

Outbreak Simulator: Building systems thinking and data science literacy with an interactive simulation of infectious disease

Carson Rueber, University of Idaho, rueb8071@vandals.uidaho.edu
 Terence Soule, University of Idaho, tsoule@uidaho.edu
 Landon Wright, University of Idaho, landonw@uidaho.edu
 Roger Azevedo, University of Central Florida, Roger.Azevedo@ucf.edu
 Daryn Dever, University of Central Florida, Daryn.Dever@ucf.edu
 Barrie D. Robison, University of Idaho, brobison@uidaho.edu

Abstract: Outbreak Simulator is a tool that allows users to simulate the spread of an infectious disease and explore the impact of different interventions. The simulation can be used to promote systems thinking and data science literacy, as users can see how different factors interact to influence the spread of the disease and how data can be used to track, visualize, and understand the outbreak. Outbreak Simulator can be used in educational, research, and public health settings to help people gain a better understanding of infectious disease dynamics and the role of data in outbreak response.

Infectious Disease, Systems Thinking, and Public Education

The SARS-CoV-2 pandemic highlighted that many people were not sufficiently equipped to understand the complex nature of infectious diseases and the uncertainties involved in managing them. This lack of public understanding made people more vulnerable to misinformation and disinformation, resulting in the formation of misconceptions about infectious diseases and public health. Pandemic mitigation strategies and compliance with public health guidance suffered as a result. Equipping the public to better understand infectious diseases requires integration of multiple complex systems, including pathogen biology, the human immune system, and systems of human behavior that are influenced by economic, social, and environmental factors. Tools that improve domain knowledge, systems thinking skills, scientific reasoning, and data literacy (e.g., simulation, data visualization) could help better prepare the public to grapple with the complexity of infectious disease dynamics.

Computer simulations with interactive visualizations of dynamic phenomenon are a potentially useful vehicle for presenting infectious diseases as a complex system. This promotes systems thinking, which is the ability to see connections and relationships between different factors and can help people understand how small changes in one area can have large effects on the overall system. Adopting a systems thinking framework can demonstrate dynamic relationships between components of a larger whole (Von Bertalanffy & Sutherland, 1974), multi-level heterogeneity of agents, nonlinear dynamics, stochasticity, and emergence of new patterns (Eckhardt et al., 2020; J. H. Miller & Page, 2009; Pearce & Merletti, 2006; Roux, 2011). Systems thinking is essential in understanding the complex dynamics of infectious disease and can help to promote better decision-making and problem-solving. To meet this need, we created Outbreak Simulator (Table 1), a game-like interactive simulation that uses engaging grade-appropriate scenarios of infectious disease to improve domain knowledge (viruses and infectious diseases), systems thinking (complexity, component relationships, emergent behaviors), and data literacy (interpretation of visualizations and argumentation from evidence).

Table 1
Key features of Outbreak Simulator.

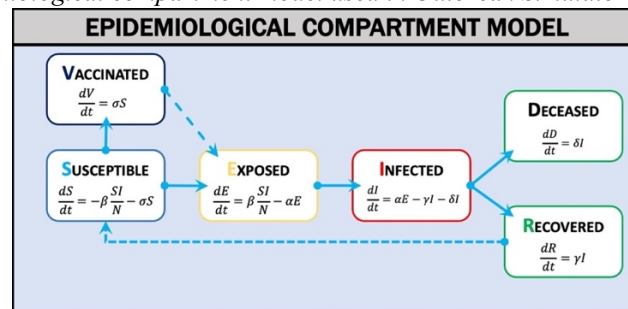
Feature	Description
Development Environment	Unity 2020.3
Platforms	PC (Mac and Linux coming soon)
Maps	Continental United States, Global
Movement Models	Diffusion, Gravitation (Simini et al., 2021), Major Airports
Disease Scenarios	Influenza (1918, custom), <i>SARS-CoV-2</i>
Fictional Scenarios	Zombies (T-virus [ref], Walking Dead [ref])
Model Types	Deterministic, Stochastic (Gillespie, 1976)
Model Compartments	Susceptible, Exposed, Infected, Vaccinated, Recovered, Removed
Data Layers	Population, Elevation, Roadways, Water, Vaccination Rate (MMR)

