Mathematical Epidemiology Goes to College

Meredith Greer and Ella Livesay

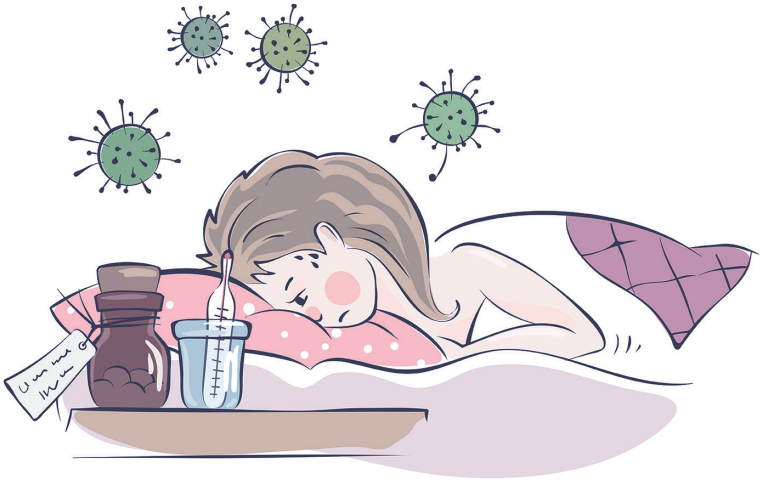
Every year waves of illnesses sweep through college campuses. This seems a natural result of sleep-deprived college students living, working, and playing together. Such outbreaks suggest questions: How many people will become infected? How can illnesses be contained? And crucially: How is mathematics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

Mathematical epidemiology is the study of modeling diseases, often using *compartmental models*. We can use such models to learn from past outbreaks and investigate theoretical future outbreaks.

In this article we present models that were inspired by two real-life outbreaks at a small residential college campus: H1N1 influenza in 2009 and, surprisingly, mumps in 2016.

Compartmental Models

Consider an illness in a population. Suppose that at time t, each person belongs to one of three subsets, or *compartments*, of the population: 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 susceptible individuals; and the *removed* compartment—people who are immune and not spreading the illness. The sizes of these compartments are given by $S(t)$, $I(t)$, and $R(t)$, respectively (time is measured in days throughout this article); hence, this is known as an *SIR model*.



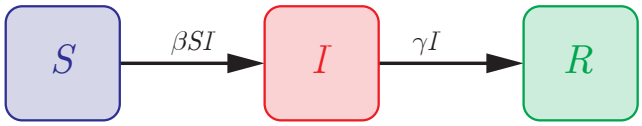


Figure 1. SIR compartmental model diagram.

Individuals can move from one compartment to another, so the sizes of the compartments change over time. "Change" suggests a derivative, and indeed the change in each compartment's size is written as a differential equation. We can express the SIR model as the system of differential equations

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

or as the model diagram in figure 1.

We will discuss the details of this model in the next section, but for now, consider these general observations. The model has only outward flow from S, meaning $\frac{dS}{dt} \le 0$; so the susceptible population cannot increase. Similarly, R has only inward flow and cannot decrease. However, I has both inflow βSI and outflow γI. If $\frac{dI}{dt} > 0$, then the infected population is increasing, such as at the start of an outbreak. If $\frac{dI}{dt} < 0$, then $\gamma I > \beta SI$, meaning there are more recoveries than new infections; the outbreak may be nearing its end. Think about what it could mean when $\frac{dI}{dt} = 0$; that is, when there is no net change in I.

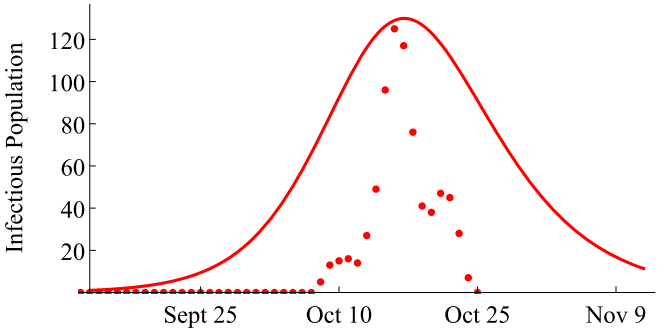


Figure 2. Influenza model compared with three-day prevalence data.

