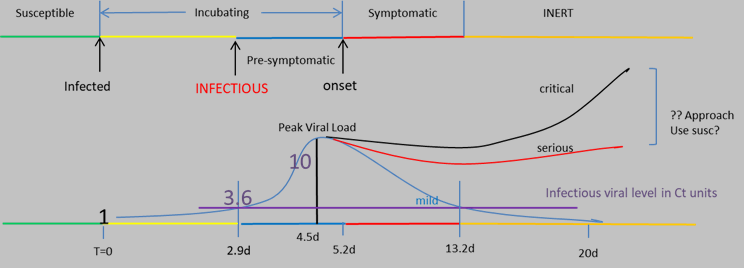
CovidSIMVL Technical Report 004. On Calibrating R0, Hazard Radius and Mingle Factors

September 20, 2020

INTRODUCTION

The agent-based simulation tool CovidSIMVL is based on a predator-prey framework, in which the movement of agents (persons) and their subsequent positions may result in transmissions of a viral load between two agents that are in contact.

The PRIMARY rules for viral growth within an agent that is carrying a viral load, acquired by contact with an infective agent, Is based on temporal dynamics from a paper published by Xi, He et al. *Nature Medicine* **26,** 672-675(2020), expressed as shown below.



Each agent has one of the states: susceptible, incubating, aymptomatic infective, symptomatic infective, and inert (e.g., recovered and assumed no longer infective; deceased; quarantined). The viral loads determine the size of an agent (called the Hazard Radius).

The SECONDARY rules for the CovidSIMVL model concern the interaction between agents who are in their individual viral states. These rules operate on three factors:

1. The population density (number of agents in the fixed arena of 800 x 600 pixels
2. The Hazard Radius of the agents which stochastically change from a base value, according to their viral loads
3. The degree of activity, or Mingling Factor, of each agent, which may change from time and place according to their roles or purported degree of activity. The lower the activity of all agents, the less likely they are to contact, and the higher, the greater the likelihood. These movements can be considered as random walks, and the extent of the arena covered by the random walks of the agents together determine the likelihood of contact between agents.

RELATIONSHIP TO R0

In standard epidemiological studies of contagions, the notion of R0 is the number of successful transmission of an infectious agent to susceptibles in the duration of infectivity of that agent. While R0 is estimated in equation based modelling, CovidSIMVL can count these successful transmissions for each infective agent, and average them at any point.

If R0 characterizes specific local epidemics (BC, Alberta, Vancouver, Maine etc), and it has been found that R0 values < 1 represent self-extinguishing epidemics, while R0 values > 2 represent epidemics that are explosive in growth, the R0 values between 1 and 2 would appear to be epidemics that are continuing, but which can be held in check by various mitigations such as social distancing, infrequent mingling, the wearing of masks, reducing time indoors, etc.

CovidSIMVL Trials

Each set of parameters in CovidSIMVL represents a specific starting state. The stochastic nature of move generation, viral transference, initial spatial arrangement of the population, makes each such trial unique,within confines set by the range of values that a given parameter can assume. The allowable ranges of values for parameters for a given series of trials locates the a stochastically-varying set of results for those trials to be grouped together into a class.

In keeping with basic characteristics of repeated sampling in probability and statistics, we assume that the outcome of trials based on a certain set of parameters converge to a mean value which may differ from the outcome of other parameter settings.

We have run a number of trials with different parameter settings, to obtain two sets of outcomes:

1. The value of R0 at the time of the termination of a trial
2. The values of the Critical Exposure Times in the deciles 1 to 5 (explained below in the discussion of “goals”, in the sense of statistical benchmarks or cut-offs).

We have used as parameters:

1. The population size
2. The Hazard Radius of the agents (uniform to start with at 2, 3, 4 and 5).
3. The Mingle Factor (degree of activity) of the agents: individually they are preset to 3, and then these are modified by a universal Mingling Factor for the space, which modifies the activity of the individual as a product.

In other words, the final Mingle Factor of an agent is the individual MF X Universal MF

Iterations with Trials

Each trial entails a number of iterations, each of which culminates in a description of number of key quantities, namely, the numbers of susceptibles, incubating persons, asymptomatic infectives, symptomatic infectives, and inerts. Changes in those quantities over time reflect the dynamics of spread of the infectious agent.

The duration of the iterations can be set within a series of trials. For example, the duration could be set to an hour, which is useful when modeling dynamics that are associated with changes to agents within a 24 hour period, e.g., students spending a portion of the day in school, and a portion of the day at home.

