COVID-19 Projection Model Proposed

17 June 2020 Working Draft

Richard D. Hamlin
rdhamlin@pandemonium.dev

Model Overview

Demographically-structured (age, sex, etc.) SEIRD(S) Compartmental stochastic model that is AI-enhanced and spatially distributed and refinable with spatial dynamics and micro mechanistic submodeling.

The basic framework of the model will allow for an arbitrary degree of demographic refinement where data is available. For example, if aggregate nursing home data are known, nursing homes can be separated out from the general population for modeling as a subgroup. If (at least some of) the individual facility data is known, then those particular facilities can each be modeled separately.

The framework will also allow for an arbitrary degree of spatial refinement where data is available. For example, if there is data for a city or town, it can be separated out from its encompassing county.

In both cases, the interaction between the more specific entity or subgroup and its encompassing entity or group will be modeled. And the interactions with other subgroups or entities that differ will be supported too.

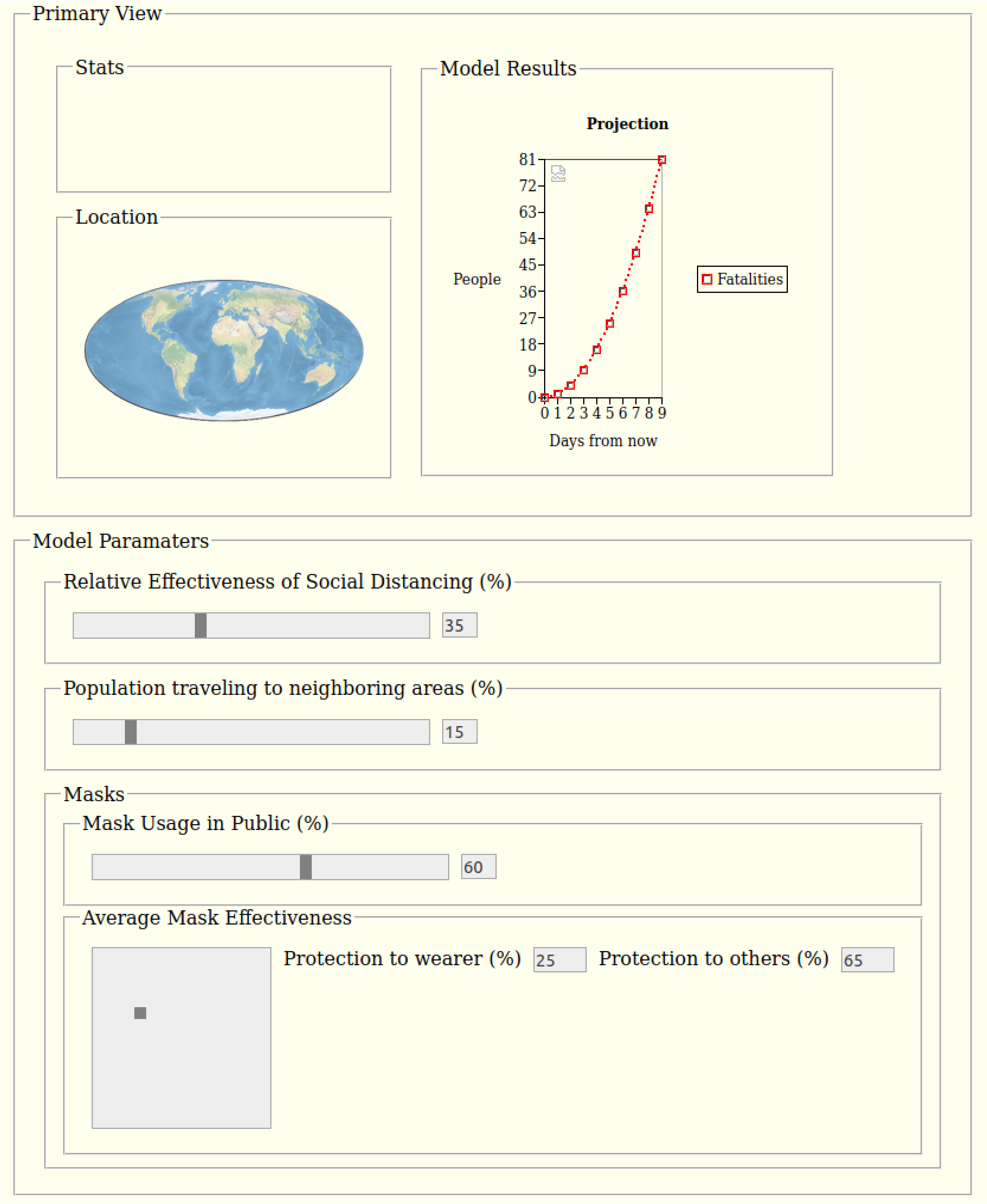
The framework will automatically deduce the compensation needed due to refinements at the next higher level to remain self-consistent, and detect cases in which that is not possible, indicating that the data is inconsistent (errant). Data considered more reliable will then be given precedence to work around the inconsistency.