Different outbreaks may require different models. The choices of compartments, numbers 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.

Here is some useful epidemiological vocabulary: We define *incidence* as

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

and *prevalence* as

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

Data is often presented in terms of the incidence numerator, which is modeled as βSI. Notice that the prevalence numerator is $I(t)$.

The *basic reproduction number* $\mathcal{R}_0$ plays many roles in modeling. It 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, and when $\mathcal{R}_0 < 1$, the disease dies out. Epidemiologists gather data to estimate $\mathcal{R}_0$ for diseases like influenza and mumps (L. Edelstein-Keshet, *Mathematical Models in Biology*, Random House, 1988).

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

Influenza

The SIR model in figure 1 is a good choice for a basic flu model. 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 that person could infect others was quite short, so we leave this time period out of our model—students move directly from susceptible to infectious. (In the mumps section, we introduce a compartment for infected, noninfectious students.) Students who are no longer contagious have gained immunity to that flu strain, so they go to the removed compartment and stay there (see the Centers for Disease Control and Prevention's website on the H1N1 flu, *http://bit.ly/CDCH1N1Flu*).

Each differential equation term has a flu-related biological interpretation. The γI term governs movement from the I compartment to the R compartment. For H1N1, the infectious period could be as many as five to seven days, though the period of highest infectivity lasts just two to three days. If we assume a three-day infectious period, then in a typical day, about $\frac{1}{3}$ of the members of I move to R, which indicates $\gamma = \frac{1}{3}$.

The model's βSI term describes interaction between susceptible and infectious individuals. The student population was a fixed 1,714 throughout the flu outbreak. When I is very small—one or two students, say—βSI is relatively small. As the 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 βSI again smaller.

Figure 2 shows a modeled infectious population (the smooth curve) compared with real prevalence values computed from campus health center incidence data by assuming students were infectious for three days. The model sets $\gamma = \frac{1}{3}$, $\beta = 0.000305$, $S(0) = 1{,}713$, $I(0) = 1$, and $R(0) = 0$.

In real life, not every flu sufferer seeks treatment. Indeed, anecdotal evidence from faculty and health center employees indicates that there were many

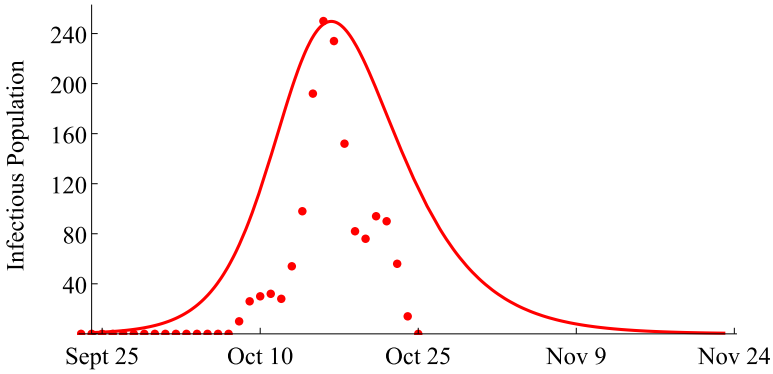


Figure 3. Influenza model compared with double the three-day prevalence data, to account for unreported cases.

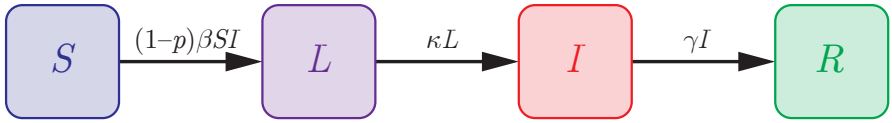


Figure 4. SLIR compartmental model diagram.

unreported cases of the flu. Figure 3 shows a model that assumes double the number of reported cases ($\gamma = \frac{1}{3}$ and $\beta = 0.00038$).

