1-2018

Mathematical Epidemiology Goes to College

Meredith L. Greer
Bates College, mgreer@bates.edu

Ella Livesay
Bates College

Recommended Citation
Greer, M., & Livesay, E. (2018). Mathematical epidemiology goes to college, Math Horizons, 25(3), 8-11. https://doi.org/10.1080/10724117.2018.1424457

Mathematical Epidemiology Goes To College

Every year waves of illnesses sweep through college campuses. This seems a natural result of sleep-deprived college students living, working, and playing together. Such outbreaks suggest questions: How many people will become infected? How can illnesses be contained? And crucially: How is mathematics involved?

Mathematical epidemiology is the study of modeling diseases, often using *compartmental models*. Read on to see how to build compartmental models so you can use them to learn from past outbreaks and investigate theoretical future outbreak scenarios. The models shown are inspired by two real-life outbreaks with different dynamics at the same small residential campus: H1N1 influenza in 2009, and a surprising outbreak of mumps in a highly vaccinated population in 2016.

Compartmental Models

Consider an illness in a population. An *SIR model* assumes that at each time t, each person in the population belongs to exactly one of three subgroups of the population, called *compartments*: the *Susceptible* compartment, people who have not contracted the illness but could get it, the *Infectious* compartment, people who have the illness and can spread it to Susceptibles, and the *Removed* compartment, people who are immune and are not spreading the illness. People can move from one compartment to another, so population sizes of compartments change over time. "Change" suggests a derivative, and indeed the change in each compartment's size is written as a differential equation. Figure 1 shows one such model as a diagram; below is the same model as a system of differential equations. In both, parameters β and γ affect outbreak dynamics. The next section describes their roles and the form of the equations.

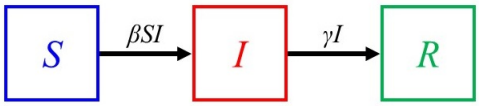

Figure 1: SIR compartmental model diagram.

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

Different outbreaks may require different models. The choices of compartments, number of arrows, and formulas associated with each arrow can all change, depending on the outbreak's

biology and the modeler's focus. Yet the connection between diagram and equations stays consistent: one differential equation per compartment, each arrow showing flow into one compartment and/or flow out of another compartment.

The model in Figure 1 has only outward flow from S, meaning $\frac{dS}{dt} \leq 0$; in words, population S cannot increase. Similarly, R has only inward flow and cannot decrease. However, I has both inflow βSI and outflow γI. Think about what this might mean. If $\frac{dI}{dt} > 0$ then population I is increasing, for example at the start of an outbreak. If $\frac{dI}{dt} < 0$ then $\gamma I > \beta SI$, meaning more recoveries than new infections: the outbreak may be nearing its end. Also possible is $\frac{dI}{dt} = 0$, meaning no net change in I. Ponder what $\frac{dI}{dt} = 0$ could signify; there are multiple options.

Some useful epidemiological vocabulary:

$$\text{Incidence} = \frac{\text{Number of new cases in one time unit}}{\text{Total population size during time unit}};$$

$$\text{Prevalence} = \frac{\text{Number of existing infectious cases at one time unit}}{\text{Total population size during time unit}}.$$

Data are often in terms of the incidence numerator: new cases per time unit, modeled as βSI. The prevalence numerator is $I(t)$, total current cases on a given day. Time t has units of days throughout this article.

The *basic reproduction number* $\mathcal{R}_0$ is the average number of new disease cases caused by a single Infectious person in an otherwise Susceptible population. When $\mathcal{R}_0 > 1$, the disease initially spreads. When $\mathcal{R}_0 < 1$, the disease dies out. Epidemiologists gather data to estimate $\mathcal{R}_0$ for diseases like influenza and mumps [7] and $\mathcal{R}_0$ plays many roles in modeling.

With these ideas in mind, we use the SIR model for a common campus outbreak: the flu.

Influenza

For a basic flu model, Figure 1 is a good choice for the following reasons. The novel virus strain in 2009 meant every student was initially considered Susceptible (except the first Infectious student). Once a student contracted the flu, the time till they could infect others was quite short [3]. We therefore leave this time period out of our model, calling these students Susceptible until they start infecting others. (Another option is introducing a new compartment for these students. This idea resurfaces in the mumps section.) Once no longer contagious, students go to the Removed compartment and stay there: they have gained immunity to that flu strain, hence cannot return to the Susceptible compartment [3].

Each differential equation term has flu-related biological meaning. The γI term governs movement from the I compartment to the R compartment. To estimate γ, consider the infectious period. For H1N1, this could be as much as 5 to 7 days, though the period of

highest infectivity lasts just 2-3 days. Supposing a 3-day infectious period, then in a typical day about $\frac{1}{3}$ of the members of I move to R, which indicates $\gamma = \frac{1}{3}$. To be biologically reasonable, any modeled gamma should correspond to real-life values.