Simulation Description

Mathematical Model of Infectious Disease

Outbreak Simulator uses a compartment model of disease (Weissman et al., 2020) in which the population is divided into categories (Figure 1): Susceptible (S), Exposed (E), Infected (I), Vaccinated (V), Recovered (R), or Deceased (D). The model estimates the rates of exchange between categories over a given time interval (t) using differential equations. When the model parameters are known and key assumptions are met, the differential equations can estimate the epidemic curve of an outbreak. The two most critical assumptions are that the population is homogeneous and well mixed and is fixed in size. Deterministic and stochastic discrete time (Gillespie, 1976) versions of the models offer simpler but computationally performant (deterministic model) and mathematically rigorous but computationally slower (stochastic model) options to use in different educational contexts. We have pre-parameterized scenarios for influenza (multiple strains) and SARS-CoV-2 (multiple variants) and are creating more for additional diseases including Ebola, Measles, Mumps, Rubella, and the common cold. Outbreak Simulator can also be used to model fictional disease systems. For example, many science fiction and horror fiction treatments of zombies are based on fictional viruses (e.g. World War Z, Resident Evil, Walking Dead), and these scenarios could prove to be engaging learning environments for some users (E. T. Lofgren et al., 2016; Smith, 2009).

Figure 1
Epidemiological compartment model used in Outbreak Simulator.



Geospatial Representation:

The compartment model described above assumes a homogeneous and well mixed population - an assumption that is clearly violated when we consider the geographic scope of many infectious diseases. For example, consider the population of the continental United States. There are many factors within the US population that affect the mixing of individuals, including transportation patterns at the local and national scale. While modern transportation increases the mixing of the US population compared to previous pandemics (such as the 1918 influenza outbreak), we have seen that social distancing and travel restrictions can profoundly affect transmission dynamics at the national scale. There can also be considerable geospatial variation in the parameters of the model itself. For example, contact rate (number of individuals that an infectious individual contacts per time interval t) is likely to vary between urban and rural environments. Even seasonal outbreaks of influenza and the common cold can vary depending on spatial and temporal variation of temperature and host health (E. Lofgren et al., 2007). These spatial effects on the modeling of disease can be important in showing learners how outbreaks might affect their region, and how spatial heterogeneity adds complexity to model predictions.

To represent the spatial effects of infectious disease dynamics, our simulation models the population as a gridded set of subpopulations. Outbreak Simulator incorporates also uses spatial data layers (Table 1) to modify the parameters of the disease and movement models for each grid cell. We have installed two movement models - a diffusion model and a gravitation model (Simini et al., 2021). We use these models for commuting scale movements between adjacent grid cells. For longer (national) movement, we use a data layer of the travel rates between the 150 largest airports in the continental US.

Embedded Data Visualizations

Outbreak Simulator uses interactive visualizations intended to help users understand the dynamics of disease ecology, practice scientific inquiry skills, and build data literacy. Figure 2 is a screen shot of Outbreak Simulator, annotated with letters which indicate system components and different visualizations. *INFECTED*, *EXPOSED*,

and *DECEASED* individuals are indicated in area (I) on the map by color. The interactive map (A) can be panned, zoomed, and displayed with different data (satellite with population data overlay shown). Elapsed time, speed controls, help, and settings are accessed with a collapsible panel (B). Display settings such as color and visibility for each model compartment on the map can also be customized (C). Other features include a population health bar (D), buttons for initiating an outbreak, an Interventions menu (E), and an Event Log that describes simulation milestones such as initial outbreak location and timestamps of infection events (F). The intervention option allows the user to test how choices related to public health guidance (such as air travel restrictions, local lockdowns, masking, and vaccination) might impact outbreak dynamics.

Figure 2
Outbreak Simulator screenshot.



Data visualizations in Outbreak Simulator are interactive, dynamic, and inter-related. For example, the spatial dynamics of the outbreak can be observed by watching the expansion of red (infected) color across the map (Fig 1 I), by scrolling through the event log (Fig 1 F), or by viewing the epidemic curve of a selected state (Fig 1 G). The dynamic graphs (Fig 1 G and H) are interactive in that each model compartment can be toggled and the scale dynamically adjusted. Hovering over the graph shows the exact values for all model component at the hovered time point (Fig 1 G).