Models can also be fit to data. The *residual sum of squares* (RSS) is one way to measure how close the fit is: Given data points $y_1, y_2, \ldots, y_n$ and modeled values $I(1), I(2), \ldots, I(n)$ at corresponding times $1, 2, \ldots, n$,

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

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 be used to estimate $\mathcal{R}_0$. In the SIR model, $\mathcal{R}_0 = \beta S(0) / \gamma$, where $S(0)$ is the initial susceptible population (see F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, 2nd ed., Springer, 2012). The model 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 1 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. (In contrast, in the era before vaccinations, $\mathcal{R}_0$ was 12 or greater for measles outbreaks—hand washing had much less effect!)

We used β to estimate $\mathcal{R}_0$, but given insufficient data to compute β, we can use epidemiologists' $\mathcal{R}_0$ values to estimate β. This approach makes sense when modeling an ongoing outbreak, which happened when mumps appeared in fall 2016.

Mumps

Only a few hundred to a few thousand cases of mumps occur in the United States each year. When the cases do appear, they cluster at places like residential college campuses, even when nearly 100 percent of students have been vaccinated.

The trajectory of mumps illness contrasts notably with that of influenza. (Information in this section was obtained from the CDC websites *http://bit.ly/CDCMumpsCases*, *http://bit.ly/CDCMumpsSigns*, and *http://bit.ly/CDCMumpsTransmission*.)

When a susceptible individual contracts mumps, there is a long time lag till he or she becomes infectious. Thus, we introduce a new compartment: L, or *latent*. (Note that epidemiologists' definition of latent differs somewhat.) The time from infection till symptom appearance is typically 16 to 18 days. Symptoms last about five days. Infectiousness begins about two days before symptoms appear and ends about when symptoms end. Figure 4 shows one possible diagram.

Most parameters are familiar from the SIR model. New is the factor $1 - p$ multiplied by βSI; here, p is vaccination effectiveness. If everyone receives two doses of the mumps vaccine, then it averages 88 percent effectiveness ($p = 0.88$) across the population.

Figure 5 shows outbreak data and the curve for the modeled R population. Parameters match given biological data ($\kappa = \frac{1}{15}$, $\gamma = \frac{1}{7}$, and $p = 0.88$); initial conditions are $S(0) = 1{,}793$, $L(0) = 6$, $I(0) = 1$, and $R(0) = 0$. The $\mathcal{R}_0$ equation for SIR fits the SLIR model too, yielding $\beta = \gamma \mathcal{R}_0 / S(0) \approx 0.00056$, using $\mathcal{R}_0 = 7$. (Epidemiologists report that for mumps, $\mathcal{R}_0$ is approximately 4 to 7 [see Edelstein-Keshet]. Given many campus risk factors, we use the largest value.) With these parameters and more time, the curve for R reaches 40 students. In reality, 28 students reported sickness. The difference between 40 modeled and 28 reported cases may be because the holiday break halted the outbreak, along with incomplete reporting of mumps by students.

Such models enable us to explore various scenarios. For instance, what if mumps came to a completely susceptible campus? Setting $p = 0$ in the model leads to mumps infecting nearly the entire student body by late December! Note, however, that before vaccination, most people contracted mumps as children, and thus they were immune before college.

How do we know which model to use for which disease? There is no one correct answer. We can model the same outbreak in different ways. Let's do that.

A campus may isolate symptomatic students so they cannot infect others, so let's include that in our model. The *infectious* compartment (I) consists of not-yet-symptomatic students who can mix freely

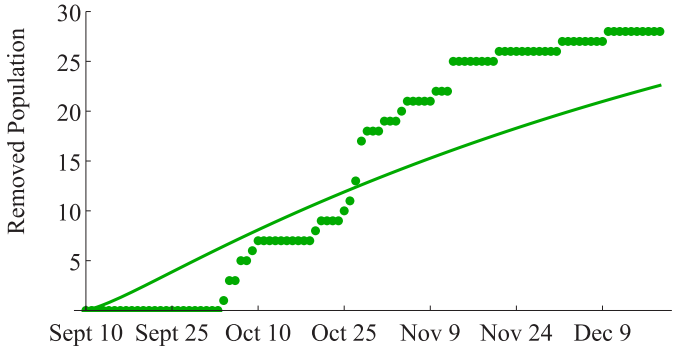


Figure 5. Mumps model compared with data.

