releases.

# Model Description

### How the model works

- Overall, the model keeps track of the number of susceptible, infected, and 'recovered' for each population subgroup in our model on each day. Note that we do not model deaths, the SIR model shifts everyone who is infected to a non-infectious state called 'recovered' which includes those who fully recover, those who still experience COVID-related morbidity, and those who die.
- To generate **group population sizes**:
    - We used national data, sometimes skewed toward large cities, with the goal of constructing a synthetic metropolitan region with a large city ("synthetic city").
    - We used national averages to estimate how many of the 5M residents would fall into each group of interest. The eight groups are listed in the table below.

**Table 1. Synthetic City Population Subgroups**

| | Short Name | Detailed description | Group size (see "Model Inputs and Source Data" section for details) |
|---|---|---|---|
| 1 | White non-essential-labor | White *higher-wage and/or non-essential* workers and all White non-workers (including children and the elderly) | (5M * .604) *(1-.11) |
| 2 | White essential labor | White *low-wage essential workers* | (5M * .604) *.11 |
| 3 | Black non-essential-labor | Black *higher-wage and/or non-essential* workers and all Black non-workers (including children and the elderly) | (5M * .134) *(1-.13) |
| 4 | Black essential labor | Black *low-wage essential workers* | (5M * .134) *.13 |
| 5 | White police | White *patrol officers* [assuming patrol officers are 70% of all sworn officers] | 5M*(24.3 /10K)*.7*.74, where .74 = # White patrol officers/(total # White and Black patrol officers) |
| 6 | Black police | Black *patrol officers* [assuming patrol officers are 70% of all sworn officers] | 5M*(24.3/10K)*.7*0.27, where .27 = # Black patrol officers/(total # White and Black patrol officers) |
| 7 | White people released from jail and prison same-day | White people who had been incarcerated in jail or state prison and are released that day | The number of White people returning to the community from jail/prison each day exactly equals the number going into jail/prison each day. |

| | | | |
|---|---|---|---|
| 8 | Black people released from jail and prison same-day | Black people who had been incarcerated in jail or state prison and are released that day | The number of Black people returning to the community from jail/prison each day exactly equals the number going into jail/prison each day. |

- We set the number of infected individuals in each subpopulation at the beginning of the simulation (i.e., on day 0). The number of susceptible individuals on day 0 is then equal to the population of this subgroup minus the number of initial individual infections in the subgroup. We begin with no recovered individuals.
    - To generate **initial infection rates**:
        - We begin with the same rate in every group (0.000011).
        - Operationally, we start with 10 positive cases in the smallest of the four main population groups (Groups 1 through 4 in Table 1), which is the Black essential labor group,  and scale up the number of cases in all other groups proportionally based on population size.
- We begin with a contact matrix, which is an 8 x 8 matrix quantifying the contact rates within and between the eight subpopulation groups in the Table above.
    - To generate our **contact matrices -- which determine probabilities that people move from Susceptible to Infected**:
        - We consider a **standardized unit of contact** that can be thought of as 15 minutes of standing next to someone on public transportation. We assume a constant transmission rate over these standardized units of contact. We are aware that the transmission rate and the contact numbers can be scaled in many ways relative to each other. We set the disease transmission rate at 0.015.
        - We handled police and prison contact differently, which we discuss below. **For all other groups**, we are guided by the transmission rates under conditions of “social distancing” and “no social distancing.” We use an infection rate of 2.7 under no social distancing and an infection rate of 1.2 under social distancing. Dividing by the product of our assumed 0.015 transmission rate and our assumed 21 days until recovery, we assume approximately 13.9 daily contacts under no social distancing and 5.7 under social distancing.
            - Before social distancing occurred, most groups operate under the same contact levels. Essential workers are set to have higher average at work contact levels than non-essential due to different job types and police are set to have higher contact rates at work than other essential workers on average.
            - After social distancing occurs, only essential workers operate under pre-social distancing contact levels (this includes police). All other groups have greatly reduced contacts.
            - The stylized simulation uses strong racial residential segregation, where we set contact rates between Black and White residents who are not essential workers to 0.
            - We then assume no racial segregation among workers while they are at work, distributing contacts proportionally across the Black and White worker groups.
        - For **jail and prison contact**, we do not model spread of the disease within jail or prison.

