SIMULATING VIRUS TRANSMISSION: A STEM INVESTIGATION

Nicolle Dexter, Holmen High School, Holmen, Wisconsin

I am an advocate for true STEM activities: activities that genuinely combine two or more disciplines at grade level. As a dual-licensed physics and math educator, I often find "STEM" activities that were created with good intentions, but they are not meaningful; one field is often de-emphasized to spotlight the goals in another. STEM activities, as real work in the fields often does, should blur the lines between content to an extent that they are indistinguishable from one another. When Covid-19 first started to spread and schools were transitioning online, I knew this would be the perfect time to start incorporating some virology into my math curriculum; students were excited to learn about it and society as a whole needed a refresher on the relevant science. In late June through early July 2020, I taught a precalculus class for my local Upward Bound program and I centered all the content around viruses and how they spread, specifically focusing on COVID-19 and similar pandemics. We analyzed reported and projected data, explored herd immunity (how, scientifically and mathematically, we think about protecting society through immunity), discussed the different numbers that model the biologically and environmentally determined ability of a disease to spread (effective and basic reproduction values), and simulated the spread of the virus under specific conditions. Our time frame was eleven 1-hour sessions over two weeks, so some of the true exploration in considering multiple biological variables was cut short -- but I have considered many ways to add to the content and be truer to the biological themes in the future.

I do not have a degree in biology, so I did a lot of research while creating this unit. There are two values that are often reported to describe an infectious disease: the basic and effective reproduction numbers. The basic reproduction number (R_0) describes the average number of individuals who catch the disease from each infected person over the course of their illness assuming everyone is susceptible to infection (no immunity). The effective reproduction number, in comparison, is a calculation of the average number of transmissions per person that a community in which some are immune might experience (immune response post recovery, vaccinations, etc.). While researching the effects of the basic and effective reproductive numbers on herd immunity, I came across this Bozeman Science video that illustrated how a virus's spread through a community can both create and be hindered by herd immunity. I found this visual for herd immunity and the ability for a virus to move amongst a community really effective; so, I used the same concept to create a simulation for my class.

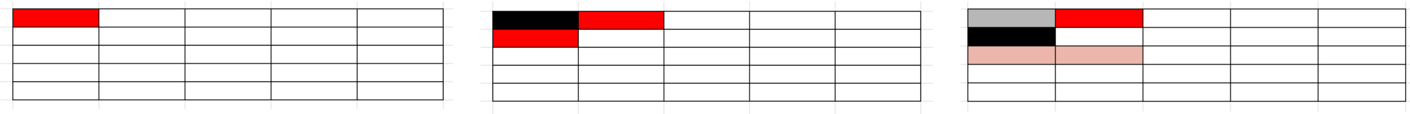
In the initial simulation, students begin in small groups with a grid of 25 "subjects." Subjects do not move and can only spread the disease to others who are not immune and are directly adjacent (maximum of eight). The initial infected subject, for our simulation, was in the upper left corner. While you could begin with any person infected (I discuss this as a modification for future simulations later), the cognitive requirement of the task, especially when given in a virtual environment, drove my decision to begin in a corner and slowly escalate the requirements for each round of the simulation. Students then work their way through the grid and collect data over rounds or "generations." One round consists of testing susceptible subjects adjacent to any diseased subject at the beginning of the round for transmission of the virus based on the predetermined probability (we used 50%- a fair probability that is relatively simple to simulate). After this test, any subject infectious at the beginning of the round becomes immune (simulates transmission over the course of a person's illness and assumes some level of immune response after recovery). Some subjects, if adjacent to multiple diseased subjects, may need to be tested more than once in a round though can obviously only be infected once. Throughout the simulation, students collect data on the overall number of folks infected (or recovered) after each round, new infections that occur per round, and the number of new infections caused by each infected individual.

There was a great deal of information to keep track of as the simulation progressed. In an attempt to scaffold this high level of cognitive demand, I created spreadsheets in Google Sheets for each group (that I could easily access as I moved from group to group) and specific roles for each student in their small group. I outlined the roles as well as some helpful tips in the following table (Table 1) that was given to students in their instructions.