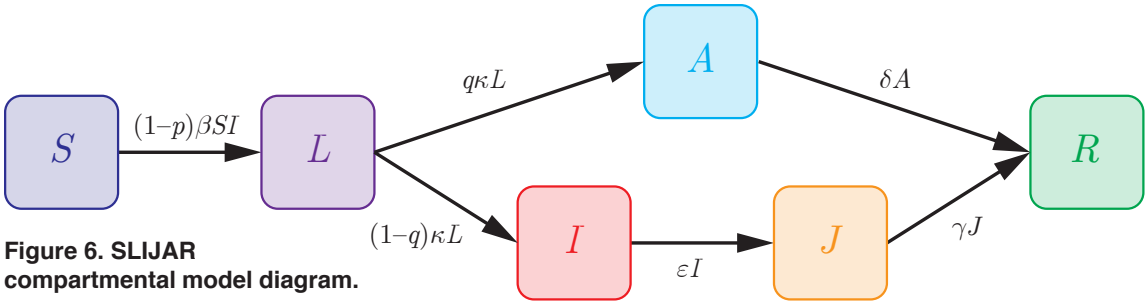


Figure 6. SLIJAR compartmental model diagram.

and spread mumps. The *isolated* compartment (J) contains symptomatic students whose isolation prevents them from spreading mumps. Based on above data, students remain in I approximately two days and in J approximately five days.

Moreover, as many as 20 percent of people infected with mumps never display symptoms (J. M. Conly, B. L. Johnston, "Is Mumps Making a Comeback?," *Can. J. Infect. Dis. Med. Microbiol.* 18 [2007]: 7–9). This suggests an *asymptomatic* (A) compartment.

Figure 6 shows the resulting SLIJAR model. Notice parameter q. When $q = 0.2$, 20 percent of students leaving L move to A, and the other 80 percent 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.

Models with more details can be helpful for understanding new aspects of disease spread. For example, whereas the formula $\mathcal{R}_0 = \beta S_0 / \gamma$ for the SLIR model shows the importance of length of time spent in compartment I,

$$\mathcal{R}_0 = \beta S_0 \left(\frac{q}{\delta} + \frac{1-q}{\epsilon} \right)$$

(see Brauer and Castillo-Chavez) for SLIJAR shows that time spent in both compartments A and I matters, with weights q and $1-q$ respectively.

Having more parameters in a model often makes 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, 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 one more than the number of model parameters. (For several AIC_c examples, see O. Akman, M. R. Corby, E. Schaefer, "Examination of Models for Cholera: Insights into Model Comparison Methods," *Lett. Biomath.* 3 [2016]: 93–118.) 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 contains people who are infectious and symptomatic, which we compare with the I population of SLIR and the $I + J$ population in SLIJAR. With SLIR parameter values from figure 5, comparable values for SLIJAR ($\kappa = \frac{1}{15}$, $q = 0.2$, $\epsilon = \frac{1}{2}$, $\gamma = \frac{1}{5}$, $\delta = \frac{1}{7}$, and $p = 0.88$), and computing β for each model using $\mathcal{R}_0 = 7$, the simpler model, SLIR, has both lower RSS and lower AIC_c.

Further Explorations

These compartmental models let us use mathematics to represent human interactions and simulate outbreak scenarios. Explore these models on your own. Consider making β piecewise constant (dropping when a campus reacts to an outbreak, say) or periodic (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! ■

This partnership formed at Bates College when Ella Livesay took Meredith Greer's course on mathematical epidemiology. It continued through the writing of this article. And since Ella stayed in Maine post-graduation, as a consultant in health analytics, we are fortunate enough to celebrate in person that the article is appearing in Math Horizons. *The authors thank Chip Ross and Karen Palin for close consultation and edits on drafts of this article.*

http://dx.doi.org/10.4169/mathhorizons.25.3.8