The model's βSI term describes interaction between Susceptible and Infectious people. To think through the SI part, note that the student population was a fixed 1714 throughout the flu outbreak. When either S or I is very small—1 or 2 students—SI is relatively small. As flu spreads, a still-sizable S population, multiplied by an I population in the dozens (or more), is much larger. Later in the outbreak, S has decreased, and some people have moved to R, making SI again smaller.

Mathematical software (e.g. *Mathematica*) turns differential equations, initial conditions, and parameter values into graphs. Figure 2 shows a modeled Infectious population (the smooth curve) compared with real prevalence values computed from campus health center incidence data by supposing students were Infectious for three days. The model sets $\gamma = 1/3$, $\beta = 0.000305$, $S(0) = 1713$, $I(0) = 1$, and $R(0) = 0$.

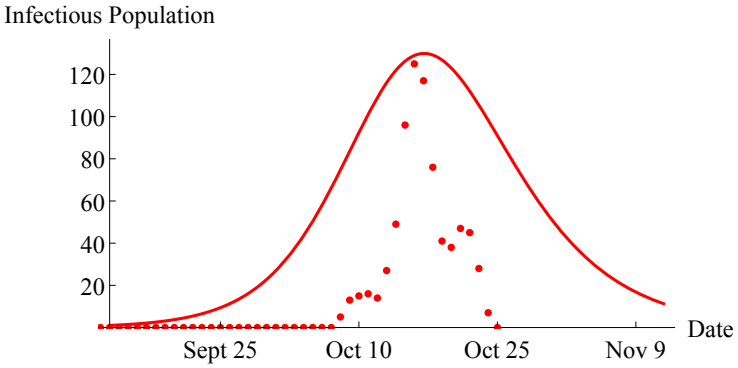

Figure 2: Influenza model compared with 3-day prevalence data.

Natural questions follow. How do you use these models? And how do they connect with real data?

One answer is to use models for mathematical understanding. For instance: How does increasing or decreasing γ or β change your model's graphs and underlying biological assumptions? How are non-integer-valued populations interpreted? What happens if parameter values are not constant? Try these yourself!

We can also ask questions with models, such as: How does student failure to report illness change the model? Faculty and Health Center employees in 2009 heard of many unreported cases; a model assuming double the number of reported cases ($\gamma = 1/3$, $\beta = 0.00038$) appears in Figure 3.

Models can also be fit to data, keeping parameters within biologically feasible ranges. One approach is *Residual Sum of Squares* (RSS): given data points $y_1, y_2, \ldots, y_n$ and modeled values $I(1), I(2), \ldots, I(n)$ at corresponding times $1, 2, \ldots, n$,

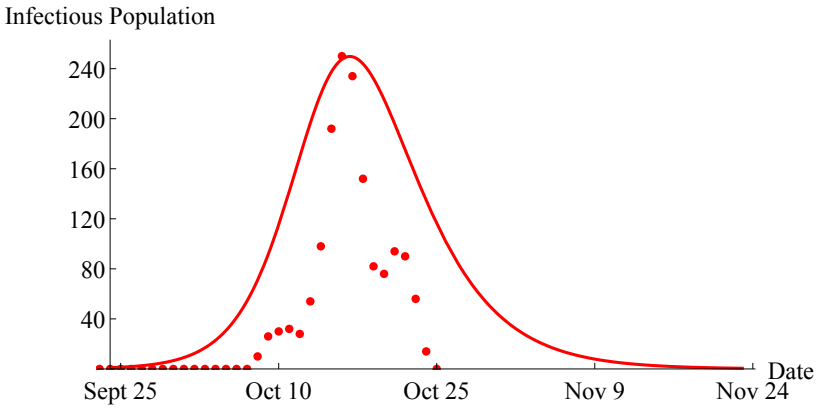

Figure 3: Influenza model compared with double the 3-day prevalence data, to account for unreported cases.

$$\text{RSS} = \sum_{i=1}^{n} (y_i - I(i))^2.$$

A model with smaller RSS fits the data more closely than a model with larger RSS.

Once a model is fit to data, it can help estimate $\mathcal{R}_0$. The *Next Generation Method* uses partial derivatives and linear algebra to determine a formula for $\mathcal{R}_0$ [2]. The basic SIR model has formula $\mathcal{R}_0 = \frac{\beta S(0)}{\gamma}$, where $S(0)$ is the initial Susceptible population. The curve in Figure 2 has $\mathcal{R}_0 \approx 1.57$. Because $\mathcal{R}_0 > 1$, disease spreads, but $\mathcal{R}_0$ is close enough to one that changes in human behavior make a difference. Hand washing and social distancing reduce β, bringing $\mathcal{R}_0$ closer to 1 and reducing new flu cases significantly. (For contrast: $\mathcal{R}_0$ for pre-vaccination measles outbreaks was 12 or greater! Hand washing had much less effect on those outbreaks.)