Card Keeper	Probability Master	Table Keeper 1	Table Keeper 2
Your job is to keep track of the grid. **Guide to Success** You should use the following color-coding scheme: Fill Red: Infected Fill White: Not Infected Fill Grey: Immune When you are focusing on an infected cell, color it black until you're done testing adjacent cells. Then, color it grey as immune.	Your job is to handle determining whether or not an infected cell infects an adjacent cell. There is a 50% chance of spreading the disease. **Guide to Success** You can use a coin (Heads-Infected, Tails-Not infected) You can use a six-sided die, (Even-Infected, Odd- Not Infected) You can use the random number generator (1-10); (Even-infected, Odd-Not infected)	Your job is to keep track of two things: How many NEW infections occur each round. How many total infections occur after each round. (This includes people who have recovered.) **Guide to Success** Use the table in the spreadsheet to help you organize your data. Notice that if you add up new infections from the entire activity,, you should always have the total for each round.	Your job is to keep track of how many new infections occur **per infected person.** **Guide to Success** Use the table in the spreadsheet to help you organize your data. The table keeper will highlight a current cell black when it's being run through the simulation. You should keep track of how many new boxes get infected from each black box.

I created these roles based on supplies I knew my students had access to. Every student was assigned an iPad for the summer and was supplied with routers if they did not have access to wi-fi (we met synchronously over Zoom). I was also intentional in requiring only one student to have access to materials for the probability tests -- and included multiple options (including one solely online) in case none of the students in a group had access to a coin or die. As seen in the table, I gave very specific suggestions in the "Guides to Success" so that there was some consistency as I transitioned between breakout rooms. This consistency also allowed students to communicate more easily across groups. It is important to note that doing a binary random number generator often makes outcomes confusing (getting the same output twice in a row can leave you wondering if you are seeing a new output or if you missed the button.)

After determining roles, each group was assigned a Google Sheet on which to collect data and run their simulations. These spreadsheets included a grid in which each cell was considered a different person. The card keeper would highlight which cell they were working on and used colors to indicate different statuses for each cell. An example of the color-coding system for a few rounds is demonstrated in Figure 1.



This was the initial grid presented to students, indicating that the person in the upper left was infected.

In the first round, the card keeper would highlight which infected person was being simulated at the time (by coloring it black). Then, the card keeper would indicate each adjacent cell one at a time. The probability master would then flip for each adjacent person and communicate with the group and the card keeper colors the cells red if necessary.

In the second round, the first person tested has become immune (grey). The card keeper has highlighted which of the two infected people at the start of the round is being tested first. They are also using a different color red for new infections to distinguish them from the infected persons from the current round. After testing those adjacent to the black cell, they will turn the cell grey and focus on the other infected individual at the start of the round before ending the round and moving on to round 3.

Figure 1: An Effective Color-Coding System

As the card keeper and probability master were working through the table, the two data keepers were collecting data in the same spreadsheet. Completed grids looked different depending on how the simulation played out in each group.

After running the simulation, students plotted their infection data (Refer to Figure 3) in Desmos and averaged the number of infections per individual. I encouraged students to try a few regressions to see which fit their data best. All three groups (class size of 12) decided on a logistic regression. Discussions centered around the logistic curve and why it's a good fit for infection spread and whether or not the immune response in infected individuals meant our averages were better representations of the basic or effective reproduction value. An example of collected data and regressions can be found below in Figures 2 and 3.

Generation	Total Infections	New Cases
0	1	-
1	3	2
2	6	3
3	8	2
4	14	6
5	17	3
6	22	5
7	23	1
8	23	0
9	23	0
10	23	0

	New Infections per Infected Person
	2
	2
	1
	2
	2
	0
	3
	3
	0
	0
	0
	2
	0
	1
	1
	2
	2
	0
	0
	1
	0
	0
Average >	1.090909091

Figure 2: Raw Data Working in the same Google Sheet allowed students to help one another and hold each other accountable throughout the simulation.