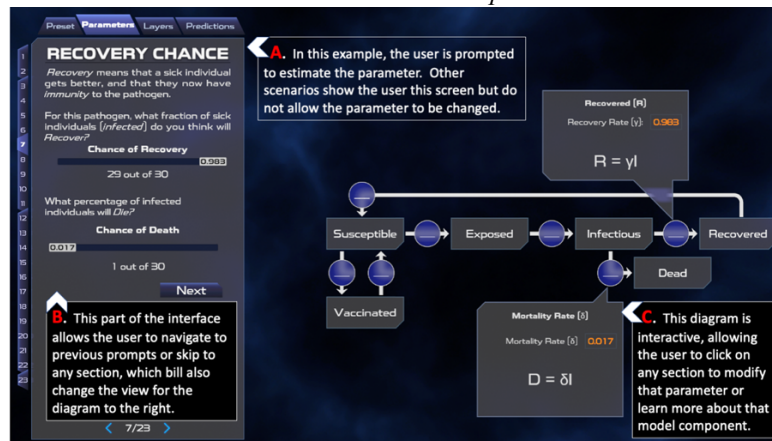
Outbreak Simulator parameterizes infectious disease dynamics with an interactive mental model. The beginning of each scenario provides grade appropriate information about the disease. The user is then guided through model parameterization using a drag and drop interface and plain language narrative-based scaffolding (Figure 3). Figure 3 shows a scenario in which the model parameters represent the 1918 influenza strain (Chowell et al., 2007). The interface allows the user to change model parameters using sliders or direct numerical input.

Interactive Demonstration Structure.

Using one of several provided laptop computers, participants will run through pre-parameterized scenarios for real diseases (Influenza, SARS-CoV-2) or fictional pathogens (such as viral zombie outbreaks). The scenarios provide (optional) pre and post simulation worksheets regarding general knowledge of disease dynamics, systems thinking skills, and data literacy. The worksheets are intended to solicit feedback from the participants on how Outbreak Simulator could be improved or expanded. We are not collecting educational outcome data. A downloadable build of Outbreak Simulator will also be available on our website, allowing participants to use their own computers to try the software.

Figure 3

Model Parameterization and scenario set up in Outbreak Simulator.



Potential Instructional and Research Applications

Ultimately, Outbreak Simulator will collect data on learning gains from the user by embedding pre and post scenario evaluation instruments into the software. We are interested in embedding content in the activities to fill knowledge gaps and exploring interactive mental model construction as a mechanism to detect user misconceptions. Our goal is to have Outbreak Simulator present data visualizations that are responsive to individual learner's needs in terms of timing, information content, and interface location by collecting, analyzing, and synthesizing pretest and post-test measures (i.e., mental models, learning outcomes, and self-report measures) and multimodal trace data (such as user interaction patterns, log files, eye movements, facial expressions of emotion, physiological arousal, and concurrent verbalizations; see (Azevedo et al., 2018; Dever et al., 2022).

References

- Azevedo, R., Taub, M., Mudrick, N. V., Martin, S. A., & Grafsgaard, J. (2018). Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. *Handbook of Self-Regulation of Learning and Performance, 2*.
- Chowell, G., Nishiura, H., & Bettencourt, L. M. A. (2007). Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface, 4*(12), 154–166. <https://doi.org/10.1098/rsif.2006.0161>
- Dever, D., Amon, M., Vrzáková, H., Wiedbusch, M., Cloude, E., & Azevedo, R. (2022). Capturing Sequences of Learners' Self-Regulatory Interactions With Instructional Material During Game-Based Learning Using Auto-Recurrence Quantification Analysis. *Frontiers in Psychology, 13*.
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics, 22*(4), 403–434. [https://doi.org/https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/https://doi.org/10.1016/0021-9991(76)90041-3)
- Lofgren, E., Fefferman, N. H., Naumov, Y. N., Gorski, J., & Naumova, E. N. (2007). Influenza Seasonality: Underlying Causes and Modeling Theories. *Journal of Virology, 81*(11), 5429–5436. <https://doi.org/10.1128/jvi.01680-06>
- Lofgren, E. T., Collins, K. M., Smith, T. C., & Cartwright, R. A. (2016). Equations of the End: Teaching Mathematical Modeling Using the Zombie Apocalypse. *Journal of Microbiology & Biology Education, 17*(1), 137–142. <https://doi.org/10.1128/jmbe.v17i1.1066>
- Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2021). A Deep Gravity model for mobility flows generation. *Nature Communications, 12*(1). <https://doi.org/10.1038/s41467-021-26752-4>
- Smith, R. (2009). Faux review: A report on the zombie outbreak of 2009: How mathematics can save us (no, really). *Cmaj, 181*(12), 297–300. <https://doi.org/10.1503/cmaj.091704>
- Weissman, G. E., Crane-Droesch, A., Chivers, C., Luong, T. B., Hanish, A., Levy, M. Z., Lubken, J., Becker, M., Draugelis, M. E., Anesi, G. L., Brennan, P. J., Christie, J. D., Hanson, C. W., Mikkelsen, M. E., & Halpern, S. D. (2020). Modeling Epidemics With Compartmental Models. *Annals of Internal Medicine, 173*(1), 21–28. <https://doi.org/10.7326/M20-1260>