Above, β helps estimate $\mathcal{R}_0$. Given insufficient data to compute β, we can instead use epidemiologists' estimated $\mathcal{R}_0$ values to estimate β. This approach makes sense when modeling an ongoing outbreak, which happened when the mumps appeared in Fall 2016.

The Mumps

While the flu appears often, sometimes a rarer illness emerges. In one year, just a few hundred to a few thousand cases of mumps occur in the United States... and they cluster at places like residential college campuses [4] despite nearly 100% vaccination. The campus focus heightens our interest in mumps, and the trajectory of mumps illness contrasts notably with that of influenza.

When a Susceptible contracts mumps, there is a long time lag till they become Infectious [5, 6], prompting a new compartment: L, or *Latent*. (Note that epidemiologists' definition of

"Latent" differs somewhat.) The time from infection till symptom appearance is typically 16-18 days. Symptoms last about 5 days. Infectiousness—ability to spread mumps to others—begins about 2 days before symptoms appear and ends about when symptoms end.

Figure 4 shows one possible diagram.

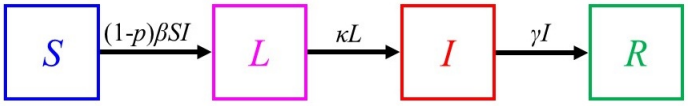

Figure 4: SLIR compartmental model diagram.

Most parameters are familiar from the SIR model. New is the factor $(1-p)$ multiplied by βSI, where p is vaccination effectiveness. If everyone receives two doses, then mumps vaccine averages 88% effectiveness across a population [4]; this corresponds to $p = .88$.

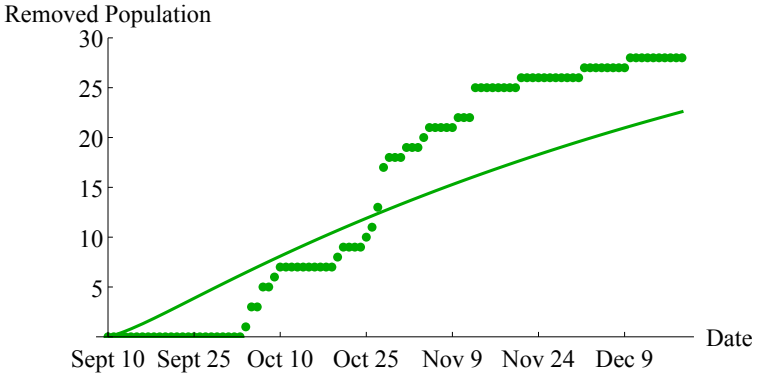

Figure 5: Mumps model compared with data.

Figure 5 shows outbreak data and the modeled Removed curve. Parameters match given biological data ($\kappa = 1/15$, $\gamma = 1/7$, $p = .88$); initial conditions are $S(0) = 1793$, $L(0) = 6$, $I(0) = 1$, $R(0) = 0$. The $\mathcal{R}_0$ equation for SIR fits the SLIR model too; rearranging indicates $\beta = \frac{\gamma \mathcal{R}_0}{S(0)} \approx 0.00056$, using $\mathcal{R}_0 = 7$. (Epidemiologists report $\mathcal{R}_0 \approx 4$–7 for mumps [7]. Given many campus risk factors [4], we use the largest value.) With these parameters and more time, the Removed curve reaches 40 students. In reality, 28 students reported sickness. The difference between 40 modeled and 28 reported cases may be due to the holiday break halting the outbreak, along with incomplete reporting of mumps by students.

A thought experiment: what if mumps came to a completely Susceptible campus? Setting $p = 0$ in the model from Figure 5 leads to mumps infecting nearly the entire student body by late December! Pre-vaccination reality was different: most children had mumps and thus were immune before college. Still, modeling lets us try such scenarios.

At this point, you may be wondering how we know which model to use for which disease. The answer is: there is no one answer. Nothing stops us from modeling the same outbreak different ways, so let's try it. We first note that a campus may isolate symptomatic

students so they cannot infect others. Then the *Infectious* compartment (I) consists of not-yet-symptomatic students mixing freely and spreading mumps, followed by an *Isolated* compartment (J) for symptomatic students whose isolation means they cannot spread mumps. Based on above data, students remain in I approximately 2 days and J approximately 5 days. Second, as many as 20% of people infected with mumps never display symptoms [8], suggesting an *Asymptomatic* (A) compartment.