A duration could be set for a day if the intention is to look at changes in rates over the course days, and there is no need or interest to factor in and portray changes at a more granular level

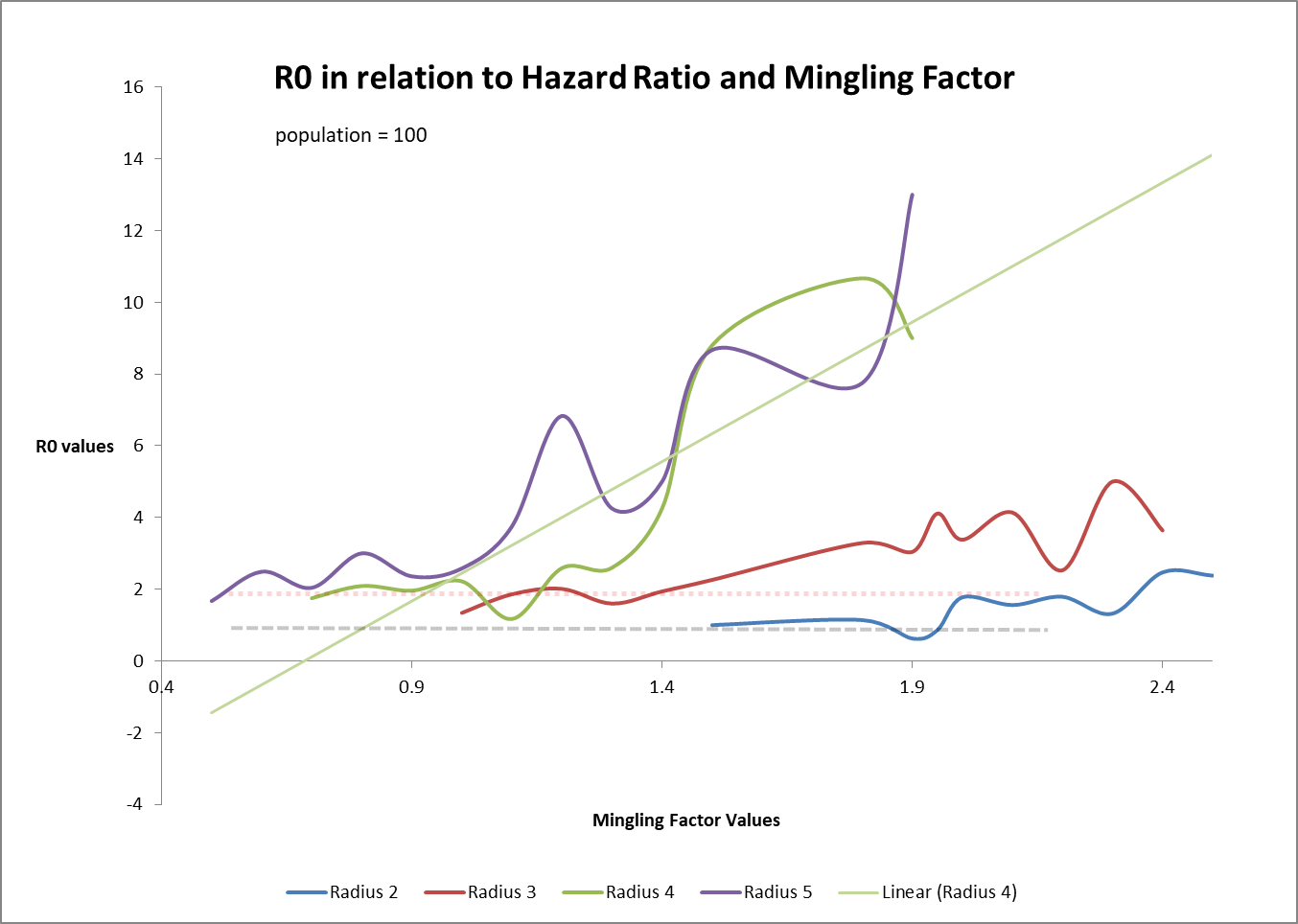
OUTCOME MEASURES OF R0

The following table shows a number of trials with variation in Hazard Radius and Mingle Factor, and the R0 values at the end of the trials.

Note that there are specific ranges in which combinations of Hazard Radius and Mingle Factor produce epidemics that terminate with the population all infected, for which we report R0 values. For Hazard Radius 4 and 5, we do not go beyond Mingle Factor values that produce R0 > 10.



