Epidemiological parameters in OO

(Document version 1.0, July 22, 2024)

This document presents a detailed derivation of the formulas that are used to calculate the epidemiological parameters that drive the OO simulations.

Model of disease progression

The model of disease progression in OO can be summarized in the following diagram:

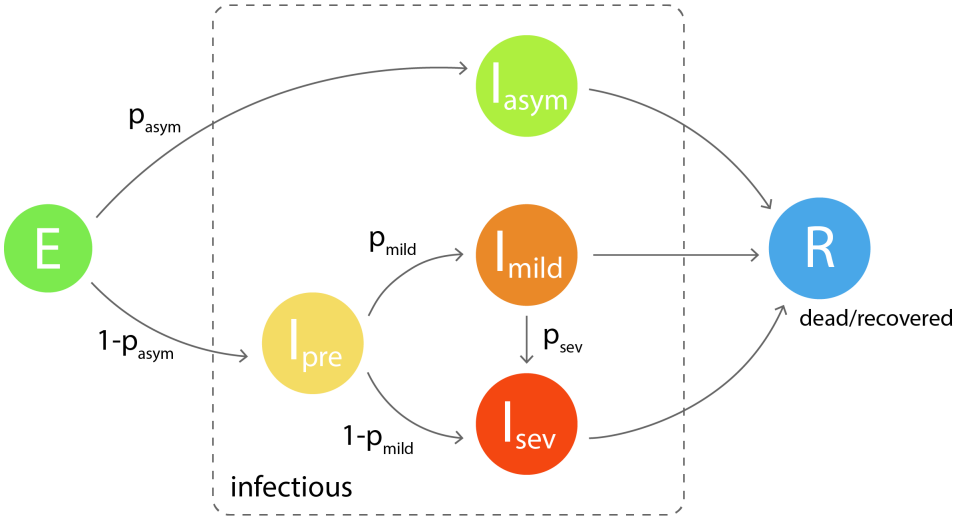


Figure 1: Model of disease progression during an infection, from exposure to removal.

According to this model of disease progression, infected individuals start in the exposed (E) state, where they are not infectious, and from there they move on to the infectious state. The infectious state contains a sequence of possible substates depending on the symptomatology of the individual; first, the individual can become fully asymptomatic but infectious (I_{asym}) with a probability p_{asym}. If this does not happen, with probability $1 - p_{asym}$ the individual will turn into symptomatic, going through a pre-symptomatic state (I_{pre}), which can evolve into symptomatic mild (I_{mild}) or symptomatic severe (I_{sev}) with probability p_{mild} and $1\text{-}p_{mild}$, respectively. An individual in the symptomatic mild state can become severe with probability p_{sev}. From all these states, I_{asym}, I_{mild}, and I_{sev}, the individual finally becomes removed, by either recovering or dying.

Parametrizing a simulation given the disease and the simulation duration and "speed"

In this section we will present a formula that will provide an easy way to parametrize an OO simulation given the parameters of a real-world disease

The mean total duration of an infection can be written as:

$$d_{infection} = d_E + d_I$$

where d_E is the mean duration of the exposed period and d_I the mean duration of the infectious period. So, we can write:

$$d_{infection} = d_E + p_{asym} \times d_{asym} + (1 - p_{asym}) \times d_{sym}$$

Here, d_{asym} and d_{sym} are the mean total durations of the fully asymptomatic and symptomatic infectious periods. Following the transitions in Figure 1 for the symptomatic route:

$$\begin{aligned} d_{sym} &= d_{pre} + p_{mild} \times d_{mild} + (1 - p_{mild}) \times d_{sev} + p_{mild} \times p_{sev} \times d_{sev} \\ &= d_{pre} + p_{mild} \times d_{mild} + (1 - p_{mild} + p_{mild} \times p_{sev}) \times d_{sev} \end{aligned}$$

With d_{pre}, d_{mild}, and d_{sev} the mean duration of the pre-symptomatic, mild, and severe infectious periods.

Given the total duration of a simulation, d_{total}, and $d_{infection}$ the mean total duration of an infection as before, we now define the following quantity:

$$N_{inf} = \frac{d_{total}}{d_{infection}}$$

This number simply indicates how many consecutive infections would fit within one simulation. The goal is to adjust the mean total duration of an infection, $d_{infection}$, to reach a desired N_{inf} for a target simulation time. The higher N_{inf}, the simulation would have a faster pace or "speed", since individual infections would need to be shorter to fit.