This makes it easier to utilize different data sources with different degrees of specificity at the same time as well as utilize different specialized sub-models.

Goals

1. Quality model projections down to the local level (where data available).
2. Help individuals understand their own & societal risks from various actions.
3. Make it easy for nonprogrammers to tweak various parameters and see the potential impact of various policies and behaviours and alternative models.
4. Extensibility to allow users to import additional/refined/corrected data as well as create submodels.

Web User Interface Conceptual Prototype



Primary View

Stats

Location

Model Results

Model Paramaters

Relative Effectiveness of Social Distancing (%) 35

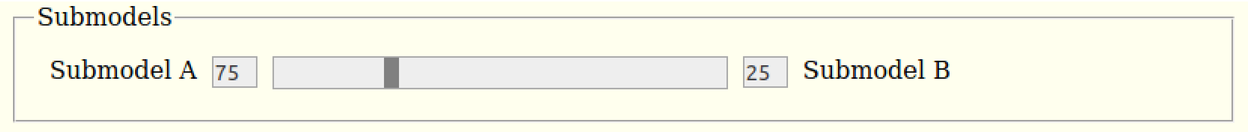
Population traveling to neighboring areas (%) 15

Masks

Mask Usage in Public (%) 60

Average Mask Effectiveness

Protection to wearer (%) 25 Protection to others (%) 65



Submodels

Submodel A 75 25 Submodel B

The user will be able to select a location (in much the same way as in many data dashboards) and alongside a map of that area, see the model projection results for that location.

The user will be able to easily vary certain model parameters (representing assumptions in the model) with sliders to be able to immediately see the effect. These parameters will be able to be adjusted locally as well as globally or any regional level in between.

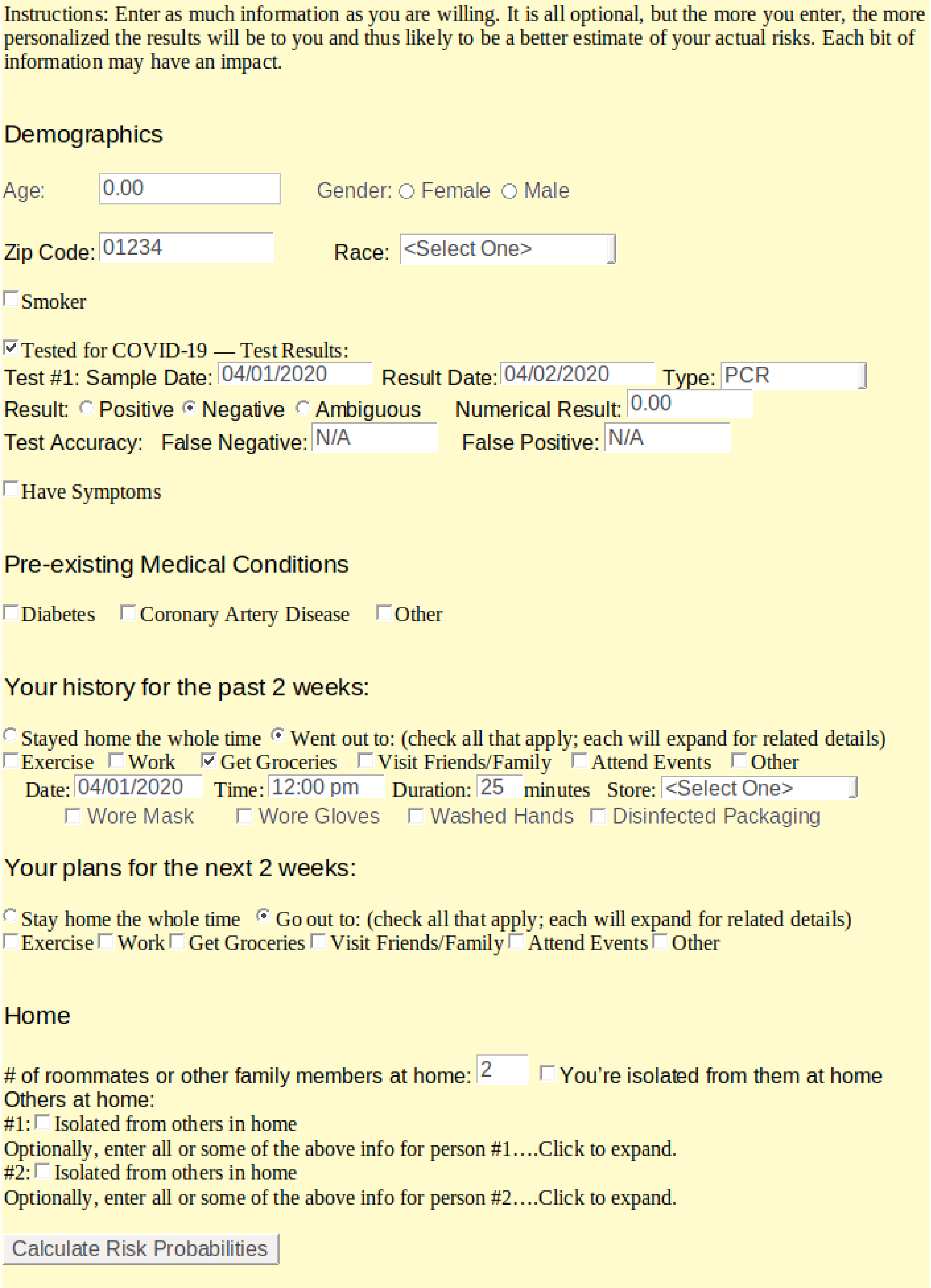
The parameters shown in the screenshot are for illustrative purposes and not necessarily the exact parameters used in the model. There will likely be many model parameters that users will be able to adjust, in which case some of the primary ones will be shown and the others will be hidden until expanded by the user.