The corresponding graph looks like this:



The important observations here are:

1. For different Hazard Radius values, the values of within 1 and 2 (black and red dotted lines) are different for the distinct Mingling Factors. For example, most values of R0 for Hazard Radius of 5 are greater than 2 when the Mingling Factors is greater than 0.5.
2. For Radius = 2, the values of R0 are consistently below 2, and for Mingling Factor < 1.8, the values of R0 are < 1. These represent Trials which terminate before the goal of 50% infected.
3. For each value of Hazard Radius, the starting point at which R0 is meaningful is different. This is because, below these limits, the Trials show that the simulations terminate very early or do not proceed at all (ie no contacts at all if the entities are very small and the movement is very small).

This kind of correlation gives us a guide to setting parameters for modelling epidemics of various characteristics. Conversely, if we had estimates of the nature of an epidemic (approximate R0), we could set the parameters of CovidSIMVL to approximate it.

ESTIMATES OF RISK PER HOUR (“RPH”)

A given trial may be set up to iterate until values are produced for a series of outcomes or goal states:

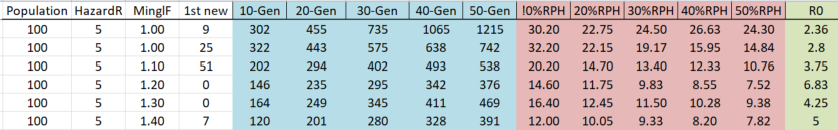
1. The first 10% of agents newly infected
2. The point at which 20% agents are infected
3. The point at which 30% agents are infected
4. The point at which 40% of agents are infected
5. The point at which 50% of agents are infected

Note that for certain configurations of parameters, a particular trial may end before a goal is reached. For example, the trial may end after only a few transmissions if there are no longer any infectious persons within the population specified for the trial.

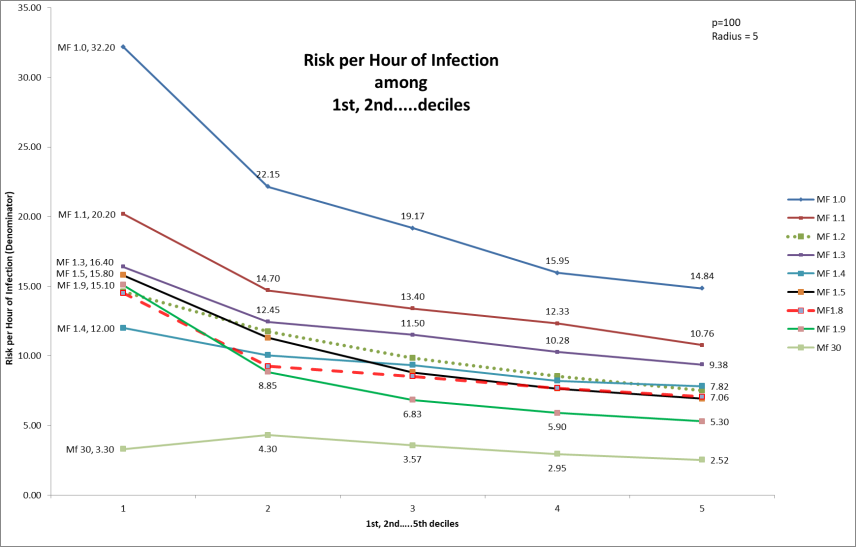
We create the metric “10%RPH” as follows: we track the number of generations in the trial to reach 10% of the population as Newly Infected. For populations of 100, the number would be 10 new infections. If this took 100 generations, then we have the following:

* the average number of generations for each newly infected = 100/10 = 10 generations
* within the first 10 new infections, the average span between infections is 10
* the risk is 1/10 that there will be an infection in a generation in an hour in this time segment
* the RiskPerHour (“RPH”) is defined as the inverse – ie 10
* the larger this is, the lower the chance that there is an infection in an hour (a generation)
* the smaller this number, the more likely an infection will occur In the hour
* with the progression of a trial, the RPH can be increasing or decreasing
* a falling RPH is an accelerating epidemic, while a rising RPH shows a decelerating epidemic

Considera trial that was set up to terminate as Newly Infected = 50%. At that point, the number of generations required to reach each interim/final goal state is recorded, as well as R0 at that point.



We calculate this metric at each of the points where 10% of the population are infected, 20%, 30%, 40% and 50%, in order to capture the number of generations the trial took to get those levels of infection.