But $d_{infection}$ can be expressed as a function several elementary quantities, by combining the formulas above:

$$\begin{aligned} d_{infection} &= d_E + p_{asym} \times d_{asym} + \left(1 - p_{asym}\right) \times d_{sym} \\ &= d_E + p_{asym} \times d_{asym} d + \left(1 - p_{asym}\right) \times (d_{pre} + p_{mild} \times d_{mild} \\ &\quad + (1 - p_{mild} + p_{mild} \times p_{sev}) \times d_{sev}) \end{aligned}$$

If we define the following ratios:

$$r_{em} = \frac{d_E}{d_{mild}}$$

$$r_{am} = \frac{d_{asym}}{d_{mild}}$$

$$r_{pm} = \frac{d_{pre}}{d_{mild}}$$

$$r_{sm} = \frac{d_{sev}}{d_{mild}}$$

we can express $d_{infection}$ as function of d_{mild} and the ratios of the rest of the infection periods with respect to the duration of the mild infectious period:

$$d_{infection} = d_{mild} \times [r_{em} + p_{asym} \times r_{am} + (1 - p_{asym}) \times (r_{pm} + p_{mild} + (1 - p_{mild} + p_{mild} \times p_{sev}) \times r_{sm})]$$

Combining this with the definition of N_{inf}, we get:

$$d_{mild} = \frac{d_{total}}{N_{inf} \times [r_{em} + p_{asym} \times r_{am} + (1 - p_{asym}) \times (r_{pm} + p_{mild} + (1 - p_{mild} + p_{mild} \times p_{sev}) \times r_{sm})]}$$

If we define the following auxiliary variable:

$$R = r_{pm} + p_{mild} + (1 - p_{mild} + p_{mild} \times p_{sev}) \times r_{sm})$$

we can finally arrive to the following expression:

$$d_{mild} = \frac{d_{total}}{N_{inf} \times [r_{em} + p_{asym} \times (r_{am} - R) + R]} \quad (1)$$

The advantage of this formula is that, given a real disease with known infection periods measured typically in days or weeks and probabilities of mild, severe, asymptomatic illness, etc., we can calculate the ratios of all those periods with respect to the duration of the mild infectious period, and then calculate d_{mild} as a function of the total simulation time, $d_{total,}$ and, given by N_{inf}. Then, we can get the rest of the periods according to the simulation scale by using the rates defined before.

Calculating the probability of transmission

The other key parameter in OO simulations is the probability of a transmission between an infectious and a susceptible player. During a contact event between two phones, one infected and the other susceptible (and the event defined as the phones being under a threshold contact distance specific for the disease being simulated), the probability that the former infects the latter is p every Δt seconds, where Δt is the simulation time-step, currently at 2 seconds.

The problem is to estimate p so that the simulation would have a desired R_0 given the total duration of a typical contact, the anticipated contact rate between players, and the mean duration of the infectious period (which we calculated in the previous section).

We have the following basic relationship from mathematical epidemiology[1]:

[1] Jones, JH. Notes On R0. (2007). https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf

$$R_0 = \tau \times \bar{c} \times d_I$$

Where τ is the transmissibility (probability of infection given a contact between susceptible and infectious individuals), $\bar{c}$ is the mean contact rate between susceptible and infected players, and d_I is the mean duration of the infectious period.

We can write τ in terms of p and the mean total duration of a contact between susceptible and infectious players, $\bar{d}_c$:

$$\tau = p \times \bar{d}_c$$

$\bar{d}_c$ must be expressed in Δt units, so the resulting quantity is dimensionless, as it is a probability (with p being a probability of infection per Δt unit of time). For example, if the mean total duration of a contact is 1 minute, and the simulation time-step is 2 seconds, then $\bar{d}_c = 30 \times 2\ sec$. And if p were to be *0.01/2 sec*, then $\tau = \frac{0.01}{2sec} \times 30 \times 2\ sec = 0.3$. (The full dimensional analysis of these formulas will be presented in the notes at the end of this document).