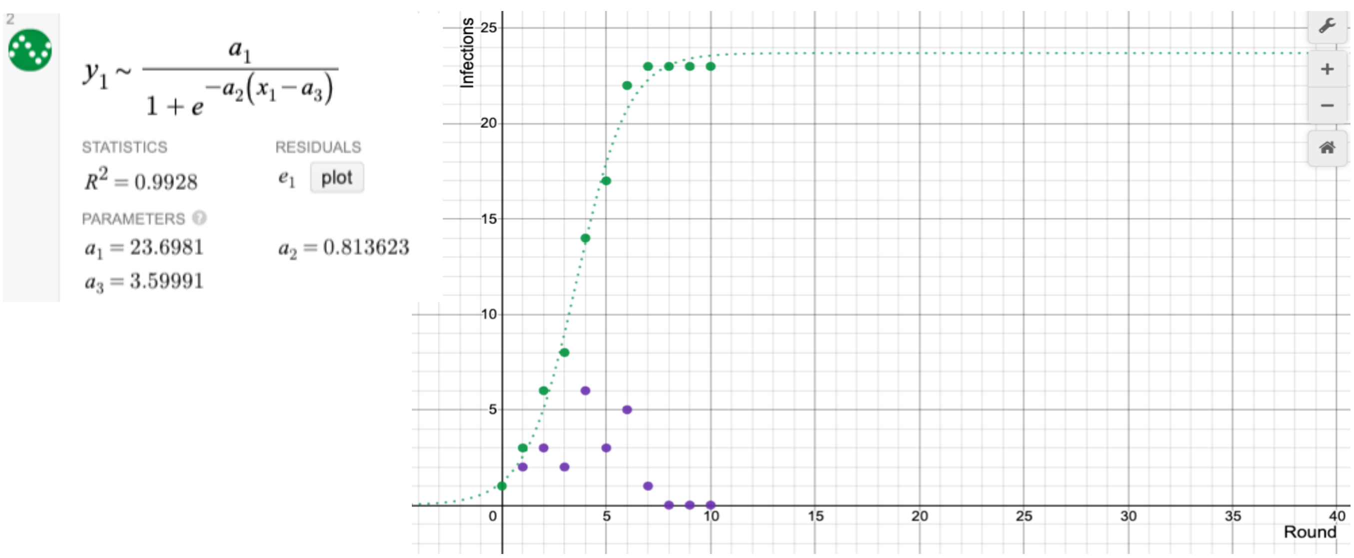
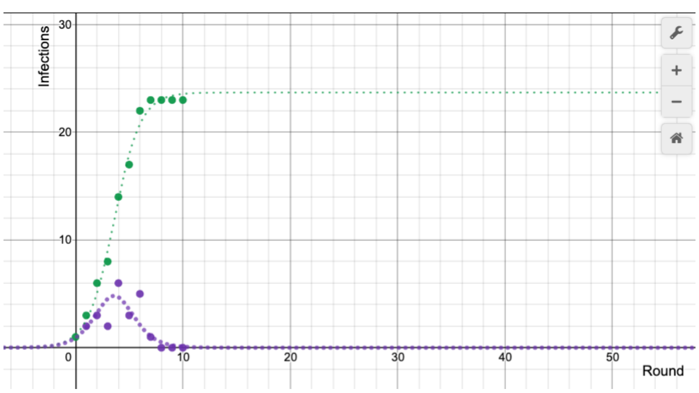
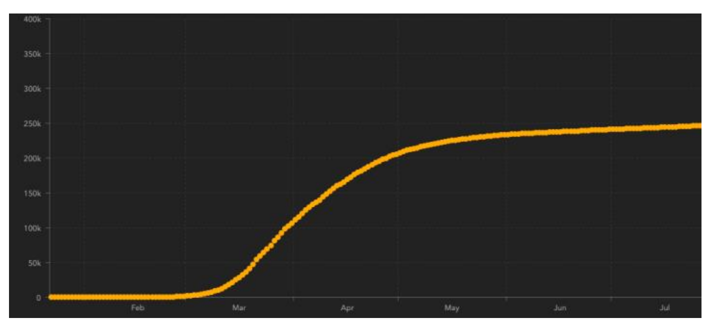
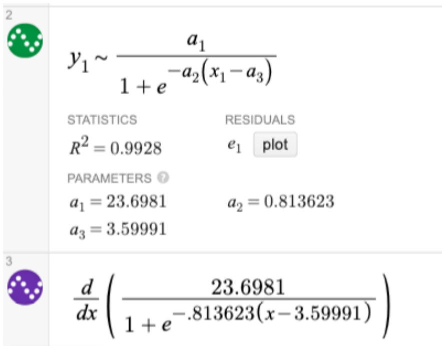


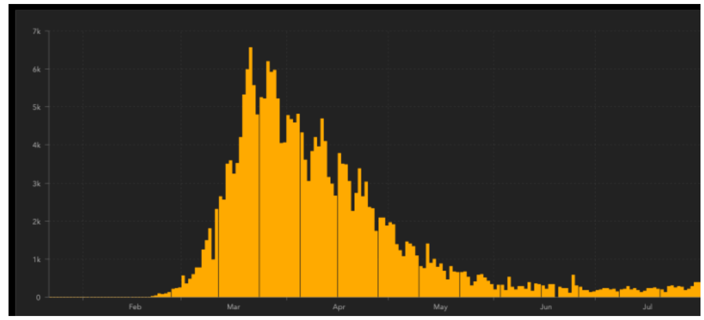
Figure 3: Regression My students were more familiar with Desmos, so we moved our data there. If students are familiar with testing and creating regressions in Google Sheets, it would also be a good option. Note that the green data points represent total infections while the purple data points represent new infections.