            - Instead, we focus only on two parameters: the number of those returning to the community from jail or prison, and the rate of infection among those people.
            - The number of in-prison infections is assumed to follow an exponential growth model written below for 35 days. On day 35, we assume the prison infection rate is 0.35. We calculate k in the below equation in the *prison_rate_build* function.
                - $Infections(day) = Initial\ Infections * e^{k*day}$
            - The number of people who had been in jail or prison returning to the community each day is constant by race. The infection rate among these people changes over time, but does not vary by race.
            - Contact rates between these returning people and the rest of the city are determined by population size and race. We assume they are not essential workers.
            - We assume that the same number of people go to jail or prison as return from jail or prison each day.
            - We are not separately modeling in-prison visits or in-jail visits, or correctional officers, although their contributions to community spread are accounted for in our non-essential labor groups.
        - For **police**, we focus on patrol officers, who on average have many contacts with other groups during the day.
            - From calls for service data, and interviews with police executives, we estimate 20 police-public contacts per day on duty, so 10 contacts per day including both on duty and off duty days (assuming officers are on duty half of calendar days).
            - We model spread of infection among police, both from police-public contacts and from police-police contacts.
            - When police are off duty, they follow the same contact patterns as any other workers of the same race.
            - Based on using calls for service data to estimate police-public contacts, we assign higher total contact levels to police (as described above) than to other types of workers.
- We construct pre- and post-stay-in-place order "contact intensity" matrices (see the Contact_Matrix_Pre_SIP… and Contact_Matrix_Post_SIP… .csv files in the input_data directory), where each element in the contact matrix for either pre- or post-stay-in-place order is multiplied by the transmission rate. Thus, the items in the contact intensity matrices represent the force of interaction between subgroup *i* and *j*, where *i* represents the row index and *j* represents the column index of the matrix. It can be interpreted as the number of contacts each person in row *i* has with people in column *j* per day. The constant transmission rate is the likelihood of the disease being passed from an infected person to a susceptible person in a given contact.
- We simulate what is happening in our synthetic city for 120 days. On each day, our model does the following (see the *build_model* in build_models.py function):
    - Up until day 28 on which we set the stay-in-place order to begin, we use a contact matrix for values based on no social distancing. Once the stay-in-place order day is reached, our contact intensity matrix is changed to reflect our new contact matrix based on a stay-in-place order.
    - We calculate a vector, *l*, that is length 8 by multiplying the contact intensity matrix by the proportion of the subpopulation group that is infected. Thus, each item in *l* represents the infection rate for each subpopulation group, known as the force of infection in the literature.

    - We calculate the number of new infections from contacts for each subpopulation group by multiplying the force of infection by the number of susceptible individuals.
    - The numbers of susceptible, infected, and recovered individuals on day *t* are calculated as follows:
        - The number of susceptible individuals in a subpopulation group is equal to the number of susceptible individuals on day *t* -1 (the initial values should be thought of as occurring on *t* = 0) minus the number of new infections from contacts.
        - The number of recovered individuals in a subpopulation group is equal to the number of recovered individuals on day *t* - 1 plus the number of newly infected individuals from 14 days ago. We assume that all individuals recover 14 days after being initially infected in the sense that they are no longer infectious.
            - For days 1-14, we multiply the recovery rate by the number of infected individuals on day 0.
        - The number of infected individuals in a subpopulation group is equal to the number of infected individuals on day *t* - 1 plus the number of new infections from contacts minus the number of recovered individuals.
    - We have three matrices, one each for susceptible, recoved, and infected individuals. In each matrix, the rows represent each day (with 120 rows total), and the columns represent each subpopulation group. For example, the 5th row of the infected group matrix contains the estimated number of people infected at day five in each of the subpopulation groups.
- With the number of people in each subgroup in each disease category for 120 days, there are a number of summary statistics we calculate that we refer to here as outcomes. **We measure all of the outcomes for this article on day 40 of the simulation.** We repeat the above SIR simulation steps varying the conditions of the model by using different police contact with civilian rates, different police group sizes, and different max prison infection rates to arrive at **uncertainty estimates** for our outcomes. See the section below on Modeling Uncertainty for full details.
- We also repeated the above SIR simulation steps with different contact matrices to arrive at estimates of what percentage of cases would be reduced by changing factors that could be impacted by policy choices. See function *create_matrix* to see where these contact matrices are created. The variants of contact matrices we used include:
    - **No contact between police officers and the public, and no infected individuals are released from jail or prison.** The functions *no_police_contact* and *no_prison_churn* are used to create this variant of contact matrices. In our output, this model run is denoted as "no_police_prison." "No contact" here can be interpreted as either no contact or as contact with sufficient PPE and distancing to eliminate risk of disease transmission.
    - **No contact between police officers and the public, no infected individuals are released from jail or prison, and no contact between essential workers and non-essential workers.** The function *no_forced_labor* is used to create a modified group size matrix where essential workers are absorbed into their respective non-essential worker racial subgroups, and the function *drop_forced_labor* is used along with *no_police_contact* and *no_prison_churn* to create the contact matrices. In our output, this model run is denoted as "no_police_prison_fl."
- **To calculate the percentage of spread attributable to policing and jail/prison churn**, we compare the outcomes using the contact matrices produced by the *no_police_contact* and *no_prison_churn* functions to the outcomes from the original model. To calculate the additional percentage of spread attributable to essential labor, we compare outcomes using the contact matrices produced by the *no_police_contact* and *no_prison_churn* functions to oucomes using the *no_police_contact, no_prison_churn,* and *drop_forced_labor* functions.