We can combine the two last equations to arrive to an expression for p:

$$p = \frac{R_0}{\bar{d}_c \times \bar{c} \times d_I}$$

The infectious period, d_I, can be written as:

$$d_I = p_{asym} \times d_{asym} + \left(1 - p_{asym}\right) \times d_{sym} = d_{mild} \times [p_{asym} \times (r_{am} - R) + R]$$

And so:

$$p = \frac{R_0}{\bar{d}_c \times \bar{c} \times d_{mild} \times \left[p_{asym} \times (r_{am} - R) + R\right]}$$

There is an important consideration with regards to the mean contact rate, $\bar{c}$. It assumes that all members of the population or group of individuals in the area where the OO simulation is taking place are part of the simulation. If not, the effective contact rate will be smaller, as result of fewer individuals coming into contact. We can name the adoption rate of OO as $r_{adoption}$, a number between 0 and 1 to denote from 0% to 100% adoption.

Another source of error is the imprecision in the Bluetooth sensing. Short-lived contacts may not be detected by the OO app, especially when running on older phones. So, we can add another factor, $r_{detection}$ (also between 0 and 1) that would quantify the contact detection rate of the app.

If $\bar{c}$ represents the contact rate for the entire population/group assuming perfect contact detection, then the effective contact rate could be written as:

$$\bar{c}_{eff} = \bar{c} \times r_{adoption} \times r_{detection}$$

If we replace $\bar{c}$ with $\bar{c}_{eff}$ in the previous formula for p, we can write:

$$p = \frac{\tilde{R}_0}{\bar{d}_c \times \bar{c} \times d_{mild} \times \left[p_{asym} \times (r_{am} - R) + R\right]} \quad (2)$$

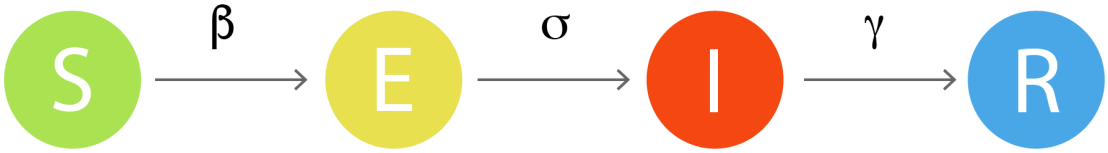
with an "scaled" R_0:

$$\tilde{R}_0 = \frac{R_0}{r_{adoption} \times r_{detection}}$$

The advantage of this formulation is that it makes clearer the effect of an adoption rate below 1 as an increase in R_0 to compensate for the lower contact rate among OO players compared with the entire population and imperfect contact detection using Bluetooth.

Calculating the parameters in the SEIR model

The OO transmission model corresponds to susceptible (S), exposed (E), infectious (I), and removed (R) compartments with the following rates between them:



The rates β, σ, and γ can be written down in terms of the exposure and infectious periods as follows:

$$\beta = R_0 \times \gamma = \frac{R_0}{d_I}$$
$$\gamma = \frac{1}{d_I}$$
$$\sigma = \frac{1}{d_E}$$

These rates can be readily calculated using the formulas for d_I and d_E:

$$d_I = d_{mild} \times \left[p_{asym} \times (r_{am} - R) + R\right]$$
$$d_E = d_{mild} \times r_{em}$$

NOTES

1. Using different time units in the calculations

Even though the time-step of the simulation is in Δt seconds unit, which we can denote as $[TS] = \Delta t\ sec$, the simulation durations are often expressed in more convenient units, such as hours or minutes, that we can represent as $[S_u]$ *(=1hr, 1 min, etc.)* On the other hand, the contact rate and contact duration will be given in another unit, most likely minutes, $[C_u]=1\ min$. How to properly apply formulas (1) and (2) given all these different units? We can apply dimensional analysis to make sure that we do not make any systematic error by using wrong units.

So, the unit of the probability of transmission per simulation time-step is the following:

$$[p] = \frac{1}{[C_u] \times \frac{1}{[C_u]} \times [S_u]} = \frac{1}{[S_u]}$$

To express p in $[TS]$ we can apply the following relationship:

$$[S_u] = C_{STS}[TS]$$

where C_{STS} is a constant that is just the ratio between the simulation unit $[S_u]$ and the simulation time-step unit, $[TS]$, to arrive to:

$$[p] = \frac{1}{C_{STS}[TS]} \tag{3}$$

So, for example, if $[S_u]=1hr$ and $[TS]=2\ sec$, we would have:

$$C_{STS} = \frac{3600sec}{2sec} = 1800$$

Dimensional equation (3) implies that after calculating p using $[S_u]$ and $[C_u]$ units, we would need to divide the resulting value by 1800 to express it in $[TS]=2\ sec$ units.