After plotting the data and finding an appropriate regression, students analyzed the model. Conversations in creating the regression included transformations and why we need to add constants into the logistic parent function in order to create a well-fitting regression. Some insightful observations included the vertical stretch (a_1 in the regression in Figure 3) being close to both the total observed and the total possible infections and the horizontal translation (a_3 in the regression in Figure 3) providing the same fit when used as a multiplier on e or included in the exponent-- an unexpected refresher on rules of exponents. Other observations came from the shape of the data itself. Students had been analyzing worldwide Covid data for about a week prior (this was in early July) and they were quick to recognize the logistic shape of infection in data from countries that had successfully slowed the spread of the virus. As students compared these countries to the United States, they noticed that the US data looked as though it were growing exponentially, not logistically, leading to a great conversation on the difference between exponential and logistic growth. Students noticed that the peak in new infections seemed to occur at the same time as half the total infections occurred (in both their simulation and worldwide data from Johns Hopkins) and indicated a slowing in the spread. After this observation, students seemed to understand that the US data looked different as we had not yet reached our peak in new cases -- a humbling discovery. Our conversation centered around how the new cases represented the change in growth in the overall number of cases, which would be an interesting extension for students at a calculus level. Calculus students could investigate how the new cases show an approximation for the derivative of the total cases and use that understanding to better compare the regressions' behaviors and recognize the point of inflection. These shapes are very well represented (as, unfortunately, there is a lot of data being reported around the world on new cases) in past coronavirus data.

Student Simulation (Below)	Johns Hopkins Dashboard (Below)
 Student simulated data for total cases (green) and new cases (purple)	 Italy-Total Confirmed Cases



Student Regressions (calculus students can further explore the relationship with derivatives.)



Italy-New Cases Reported

Figure 4: Comparison with Real Data Students compared the shapes of their data with real world data. On the left, student simulated data for total (green) and new cases (purple). On the right, data for total confirmed and new cases for Italy as reported by Johns Hopkins. We often compared the shapes of our data to Italy or China because they, at the time, had seen a decreased spread of Covid-19.

There are some fascinating biological and mathematical investigations students can embark on with respect to this activity. When asking students to brainstorm factors that might contribute to how a virus spreads, students came up with: the number of interactions infected people have, the probability of spreading the disease to adjacent people, whether or not people become immune after recovery, the intentional efforts to slow the spread, and more. In the future, I plan to have all students complete the simulation as stated here in the activity. Then, after becoming familiar with the process and running the simulation once, each group will be assigned a different modification for comparison to the original simulation. I plan to let students decide these modifications, but some examples could be:

- Have a few of the boxes be immune before running the simulation to simulate a vaccination process.
- Demonstrate different reproduction numbers by increasing or decreasing the probability of infection.
- Begin with a different cell infected to determine if the initial number of interactions (max in the first simulation was only 3 compared to a possible 8) affects the overall shape of the virus spread.
- Have multiple cells infected at the start of the simulation.
- Have some subjects scattered throughout the grid be less likely to be infected due to personal safety measures (like quarantining, social distancing, and/or wearing a mask.)
- Remove the immune response after recovery to simulate the possibility of being reinfected. You could also modify this by decreasing the probability of reinfection compared to initial infection.
- Have a separate probability for virus mutation on each roll (use a d20 or random number generator); if the mutation occurs everyone previously considered immune is once again susceptible.

Groups can present their findings (comparisons to the original simulation) and recommendations to their peers through various media, which creates an opportunity for new technology and some strong mathematical discourse. A claims-evidence-reasoning (CER) presentation would be particularly effective. Furthermore, biological and environmental factors play a huge role in the ability for a disease to spread. Brainstorming and simulating these differences can give students an idea of ways they can affect a virus's spread in their homes, schools, and communities. It also provides them with a chance to see some of the important factors being considered in public health models, giving them a stronger knowledge base for interpreting data, statistics, and graphical representations being presented in the media and through social media networks. Allowing students to present their findings through posters, community announcements, or through social media can create a real opportunity for students to advocate for public health in their schools and communities.

Using the pandemic for my summer course meant that I was creating new material each day, and that was stressful; but in doing so, I felt that I was helping my students not only develop their skills in math and science, but also helping them become more aware of the mathematics impacting decisions in public health. While I hope they took away important concepts in both fields, their willingness to learn and engage with STEM in a way that felt genuinely seamless showed that students are ready to persevere and connect their understandings across disciplines when given the chance. When asked to reflect on their experience over the summer, a student responded: "there will always be new things to pay attention to and that can be difficult, but your class helped me get an idea of what to expect." In terms of providing an opportunity for students to be more prepared to make scientifically supported decisions regarding their health and the health of their families (through a fun STEM activity, no less), I could not ask for much more than that.