Individualized Risk Assessment

Individual users will be able to enter their own demographic information and medical conditions to be able to get a personalized assessment of their own risk, showing their chances of dying from COVID-19 themselves, being intubated, having a stroke, being hospitalized, and being infected, etc., as well as the societal risk that they might spread it. The user will also be able to enter their personal routine and behaviors to customize this further as well as info for family members with whom they live or visit. With that, they can get their own risk, the risk to their own family, and risk to society at large. Then the individual user can use the web tool to evaluate the marginal risks associated with a particular activity to inform their decision of whether to participate in it or not. It will also help them optimize their personal routines to minimize their own risks and their contribution to societal risks. Some of the conditional probabilities for certain unknown variables (e.g., assuming currently infected or not infected) will also be accessible to provide greater insight.

Data privacy

Since sensitve information will be entered, it will need to be adequately protected. One possibility is to have the individualized risk assessment performed entirely on the client side. Another is to have no persistent server storage of the personal data. But store it in cookies on the client side, so that it does not need to be re-entered. These cookies can be encrypted in case the user’s machine is later compromised.



Instructions: Enter as much information as you are willing. It is all optional, but the more you enter, the more personalized the results will be to you and thus likely to be a better estimate of your actual risks. Each bit of information may have an impact.

Demographics

Age: 0.00 Gender: ○ Female ○ Male

Zip Code: 01234 Race: <Select One>

☐ Smoker

☑ Tested for COVID-19 — Test Results:
Test #1: Sample Date: 04/01/2020 Result Date: 04/02/2020 Type: PCR
Result: ○ Positive ● Negative ○ Ambiguous Numerical Result: 0.00
Test Accuracy: False Negative: N/A False Positive: N/A

☐ Have Symptoms

Pre-existing Medical Conditions

☐ Diabetes ☐ Coronary Artery Disease ☐ Other

Your history for the past 2 weeks:

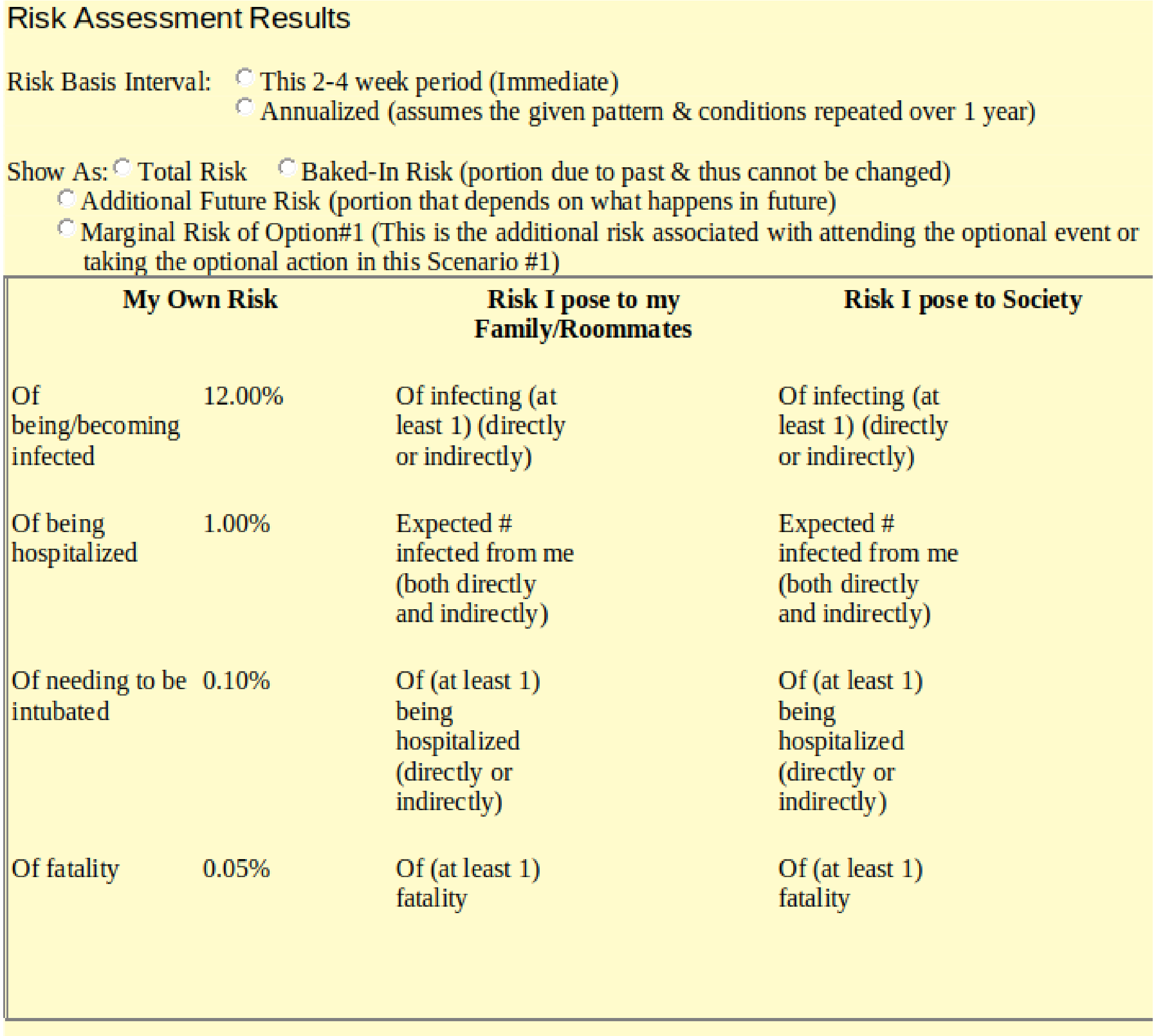
○ Stayed home the whole time ● Went out to: (check all that apply; each will expand for related details)
☐ Exercise ☐ Work ☑ Get Groceries ☐ Visit Friends/Family ☐ Attend Events ☐ Other
Date: 04/01/2020 Time: 12:00 pm Duration: 25 minutes Store: <Select One>
☐ Wore Mask ☐ Wore Gloves ☐ Washed Hands ☐ Disinfected Packaging

Your plans for the next 2 weeks:

○ Stay home the whole time ● Go out to: (check all that apply; each will expand for related details)
☐ Exercise ☐ Work ☐ Get Groceries ☐ Visit Friends/Family ☐ Attend Events ☐ Other

Home

of roommates or other family members at home: 2 ☐ You're isolated from them at home
Others at home:
#1: ☐ Isolated from others in home
Optionally, enter all or some of the above info for person #1….Click to expand.
#2: ☐ Isolated from others in home
Optionally, enter all or some of the above info for person #2….Click to expand.