2. Writing down the formulas using calculation spreadsheet labels/cells

The parameter calculation spreadsheet available in the repository below:

https://github.com/colabobio/oo-parameters/blob/main/par-calculator.xlsx

In this spreadsheet, the variables used in equations (2) and (3) are named as follows:

- F3 (Total duration of simulation) = d_{total}
- F6 (Number of consecutive infections during simulation) = N_{inf}
- C4 (exposed_to_inf_mild_ratio) = r_{em}
- C13 (ASYM_PREVALENCE) = p_{asym}
- C8 (asym_to_inf_mild_ratio) = r_{am}
- C18 (HELPER VALUE R) = R
- C20 (HELPER VALUE R0 scaled) = $\tilde{R}_0$
- F10 (number of minutes for contact rate) = denominator in calculation of $\bar{c}$
- F12 (Mean contact duration) = $\bar{d}_c$
- F9 (Number of contacts for contact rate) = numerator in calculation of $\bar{c}$
- C19 (HELPER VALUE D, duration of infectiousness) = d_I
- F15 (Conversion constant (sim-time-unit-to-sim-step)) = C_{STS}

Using these correspondences, we can write (2) and (3) as:

- C23 (MILD_PERIOD) = F3/(F6*(C4+C13*(C8-C18)+C18))
- C24 (PROB_INFECTION_FROM_PEER) = (C20*F10)/(F12*F9*C19*F14)

3. Discrepancy with older formulas

Older versions of the parameter calculation spreadsheet had a formula to calculate the probability of infection as follows:

p_{old} = C11/((F9)/(F10)*(C13*(1-C14)+C13*C14*(1+C6)+(1-C13)*C6)*C27*F12)

with the following values:

- C11 = basic_reproduction_number
- F9 = Number of contacts
- F10 = in x minutes
- C13 = PROB_MILD_FROM_ASYM
- C14 = ??? (often 0)
- C6 = inf_severe_to_inf_mild_ratio
- C13 = PROB_MILD_FROM_ASYM
- C6 = inf_severe_to_inf_mild_ratio
- C27 = MILD_PERIOD
- F12 = conversion_constant (hour-to-min) = 60

Re-writing the spreadsheet formula below

$$p_{old} = \frac{R_0}{\bar{c} \times d_{mild} \times \left[p_{asym} + \left(1 - p_{asym}\right) \times r_{sm}\right] \times 60}$$
$$= \frac{R_0}{\bar{c} \times \left[p_{asym} \times d_{mild} + \left(1 - p_{asym}\right) \times d_{sev}\right] \times 60}$$

We could identify the term $p_{asym} \times d_{mild} + \left(1 - p_{asym}\right) \times d_{sev}$ with d_I, although it is not correct. So, we could say that

$$p_{old} \sim \frac{R_0}{\bar{c} \times d_I \times 60}$$

while in the new formula derived in the pages above, we reached the following, if we ignore $\bar{d}_c$ for the time being (which seems to be of the order of 1 min):

$$p = \frac{R_0}{\bar{c} \times d_I \times 1800}$$

With 1800 the dimensional scaling constant to express p in $(2\ \text{sec})^{-1}$ units (probability of transmission per simulation time step).

A similar result appears if one looks the probability from peer calculation in the old OO admin code:

https://github.com/colabobio/oo-editor/blob/main/src/utils/equation.utils.ts#L7

which results in the following equation:

$$p_{old} = \frac{R_0}{\bar{c} \times [p_{mild} \times d_{mild} + p_{mild} \times p_{sev} \times d_{sev} + (1 - d_{mild}) \times d_{sev}] \times 60}$$

Here, the expression $p_{mild} \times d_{mild} + p_{mild} \times p_{sev} \times d_{sev} + (1 - d_{mild}) \times d_{sev}$ is a more accurate calculation of d_I, but still not correct since it's missing the duration of the pre-symptomatic and asymptomatic periods. However, the main discrepancy between the old and new calculations of p is the scale factor that results in:

$$p_{old} = 30 \times p$$

However, this discrepancy can be alleviated by setting the $r_{adoption}$ and $c_{missing}$ parameters to reasonable values that reflect the setting of the simulation. For example, if only 30% of the students in a school are participating of an OO simulation (with the participating students still interact with the non-participating students during the simulation) and we assume that only half of the contacts are being detected, then $\tilde{R}_0 = \frac{R_0}{0.3 \times 0.5} \sim 7 \times R_0$, so the resulting p would be 7 times larger than the original value without the scaling, and now only ~4 times smaller than p_{old}.