**Simulated Outbreaks for testing randomized NPI evaluation methods**

Lee Kennedy-Shaffer

May 2020

*Simulation Code Details:*

Simulated\_Outbreak.R code runs an outbreak simulation with an option to add individually randomized trial (IRT), parallel-arm cluster randomized trial (CRT), or stepped-wedge cluster randomized trial (SWT).

* Used by [Kennedy-Shaffer and Lipsitch 2020](https://www.medrxiv.org/content/10.1101/2020.05.01.20087429v1), based on [Hitchings et al. 2018](http://doi.org/10.1093/aje/kwy047) and adapted from [code by Matthew D. Hitchings](https://github.com/mhitchings/Code).

**Key Simulation Features and Corresponding Parameters:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Simulation Feature** | **Relevant Parameters in Code** | **Code Lines with Relevant Parameters** | **Values Used in K-S & L 2020** | **Potential Values/Data Sources for NPIs** |
| Number of Simulated Trials | nsim | 24 | 1,000 | Depends on goal |
| Community Size | ave\_community\_size, community\_size\_range, community\_sizes | 28, 30, 119 | Mean of 100, range of 40, uniform distribution | Distribution of city, county, or region sizes |
| Number of Communities | num\_communities | 32 | 40 | Depends on goal; may want to vary to see power under different combinations of sample proportion and number of communities |
| Network structure | rate\_within, rate\_between, within\_rates, between\_rates, g, potential\_contacts | 34, 36, 123, 124, 125, 237 | 0.15, 0, all communities have same rate, natural clustering within communities | Depends on contact rate assumptions and community size |
| Force of Infection/R0 | beta, R0 (calculated from beta, contact rates, community sizes, infectious periods) | 41, 84 | 0.02, 0.03, 0.04, 0.05 (for R0~1.5,2,2.5,3) | Tuned to give desired R0, vary for different assumptions about other control measures |
| Importations to seed community outbreaks | num\_introductions | 43 | 80 (expectation of two per community over two years) | Depends on goal; may wish to start instead with deterministic seeding of 1 or 10 individuals within each community, with seeding at random times |
| Vaccine efficacy | direct\_VE | 45 | 0, 0.6, 0.8 | Lower range of values; may also want to apply effect to number of contacts instead |
| Incubation Period | incperiod\_shape, incperiod\_rate, ave\_inc\_period, FixInc, incperiods | 47, 48, 51, 52, 313, 315 | 5.807, 1/.948, 6, 0, Gamma distribution (based on Lauer et al. 2020) | Likely similar, although may want to separate incubation and latent period |
| Infectious Period | infperiod\_shape, infperiod\_rate, inf\_periods | 49, 50, 202 | 1.13, 0.226, Gamma distribution (gives 5-day average) | Probably increase somewhat based on newer data |
| Trial timeline | trial\_startday, trial\_length | 54, 56 | 56, 308 | Start day depends on importations for where it is in outbreak timeline; length depends on question of interest |
| CRT and SWT: proportion of each community enrolled in trial | cluster\_coverage | 60 | 0.5 | Depends on intervention. May need to split this into coverage of intervention and coverage of outcome measurement |
| CRT: enrollment timeline | num\_clusters\_enrolled\_per\_day, enrollment\_period | 65, 68 | 40, 1 | Depends on intervention timing. |
| SWT: enrollment timeline | step\_interval, first\_crossover, num\_clusters\_per\_step | 72, 74, 76 | 28, 84, 4 | Depends on intervention timing. |

* Overall Function of the Code:
  + Running the simulations begins on line 900. Right now, it’s set up for each simulation to generate a network structure and then run an IRT, a CRT, and an SWT. It stores the results (total number of infected nodes and days of infection, as well as full accounting of nodes enrolled into the trial) in fullRes\_IRT, \_CRT, and \_SWT. Note that DayInfected is the day that the individual became infectious/symptomatic (assumes no asymptomatic infections and no presymptomatic transmission right now). If you’re running more than one simulation, you need to designate a way to save the results of each simulation.
  + It also stores the effect estimates and p-values for the IRT and CRT, as well as total number infected in each arm for all 3 trial types, for each simulation in OverallRes. This code does not analyze the SWTs—that code is a different file that I’m cleaning and can send when it’s ready, although may not be needed for this project.
  + It takes about 1.5 minutes to run one simulation with the three trial types on my computer.

*Use Cases for the Simulations for NPI evaluation:*

1. For the estimation of beta by the ratio of infections in the treated clusters compared to the control clusters, the variance of the estimator will depend on the variance of the number of infections across clusters, in both the treatment and control conditions. We may be able to tune this model to match some empirical data for the control condition, and then use it, implementing the NPI in the model, to get at the variance for the treated clusters.
2. In the estimation with a pre- and a post- measurement, this model can inform the relative variances of the pre- and post- outcomes depending on the various parameters (and thus how much variation the pre-measurement can soak up from the estimator). Similarly, it could be used to see how much variation stratification or matching on cluster-level characteristics may soak up.
3. In both of these cases, assess the tradeoffs between measuring more clusters and measuring a larger portion of outcomes within each cluster.
4. Evaluate how well estimates of final size based on two measurements perform.
5. Adding more complications, this model may be useful to see how various levels of adherence to interventions affect the impact on beta and the estimate of that. E.g., if transmission is largely spread by super-spreading events and adherence to some intervention is uneven, how does that translate into overall effect on beta/R0?