Risk Assessment Results

Risk Basis Interval: ○ This 2-4 week period (Immediate)
○ Annualized (assumes the given pattern & conditions repeated over 1 year)

Show As: ○ Total Risk ○ Baked-In Risk (portion due to past & thus cannot be changed)
○ Additional Future Risk (portion that depends on what happens in future)
○ Marginal Risk of Option#1 (This is the additional risk associated with attending the optional event or taking the optional action in this Scenario #1)

My Own Risk		**Risk I pose to my Family/Roommates**	**Risk I pose to Society**
Of being/becoming infected	12.00%	Of infecting (at least 1) (directly or indirectly)	Of infecting (at least 1) (directly or indirectly)
Of being hospitalized	1.00%	Expected # infected from me (both directly and indirectly)	Expected # infected from me (both directly and indirectly)
Of needing to be intubated	0.10%	Of (at least 1) being hospitalized (directly or indirectly)	Of (at least 1) being hospitalized (directly or indirectly)
Of fatality	0.05%	Of (at least 1) fatality	Of (at least 1) fatality

Event/Scenario Assessment

What would be the impact of holding a specific large event or a generic local event? The prestigious Boston Marathon, for example, attracts runners and spectators from around the world and brings them into very close contact. The starting line is so crowded that it can take minutes to cross and is already split into multiple waves. What would be the impact both locally and globally if the event was held as usual? What if certain precautions were taken?

A large event could be specifically modeled, but most small events would not likely get that attention. So, small events can be modeled generically but allow the user to tweak the parameters. Most small events would tend to be local, but some might have people traveling a great distance to attend. So the distribution of how far people travel would be adjustable as well as the expected attendance count.

Likewise, what are the COVID-19 risks in the workplaces for business-as-usual? Then what are these if businesses and/or individuals take various precautions? How does this differ by the type of business? That is to be modeled.

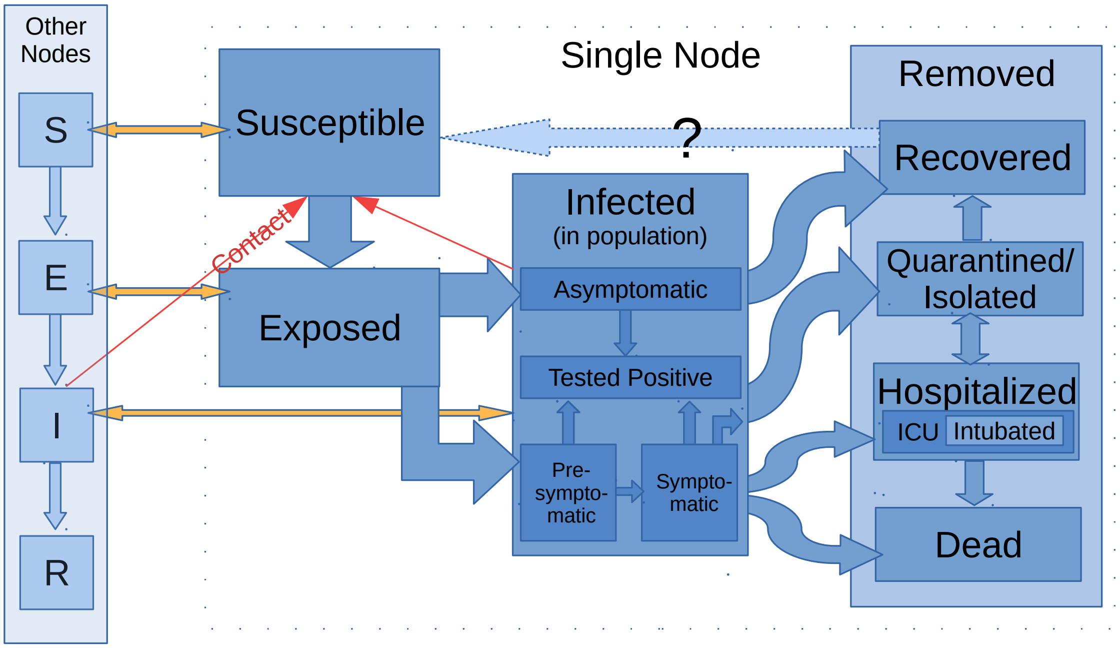
These can be combined to derive the projected impacts of certain regions, states, or countries implementing or considering “re-opening” policies or further restrictions. The interface will allow for the selection of various individual policies to be built-up into an overall policy and view the overall impact. The sensensitivity to individual policies and parameters can be analyzed but must be mindful of the inherent non-linearity. (The impact of 2 policies together is not simply the sum of both alone.)

It will be important to have some human psychology submodel, because the actual impact will result from what people actually do, which is not necessarily what they are told they can or cannot do nor must do! This is tricky and particularly speculative, so the ability to allow alternative such submodels will be valuable. That way these can be treated as another scenario variation and a weighted concensus projection can be used (similiar to what is done with hurricane forecasts).