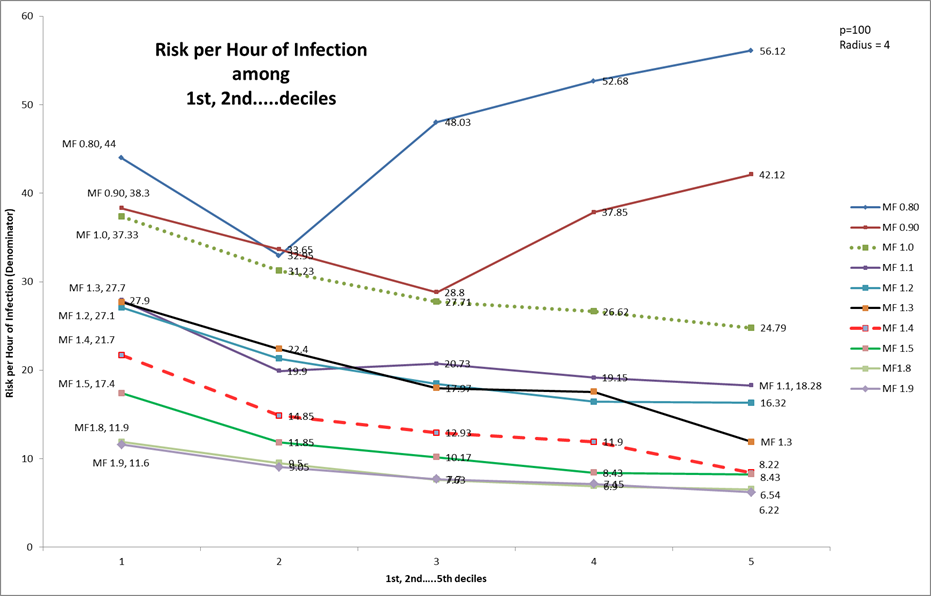


This graph is for a set Hazard Radius of 5. Each line is for a different value of Mingle Factor. The higher the Mingle Factor, the more rapid the epidemic, and our expectation is therefore we would get lower RPH values. Indeed this is what is charted.

The leftmost set of points, for the 10% decile, shows that a Mingle Factor of 1.0 produced an RPH of 32.20, monotonically decreasing to Mingle Factor of 30 (rapid epidemic) with an RPH of 3.30 (hours between infections) compared to 32.20.

For each line, as we advance in deciles along the trial, the values of RPH continue to fall. This indicates that the epidemic is accelerating (lower hours between infections) as it progresses. This is despite the removal of susceptibles due to transmissions, which shows that as the epidemic progresses, the ratio of transmitters to susceptibles continues to increase.

This is not necessarily the case for all epidemics. Consider the following graph for Hazard Radius = 4.



In this chart, the lines for Mingle Factor 0.80 and 090 fall to the second decile and then have upward slopes. This indicates that the values of RPH are rising as the trials progress. Recall that a higher RPH means that it takes longer to get the next infection in this time span.

This rising RPH means that as the epidemic progresses, new infections come slower and slower, so that this is the sign of an epidemic that is headed to self-extinction at best, and stability at worst.

It might be useful to observe that these Trials are for continuous hours for 24hrs per day, and with CovidSIMVL, we can model specific hours in which different groups of people (agents) with characterizably distinct transmission risk profiles (e.g., younger children vs high school aged children vs staff in long-term care vs patients in long term care) are located in a particular universe. At other times during the day, they may or may not move to other universes (and carry the infection with them if they are incubating or infectious).

Take the example of a classroom, which has a student in it for say 6 hrs a day. If the Risk per Hour is 30, or 1 in 30 hours, this might be optimistically interpreted to be 5 days at 6 hours/day with the risk being 1/30 per hour for an infection to take place. In 30 hours, we will get ONE infection, from this population.

By contrast, if the Risk per Hour for that student is 6, or 1 in 6 hours, this might be interpreted to be 1 infection per 6 hours of classroom time (1 school day), with a risk per hour of 1/6.

High Mingle Factors produce higher Risk per Hour metrics, because the agents are more mobile and they are more likely to make contact in any particular generation within the model. The susceptibles that are contacted and transform into infections diminish the total number of susceptibles in the starting population and add to the number of infectives. This results in a greater effective density of infectives within the population, and greater risk for any of the uninfected (susceptible) persons.

Where the decline is steep in Risk per Hour, we would expect R0 to be large, and where the Risk per Hour increases with deciles, we would expect R0 to be certainly less than 2, if not less than 1.

DISCUSSION

Recalling that CovidSIMVL is a simulation engine that models a PRIMARY set of rules (within-agent viral growth dynamics), a SECONDARY set of rules (between agent interactions in a fixed space – the subject of the above), it also has a TERTIARY set of rules, which govern the movement of populations between common spaces (which we call Universes).