## Model Inputs and Source Data

- **Overall Population and Population Subgroup Overview**:
  - We used national data, sometimes skewed toward large cities, with the goal of constructing a synthetic metro region with a large city ("synthetic city").
  - We used a total population of 5M residents, rounding up from median U.S. state size of 4.467M residents.
  - We included all ages of residents in our synthetic city (not just ages 16+).
  - Because (a) non-Hispanic White and (b) Black residents together make up <100% of the US population, and we used the national racial/ethnic population distribution, the total population across our 8 population subgroups sums to less than 5M.
  - We used national averages to estimate how many of the 5M residents would fall into each group of interest (see below for details on essential labor, police, and people returning home from jail/prison). We used 8 mutually exclusive population subgroups in our model (see Table 1 above for additional details on how group sizes were calculated).

- **Infection and Transmission Rates**
  - We used infection rates from an Institute for Disease Modeling study in King County, WA in February and March 2020 (Thakkar et al. 2020):.
    - We used 2 effective reproduction number (Re) values, which capture the average number of cases caused by each infected person:
      - Without social distancing: Re = 2.7
      - With social distancing: Re = 1.4
  - Our Re estimate of 1.4 under social distancing was corroborated by an Imperial College London study, which found an average Re with social distancing of 1.43 across 11 European nations.
  - We assume a constant **transmission rate** of .015.

- **Essential Labor Group Sizes**
  - We set the **race-specific shares of essential laborers** within our synthetic city using Bureau of Labor Statistics (BLS) Current Population Survey (CPS) data from 2019. We classified each CPS occupational category as (a) essential and low-wage, (b) non-essential and/or higher-wage, or (c) in between. Then we summed counts of Black essential workers and White essential workers (separately) across all occupational categories and divided these counts by race-specific US population counts to generate rates of low-wage essential work among the Black and White populations.
  - In classifying each occupational category as *essential* or nonessential, we consulted executive orders issued in Kentucky and New York State.
  - Our procedure is conservative with respect to assigning essential worker designations because we exclude workers in occupational categories where the total count of workers is too low for BLS to provide reliable estimates by race/ethnicity.
  - In addition, we aimed to exclude **higher-wage workers** from our essential labor counts based on the assumption that they will have greater access to PPE and private transportation (i.e., be at lower risk for contracting COVID through their employment) than will lower-wage workers. For example, our counts exclude professional/managerial health care occupations (physicians, health care administrators) but include health care support occupations (nursing aides and medical equipment preparers).

- **Police Presence – Patrol Officer Group Sizes**

  Police patrol presence estimates come from the number of officers per 10K residents in large U.S. cities as reported through the FBI UCR reporting program. We estimate that the share of patrol officers is between ⅔ and ¾ of all sworn officers. Our preferred estimate of the share of patrol officers is 0.7, resulting in an estimate of patrol group size of 24.3*0.7 officers per 10,000 residents. Due to the variability in this value across large US cities (informed by the Governing anlaysis of UCR 2016 employee reporting data), we also used four alternate estimates of this value (see table 2 below), resulting in five different inputs for patrol police group sizes: **13.8069, 14.9114, 16.016, 17.1205, 18.225 per 10K residents.**

Table 2. Estimates of Patrol Officers as a Share of Population

| | Number of patrol officers per 10K residents | Source |
|---|---|---|
| Preferred Estimate: | 24.3 *.7 | **Mean** value for 32 U.S. cities with populations >500K, per "Governing" analysis of UCR 2016 employee reporting data;<br>Multiplied by our best estimate of the share of all sworn officers who are patrol officers. |

- **Daily Police-Public Contacts**

  There is no definitive data source on the average number of contacts patrol officers have with the public during a typical shift, and this is likely to vary across metropolitan areas and over time. We used two approaches to estimate this number. First, we analyzed publicly available calls for service (CFS) data sets from 2 local law enforcement agencies: the Baltimore PD and LAPD. With each data set, we divided the annual number of CFS by the number of sworn patrol officers and then by 365.25 to obtain a daily rate of contacts. We also asked experienced law enforcement professionals to estimate the number of daily contacts, which returned a range of 20-30 per shift; assuming that half of calendar days are on-duty yields an estimate of 10-15 police-public contacts per calendar day. The higher (and more believable) estimate from the CFS data was 9.4 contacts per calendar day, roughly equal to the lower estimate from the stakeholder data, which yielded our preferred estimate of 10 contacts per day. (We note that we assume only 1 resident is present per call, resulting in a conservative estimate of the number of police-public contacts.)