Custom Data

TO DO: A separate advanced interface will be needed to allow for importing custom data, etc. Gives the potential for crowdsourcing data.

Basic Model Diagram



Each model node represents a geographic area. There can be multiple tiers of sub-nodes. (See Spatial Structure, next.) Within the node, blue arrows indicate how people can move between categories. The ***Exposed*** population represents those who were exposed enough to become infected but are not yet infectious, and thus will all move into the ***Infected*** category within 2-3 days, either as ***Asymptomatic*** (never having symptoms) or ***Pre-symptomatic*** (which may very quickly or slowly become symptomatic). Everyone in the ***Infected*** category is considered Infectious (to some degree) and potentially infecting others. Yellow arrows indicate the migration of people to and from other nodes. (They retain their same categorical status, just in another node, when they move.) Red arrows indicate the contact between ***Infected*** and ***Susceptible*** that leads to exposure, and the rate at which susceptible people move into the ***Exposed*** category for each demographic is a function of the degree of contact, the infectiousness with whom they were in contact, and the susceptibility of the demographic.

*NB: Some paths are not yet shown on the diagram: *Symptomatic* directly to *Recovered*. *Pre-symtomatic* to *Quarantine*. And *Susceptible* bi-directionally to *Precautionary Quarantine*.

of Infected

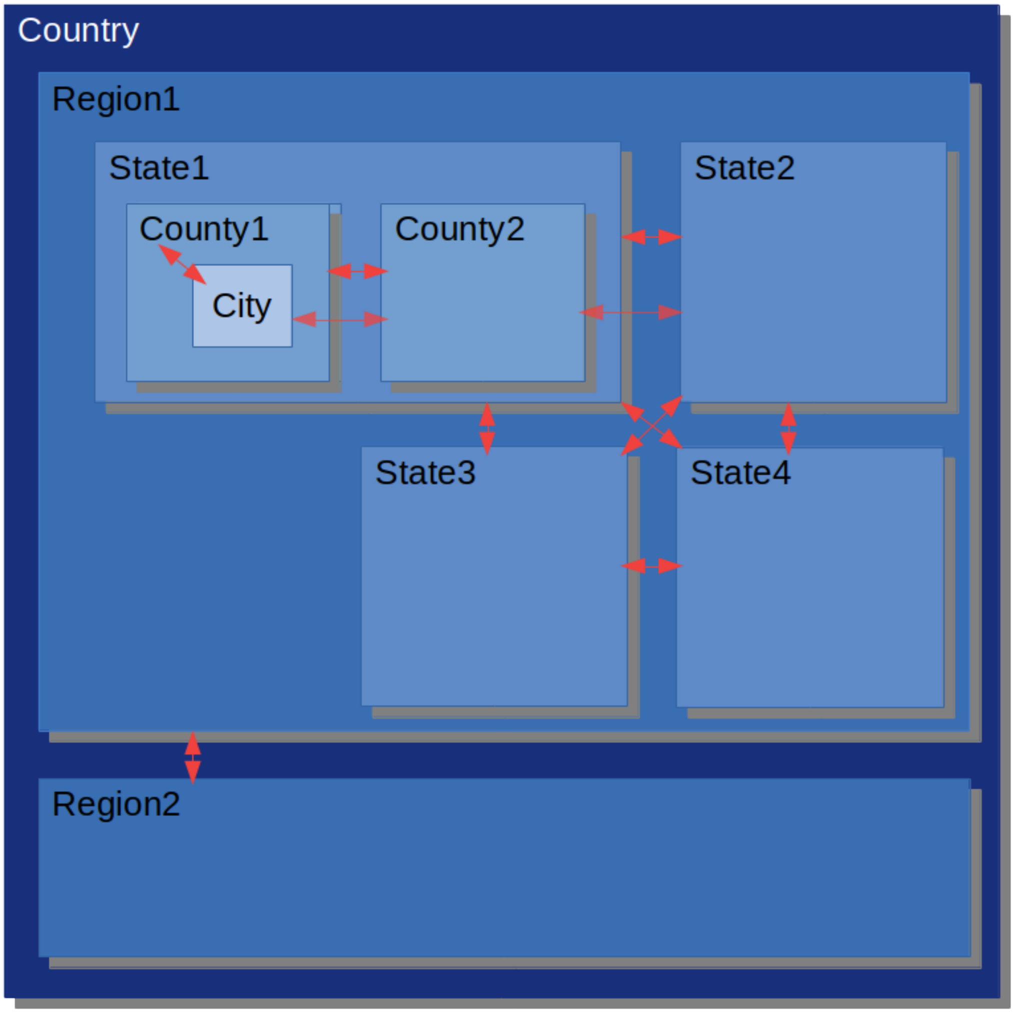
Issue: Reliability of data regarding case counts.