The simulation tool is an acknowledgment that although we can specify the rules for primary, secondary and tertiary behaviors, we cannot easily determine the end-points or even the progress of the dynamics of the systems as a whole.

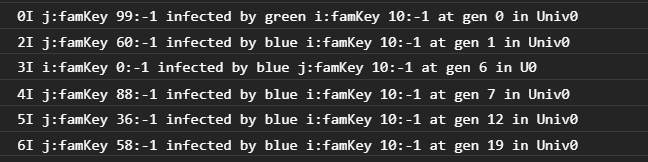
Thus, CovidSIMVL is a tool in which various parameters can be run and their resulting epidemics can be observed, with appropriate metrics. Indeed, the ability to change parameters and observe subsequent outcomes is fairly easily done where data retro-fitting models do not engage in these problem areas.

However, we understand epidemics by the traditional mathematical equation-based data retrofit methods, and thus the setting of parameters for CovidSSIMVL needs connection to the understanding of epidemics as they are commonly understood.

This is the purpose of the Calibration studies above. For a given fixed universe, and an epidemic progressing at a specific rate estimate of some R0 value, one can set Hazard Radius and Mingle Factors to reflect that R0 value (as initial conditions). The progress and prediction is not what is at stake; rather, the consequence of policies like restricting the flow between Universes, or the reduction of the duration of the infective symptomatic period by increased testing, and the quantitative results of the change in total infections, or infections per hour, can be derived.

APPENDIX. TRANSMISSION CHAINS

The console.log in CovidSIMVL produces, infection by infection, a trace of who infected whom in what generation (time), identifying the family membership, and preceded by the sequence number of the new infection. The console.log looks like this:



These data permit us to look more closely at the relationship between the length of the chains of transmission, and the frequency distribution of the numbers of infections that are incurred by infectious agents; in particular, whether there are super-spreaders as a side-effect of nature of contagion-based epidemics.

Here is a trace of infections with the parameters set with Population = 100, Hazard Radius = 5, and MingleFactor for the Universe of 2, with the agent base MingleFactor set at 3, and the resulting R0 at 50 infected persons being 4.61, which indicates an expanding epidemic.



continued…



These numbers are derived directly from the console.log, using Excel to delete unneeded text, and separating the numbers with the Excel option in “DATA” of “Text to Columns”.

The first set of columns are by order of generations, with the Person Number of the Infected in the first column, the Infecting Person Number in the second column, and the third being the generation number. In some instances, the same infection is recorded twice, as in generation 351, Person 21, but be assured it is only counted once.

The second set of columns duplicate the first set, but sorted using the “By” person number. This permits the easy tracking of a chain of transmission. For example, in the Generation set, the first entry shows Person 52 infected by Person 10. Going to the second set, and looking into the “By” column for 52, the entry shows that Person 66 was infected by Person 52, and we can continue this way.

The third set of columns is sorted by the “Infected” Person Number. This allows us to track the contagion in a basckward direction. For example, Person 95 in the first column, all the way down, was infected by Person 86, and in turn, locating 86, was infected by 60, and then finding 60, was infected by Person 6, and so on, finding 6 infected by 10, the Index case.

Using these data the longest chain was found to be:

10 🡪 30 🡪 55 🡪 41 🡪 59 🡪 0

while the shortest was just

10 🡪 71

Tracking the tree of infections from the Index Case Person 10 through the 568 generations, we find that the distribution of Infectiveness is:



The most frequent infectious intensity is one infection (16 of these events) while Person 30 infected 7 persons, and Person 10 (the Index Case) infected 6 persons. These do not tell us when these chains occurred.

The distribution of chain lengths (from the Index Case to a leaf in the transmission tree – either a node has become inert, or it has not infected any – had no descendants at the termination of the trial) are as follows for this trial:



We can see that the most frequent length of transmission trains was 4, with 10 occurrences, and next were chains of length 3, with 8 such.

In expanding epidemics, the overall rate of infection is higher, so that one would expect shorter chains compared to epidemics that progress slowly or are self-extinguishing. Similarly, one would expect a higher average of infections per infectious agent, with a more uniform distribution of infections per agent.

As infectiousness is in CovidSIMVL determined by Hazard Ratio and Mingling Factor for defined population densities, it would be interesting to identify the distributions of chain length and infectivity of agents over time, for varying parameters of Hazard Radius and Mingling Factors.

If these can be characterized, it may then be possible to take tracking data from the field as samples into these distributions, and use the distribution of the chain lengths of the samples, fit them to the distributions from the synthetic epidemics, in order to estimate the hidden true population of infected.