  To divide these daily contacts across racial groups, we multiplied the number of daily contacts by the racial distribution of all U.S. arrests per BJS arrest data from 2014 (the most recent year available as of May 2, 2020). Because the UCR data do not capture Hispanic ethnicity, they will overstate arrests for the White group (we use "White alone, non-Hispanic" population data in our synthetic city), making Black-White disparities calculated using these data conservative.

  To approximate the distribution of this police contact rate across metropolitan areas (i.e., to allow for uncertainty in this model input), we use five different inputs for daily police contact with the public: **5, 7.5, 10, 12.5 and 15.**

Table 3. Estimates of Patrol Officers' Daily Contacts with Residents, by Resident Race

| **Police Contact** | **Total** Number of daily contacts with residents | Number of daily contacts with **White** residents | Number of daily contacts with **Black** residents | Sources |
|---|---|---|---|---|
| Preferred Estimate: | 10 | Total * 0.693478703716014 | Total * 0.277981908502461 | Overlap between Baltimore Calls for Service and police executive interview estimates<br><br>Racial distribution from BJS arrest data |

- **Infection Rate among People Released from Jail/Prison**

  In our model, we assume the infection rate within jails and prisons follows an exponential growth model initially, and then holds at a peak number of infections from day 35 forward in order to estimate the other parameters of this exponential growth model. Our preferred estimate is a peak infection rate of .35. According to a Marshall Project article, mass testing in select prisons revealed infection rates exceeding 65% in several facilities. A few other anecdotal sources give rates between 10-20% (such as 11.65% in Rikers in June and 14% in NJ state prison). Testing in jails and prisons is quite limited.

  Due to the variability across metropolitan areas in the number of infections among corrections populations, we use three inputs for max prison infection rates—**0.25, 0.35, and 0.45**—to quantify the variability associated with this assumption.

- **Numbers of People Released from Jails and Prisons Daily**

- **Jails.** Per the 2018 BJS Annual Survey of Jails (tables 1 and 4):
  - There were 10,700,000 jail admissions in the US in 2018, and the average daily population at midyear has been fairly stable since 2011, centering around roughly 737K. Because the average daily population has been near-constant, we assume a similar number of releases as admissions each year.
  - We calculate an average of 10,700,000-737,000 = 9,963,000 / 365 = 27,296 releases per day. We assume the same racial distribution among people being released each day as among people currently incarcerated.
  - The racial/ethnic distribution of the jail population in 2018 was 49.9% White and 32.8% Black. So, we estimate 27,296*.499 = 13,621 White people released and 27,296*.328 = 8,953 Black people released from US jails per day.
- **Prisons**. Per the BJS National Prisoner Statistics program *Prisoners in 2017* report (tables 7 and 3):
  - There were 622,377 releases from Federal and State prisons in 2017. This yields a daily release count of 622,377/365.25 = 1704.

    - The racial/ethnic distribution of the prison population in 2017 was 30.31% White and 33.05% Black. We assume the same racial distribution among people being released each day as among people currently incarcerated. So, we estimate 1704 * .3031 = 516.47 White people released and 1704 * .3305 = 563.16 Black people released from US federal and state prisons per day.
- **Combined U.S. jail and prison releases (daily)**
    - 13621+ 516 = **14137** White people released from U.S. jails and prisons per day
    - 8953 + 563 = **9516** Black people released from U.S. jails and prisons per day
- We divide these daily release counts by race-specific U.S. population counts to generate race-specific release rates that we then apply to our synthetic population.

## Modeling Uncertainty

To allow for uncertainty in our calculations reflecting variation across cities in the US, we use different contact matrices in model runs and compare outputs. These estimates are intended to demonstrate the range of values that occur in this stylized model under the range of conditions that occur in large US metropolitan areas. We considered uncertainty (or ranges) in our key measures related to 3 key model inputs:

1. the number of people the police interact with on a daily basis,
2. the number of patrol police officers in our synthetic city, and
3. the maximum prison infection rate.

Using the ranges of values for these inputs described in the Model Inputs and Source Data section, we generated different contact matrices and/or group size matrices to be used in our model. We ran the SIR model with these variants of our original matrices and compared their outputs to the original model outputs.