Proposed: Use the number of intubations (where is it the most proximal reliable metric) to estimate historical actual cases, projecting forward from an earlier time to fill in the recent past before projecting into the future.

Discussion: A simplistic projection model assumes that the current system state is known and projects changes of that state into the future based on that. Without extensive, accurate, rapid testing, we do not know how many are actually currently infected with the virus, however. There are different approaches that can be used to deal with this issue. Most simply would be to calculate an effective reproduction rate based on confirmed cases and then project confirmed cases instead of true cases. Or percentage of tests that are positive could be used to estimate a factor to multiply confirmed cases by to estimate actual cases. But that may not be the most reliable. Alternatively, we can project back from the number of intubations (or deaths, but those are further removed temporarily and thus less useful) to estimate the actual number of infected at a sufficiently earlier time (the duration most newly intubated patients were previously infected), and then project forward from then. (This can be iteratively optimized to find the best fit to the data.)
If there are sufficiently large random samples of antibody tests that are reliable, those can also be used, but those also have an associated lag.

Spatial Structure

Simplified Illustration:

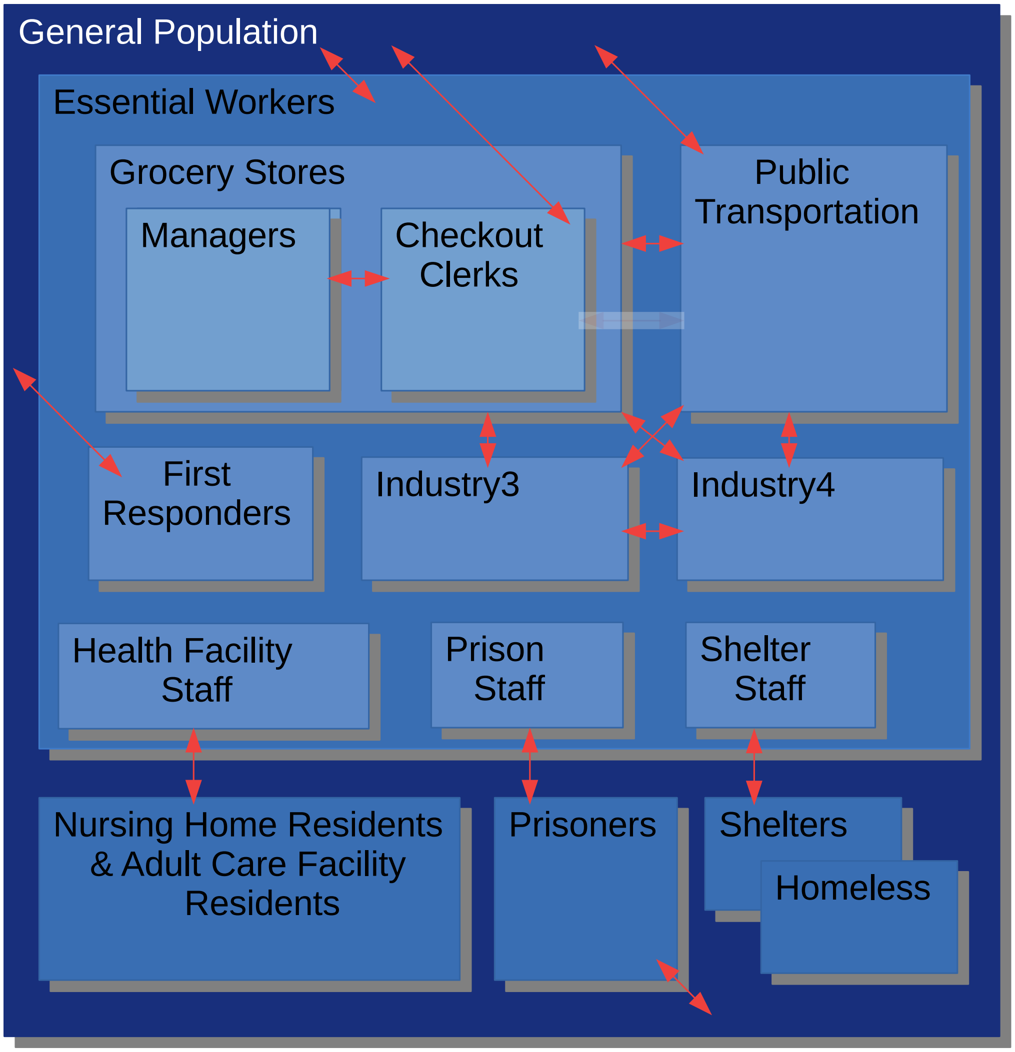


This shows a completely hierarchical spatial structure in which additional levels are possible (e.g. regions within a state), but we may want to also accommodate partially overlapping zones such as PUMA can often be. This would be significantly more complicated, so plan to leave that capability out of the first version for the sake of expediency.