The resulting model, SLIJAR, appears in Figure 6. Notice parameter q. When $q = 0.2$, 20% of students leaving L move to A, and the other 80% move to I. The sum of students leaving compartment L, $q\kappa L + (1-q)\kappa L$, equals κL, with κ computed in a similar way in SLIJAR as in SLIR.

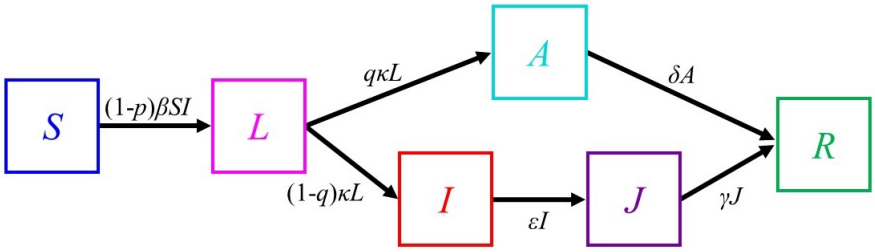

Figure 6: SLIJAR compartmental model diagram.

Models with more details can be helpful for understanding new aspects of disease spread. For example, whereas $\mathcal{R}_0 = \beta S_0/\gamma$ for the SLIR model showed the importance of length of time spent in compartment I, $\mathcal{R}_0 = \beta S_0 \left(\frac{q}{\delta} + \frac{1-q}{\epsilon}\right)$ [2] for SLIJAR shows that time spent in both compartments A and I matters, with weights q and $1-q$ respectively.

Different models raise questions about fitting data. More parameters often make a closer fit possible—but is this necessarily better? There may not be adequate information for estimating more parameters or ensuring they are biologically reasonable. Additionally, the centuries-old principle Occam's Razor encourages us to use the simplest appropriate model. To balance closer data fitting (usually with more parameters) with the goals of simplicity and of keeping models biologically meaningful, modelers use RSS to compute the *corrected Akaike Information Criterion* (AIC_C):

$$\text{AIC}_\text{C} = n \ln\left(\frac{\text{RSS}}{n}\right) + \frac{2Kn}{n-K-1}$$

where n is the number of data points, and K is 1 plus the number of model parameters. For several AIC_C examples, see [1]. The model with lowest AIC_C is considered to have the best combination of RSS and number of parameters.

Proper use of AIC_C requires comparing the same data to different models. For the SLIR and SLIJAR models, our data set is all known currently infectious people, compared to the I population of SLIR and $I + J$ in SLIJAR. With SLIR parameter values from Figure 5,

comparable values for SLIJAR ($\kappa = 1/15$, $q = .2$, $\epsilon = 1/2$, $\gamma = 1/5$, $\delta = 1/7$, $p = .88$), and computing β for each model using $\mathcal{R}_0 = 7$, the simpler model, SLIR, has both lower RSS and lower AIC_C.

Conclusion

We hope you have enjoyed this introduction to mathematical epidemiology. These models allow for mathematically representing human interactions and simulating outbreak scenarios. Consider making β piecewise constant (perhaps dropping when a campus reacts to an outbreak) or periodic (maybe showing increased student interaction, hence infection, on weekends). Try new diseases, new compartmental models, and new data sets. Model past diseases, and try to predict the outcomes of new outbreaks. And enjoy the close connection between mathematics and current campus events!

References

[1] O. Akman, M.R. Corby, E. Schaefer, Examination of Models for Cholera: Insights into Model Comparison Methods, *Letters in Biomathematics* **3** (2016) 93–118.

[2] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*. Second edition. Springer, New York, 2012.

[3] Centers for Disease Control and Prevention, H1N1 Flu (Interim Guidance on Infection Control Measures for 2009 H1N1 Influenza in Healthcare Settings, Including Protection of Healthcare Personnel), https://www.cdc.gov/h1n1flu/guidelines_infection_control.htm, Accessed 2 June 2017.

[4] Centers for Disease Control and Prevention, Mumps Cases and Outbreaks, https://www.cdc.gov/mumps/outbreaks.html, Accessed 17 May 2017.

[5] Centers for Disease Control and Prevention, Signs & Symptoms of Mumps, https://www.cdc.gov/mumps/about/signs-symptoms.html, Accessed 6 June 2017.

[6] Centers for Disease Control and Prevention, Transmission of Mumps, https://www.cdc.gov/mumps/about/transmission.html, Accessed 6 June 2017.

[7] L. Edelstein-Keshet, *Mathematical Models in Biology*. Random House, New York, 1988.

[8] J.M. Conly, B.L. Johnston, Is Mumps Making A Comeback?, *Canadian Journal of Infectious Diseases and Medical Microbiology* **18** (2007) 7–9.