Implementation Details

1 Within Node

A node is a collection of agents assigned to a spatial location with latitude and longitude. The number of agents within a node corresponds to the total population for an area, but that area is not used in any calculations and is only present in model files as metadata.

Each agent is either susceptible, latent (i.e., infected non-infectious), infectious, or recovered. In EMOD, the latent period is incorrectly labeled as the incubation period, which typically refers to infected *non-symptomatic*. This model does not track symptomatic states.

Both the latent and infectious states are implemented using timers. On infection, an agent is assigned a latent duration and infectious duration drawn from distributions. The latent duration is drawn from a Gaussian distribution with $\mu = 3$ days and $\sigma = 1$ day. The infectious duration is drawn from a gamma distribution with a mean of 24.0 days and standard deviation of 11.3 days. The gamma distribution is implemented using shape = 4.51 and scale = 5.32. An important feature of both these distributions is that they are non-integer valued. All time steps in the model are integer time steps; to ensure that the infectious period is not asymmetrically censored, both the latent and infectious durations need to be non-integer values.

The amount of infectivity shed each day by an infectious agent is also drawn from a distribution. This value is a property of the infection and does not change with time (i.e., the random value is assigned once for each infected agent, and is not re-determined on each time step). The average total amount of infectivity shed by an infectious agent defines R_0. In the model, the value of R_0 is an adjustable parameter (currently 14.0), so the mean value of this distribution of daily infectivity is fixed at 14.0 / 24.0 days. Daily infectivity is implemented using a gamma distribution with shape = 1.0; equivalent to an exponential distribution.

For an infectious agent, the rate of shedding does not currently vary with time. The rate of shedding as a function of the time-course of infectiousness is multiplied by the probability density function of a beta distribution. This implementation was selected to ensure that cumulative shedding is always normalized. Default parameters for the beta distribution are $\alpha = \beta = 1$, so uniform shedding between the start and end of the infectious period.

On each time step, the total amount of infectivity shed by all infectious agents within a node defines the total within-node infectivity. A fraction of this within-node infectivity is removed and distributed to other nodes, and some additional infectivity is collected from other nodes. (See 2)

After this (net-zero) spatial redistribution, additional modifiers are applied. Seasonality increases the infectivity by 20% between days 165 and 245 each year (June 13 to September 1). All other days are modified by -5.7% so seasonality applies a non-zero overall bias. Geographic modifications are non-zero bias and vary by state. These geographic multipliers to node infectivity are calculated using estimates of underweight fraction from rasterized data of child growth failure from IHME:

$$\textit{Geographic } R_0 \textit{ multiplier} = \frac{1.0}{1.0 + e^{24.0(\bar{u}-u)}} + 0.2 \quad (1)$$

Fraction underweight (u) and average fraction underweight ($\bar{u}$) are values estimated from IHME data. The average fraction underweight is for Nigeria only and based on state-level estimates. Values 24.0 and 0.2 are adjustable parameters.

After seasonal and geographic infectivity multipliers, the node infectivity is used to determine new infections. The modified total node infectivity is interpretable as a Poisson rate of total node infections. (Over-dispersion of the Poisson rate may be applied here. It does not influence mean outcomes for endemic transmission, but would be relevant for simulations examining elimination probability.) This total node infection rate is divided by the total number of agents within the node and interpreted as the per-agent infection rate (λ_a). Using the probability mass function of the Poisson distribution, the probability of an agent not being infected (infected zero times) is e^{λ_a}, so the probability of an agent being infected is $1 - e^{\lambda_a}$. This expression is the base probability for each agent becoming infected on a given time step.

The probability for an agent becoming infected on a given time step is further modified. If an agent is currently infected or has recovered from infection, the probability is zero. Each agent has an individual risk multiplier that is drawn from a log-normal distribution. This distribution has mean of 1.0 and variance of 4.0; the variance is an adjustable parameter, but the mean is fixed at 1.0 to avoid biasing the total infection rate. The variance can be modified based on geography, but is currently spatially uniform.

Incorporating positive covariance between an individual's risk multiplier and the infectivity shed daily when infectious is an important component of the model. It allows for smaller outbreaks to occur within a node and avoids all-or-nothing dynamics. When conceptualizing a node as a directed graph with individuals as points and arrows as edges, arrows outbound from a point are infectivity while arrows inbound to a point are risk. (Infectivity and risk are both continuous, so maybe they're actually the underlying rates of edge generation.) Since

a node is a *directed* graph, the connections aren't symmetric. If inbound and outbound arrows are distributed independently, then all parts of the graph are functionally similar. With positive covariance, points with more inbound arrows (risk) will also have more outbound arrows (infectiousness), and sections of the graph will high connectivity will tend to sustain transmission when other sections of the graph do not.

Implementing positive covariance will depend a lot on the software. In EMOD, the risk distribution is applied first and each agent, at creation, is assigned an intrinsic risk multiplier. If that agent becomes infected, then the random value selected as daily infectivity is biased according to the agent's risk multiplier and a correlation coefficient. The correlation coefficient is adjusted to match a target covariance value. Correlated random numbers are hard to implement, and ensuring positive covariance between the distributions is the important outcome.

There is no age structured transmission (e.g., age dependent contact matrices) or route-based transmission (e.g., household, community, etc.)

2 Between Nodes

All nodes are approximated as points with a given latitude and longitude.