Red arrows indicate the contact/movement between populations of different regions and thus potential for transmission across geographic boundaries. The model framework will support mobility data at different levels and between different levels.

Demographic Structure

Simplified Illustration:



Distributions for each demographic group & subgroup above:

- Age
- Sex
- Race
- Income
- Education

Micro Mechanistic Sub-Modelling

The physical mechanisms of exposure for four means for transmission are considered in estimating the potential amount of exposure:

1. Direct contact with infected individuals.
2. Large (>5μm in diameter) respiratory droplets.
3. Aerosols (droplets <5μm in diameter).
4. Contact with surfaces contaminated by:
    a) prior direct contact of the surface with infected individuals,
    b) prior indirect contact, and/or
    c) settling of respiratory droplets.

The physical spreading must be considered in the model along with the deterioration of viral particles (in term of their infectiousness) over time.

Machine Learning

Machine learning (ML) techniques are planned to be used in multiple ways. ML offers many potential advantages, including the potential for:

- optimizing the model parameter values based on the observed data
- utilizing large, disperate, additional and updated data sets with relative ease
- finding hidden patterns that may be missed in traditional statistical analyses
- and doing so relatively quickly.

Assessing Mask Usage in Public

A parameter that is likely to have a lot of spacial and temporal variability even when there is consistency in public policy (i.e., a state-wide requirement or CDC recommendation). Deep learning (DL) would be very good at recognizing whether people are wearing masks, and potentially even do well at categorizing the masks between types, such as N-95, surgical, cloth, bandana, other homemade, etc. as well as whether it is being worn properly (e.g. covering the nose and well fitted). Then a large set of publically-posted geotagged and timestamped images could be analyzed using the trained DL algorithm to identify how the usage and types vary by location and time. The geotag would determine whether the photo was taken in a public location and the kind of location, which could then be combined with similar locations in the area to estimate the usage rate. Certain locations do not tend to have many people taking photos normally despite the high number of people, however. (Grocery stores are not the most photogenic spots!) Volunteers could be welcomed to take and upload photos at such locations directly to be analyzed for this purpose. There are other ways such random photos could be analyzed to assess the number of people, their density and how much social distancing is occurring. (Trying to use any publically-posted photo would not be good for this, because there is high chance that such a photo is of a group of people who know each other and came together just for the shot, but aren't usually that close.)

Estimating Viral Exposure Probability Distributions

One of the key challenges to doing a detailed risk assessment is not having any direct data of how many active virons (viral particles that are potentially infectious) everyone is exposed to every day. The micro mechanistic sub-model helps, but it is not enough, because we generally don't know who is infected, what their viral load is or how much virus they are shedding, nor their exact movements. But that does not mean that all hope is lost. We can deal with this problem probabilistically.

The motion for people at a particular place can be analyzed from a video recording, if available, ideally from security cameras since these will avoid sampling bias (at least if they are fixed-frame and distributed to provide full coverage). Motion tracking software is well-established AI that can extract the centroidal location of the moving object (in this case humans) as a spatial point as a function of time. Once we have that we can use machine learning to analyze the human movements relative to each other as well as the venue and as the human density increases and decreases. And we can take a set ML-trained virtual automaton bots that will randomly behave characteristically to the real motions on which they were trained and with the same distribution of behaviors, which would be encoding this very complicated joint probability distribution, and run many Monte Carlo (stochastic) simulations with them, randomly selecting which bots were infectious and by what degree, etc. And then along with the mechanistic micromodel we can analyze the level of exposure to a particular bot on each run. This will result in a probability distribution itself due to the uncertainties involved in that process. Then the combined distribution results from all the runs taken together. That is our best estimate of the viral exposure, based on the lack of knowledge of actual movements as would be the case for any future and most past estimates. But if we happen to know one person's actual track from the past, we can run the simulations with that locked-in, to improve the estimate.