28 Kids Later: Understanding Flu Spread at School

Group 211
Nathan Kai Ching Chan
7/21/2025

Abstract

This project simulates the spread of the flu in a school of 61 kids, where one particular kid is infected for three days, and has a 1% chance of spreading to other kids. R is utilized for examining the flu's progression over time and to measure the expected number of kids infected on certain days (particularly days 1 and 2).

The results indicate that most epidemics are short-lived, typically lasting less than 10 days, with a strong right-skewed distribution. The expected number of infections by Days 1 and 2 is supported by simulations that approximate the mean number of kids infected. Histograms and five-number summaries are provided to highlight patterns and distributions of an epidemic timeline. Additionally, the effectiveness of vaccination is supported by a direct comparison between the duration and expected number of infections of no immunization versus 50% chance of immunization, where duration and expected number of infections decrease drastically when there is a greater chance of immunization.

Background

Following various assumptions, the project explores the spread of influenza in a school of 61 children. One student, Tommy, is infected with the flu on the first day, and remains infectious for a total of three days. The remaining 60 kids are equally as healthy and susceptible to the infection, with no consideration of immunizations until the final scenario. It is assumed that the probability of Tommy infecting any of the kids over the next three days is $p = 0.01$. A literature review precedes five questions to be answered, before concluding with lessons learned.

Literature Review

It is beneficial to explore past literature for myself and fellow readers, which will help guide this project to unlock new ideas and expectations before the analysis.

Endo et al. (2009) explored data on 29 Matsumoto primary schools to investigate the spread of influenza in three cases: within the same classroom, different classes within the same grade level, and between different grade levels. Their model also stated:

$$\lambda_i = \sum \beta_e \times \frac{\#\ infectious\ students\ in\ environment}{total\ students\ in\ environment}$$

Where:

$\lambda_i =$ probability for any given student, i, to be infected

$\beta_e =$ rate of transmission in an environment, e

This is significant because the researchers explored varying levels of contact that a student may realistically encounter daily. The research also simulates the spread of influenza by assigning labels of “susceptible”, “infected”, and “recovered” (based on the SIR model). If a student is “infected”, they have a probability of infecting their peers, depending on the environment (e.g., a classroom). Furthermore, this is a stochastic process that is simulated many times to capture the randomness, but with consistent odds of becoming infected. Ultimately, the research found that the majority of flu cases occurred in the classroom environment, and transmission between different grades was much rarer.

The implications suggest that interventions are beneficial, but focusing on smaller-scale interventions, such as classroom quarantines, rather than shutting down the entire school, is more effective. This is significantly influential for this project, as the scope is also focused on within-classroom transmissions, and their modeling of infection events as probabilistic is a shared approach.

Main Findings

(a) Distribution of the number of kids that Tommy infects on Day 1?

The first part focuses on understanding the probability of Tommy infecting x number of kids on the first day. Assuming independence between kids and days, this is a Binomial distribution with n = 60 and p = 0.01, or a 1% chance of spreading the flu. In addition, a PMF (using dbinom in R) captures the exact probabilities of the number of infections distributed on day 1.

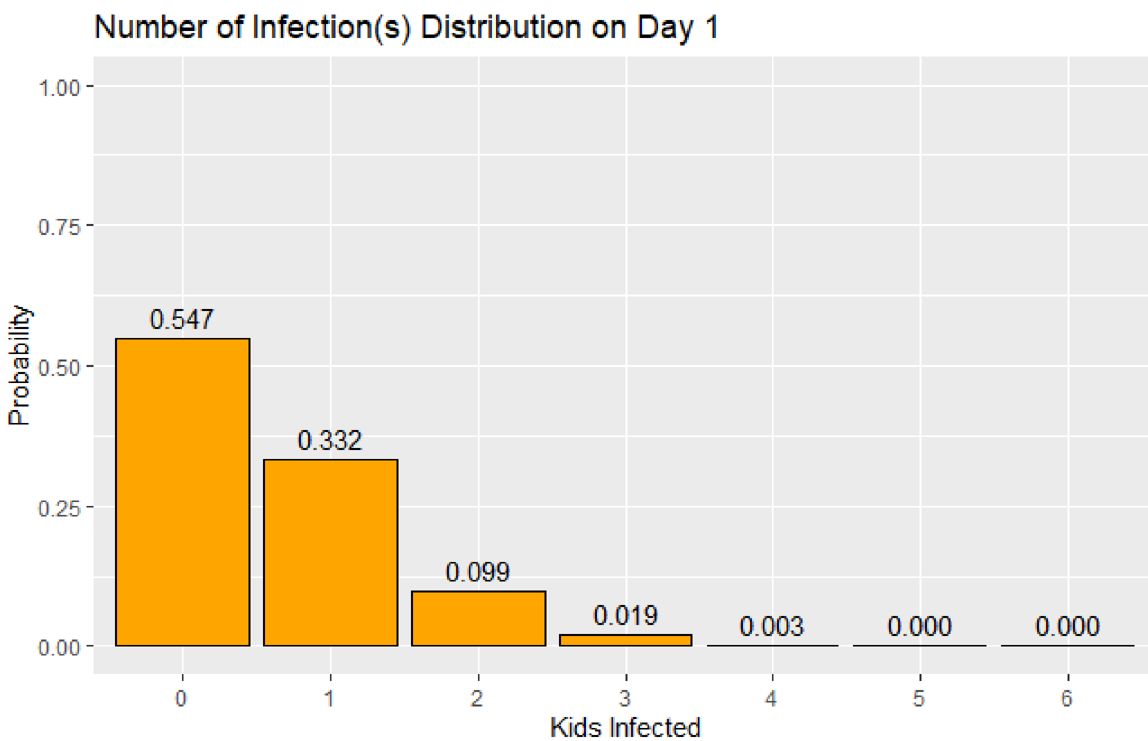

Figure 1 (above): Visualizing the probability of the number of kids infected by Tommy on Day 1

Figure 1 illustrates the distribution where the most likely outcome (54.7% of the time) is that Tommy does not infect anyone on Day 1, and 33.2% of the time, Tommy infects one kid. This implies that generally, an influenza outbreak does not begin on the first day. However, Tommy has two more chances (two more days).

(b) What is the expected number of kids that Tommy infects on Day 1?

Another way to ask this question is if this scenario were to be repeated many times, what is the average number of kids infected by Tommy on Day 1? The expected value of a Binomial variable is as follows:

$$E(x_1) = n \times p$$

Where:

- $E(x_1)$= Expected number of students infected on Day 1
- n = Number of possible victims
- p = Probability of getting infected

Given that n = 60 and p = 1%, the expected value is 0.6, meaning that on average, Tommy infects 0.6 kids on Day 1. To validate this result, a Binomial(60, 0.01) distribution is simulated 5,000 times, and thus the mean number of infections should approximate 0.6.

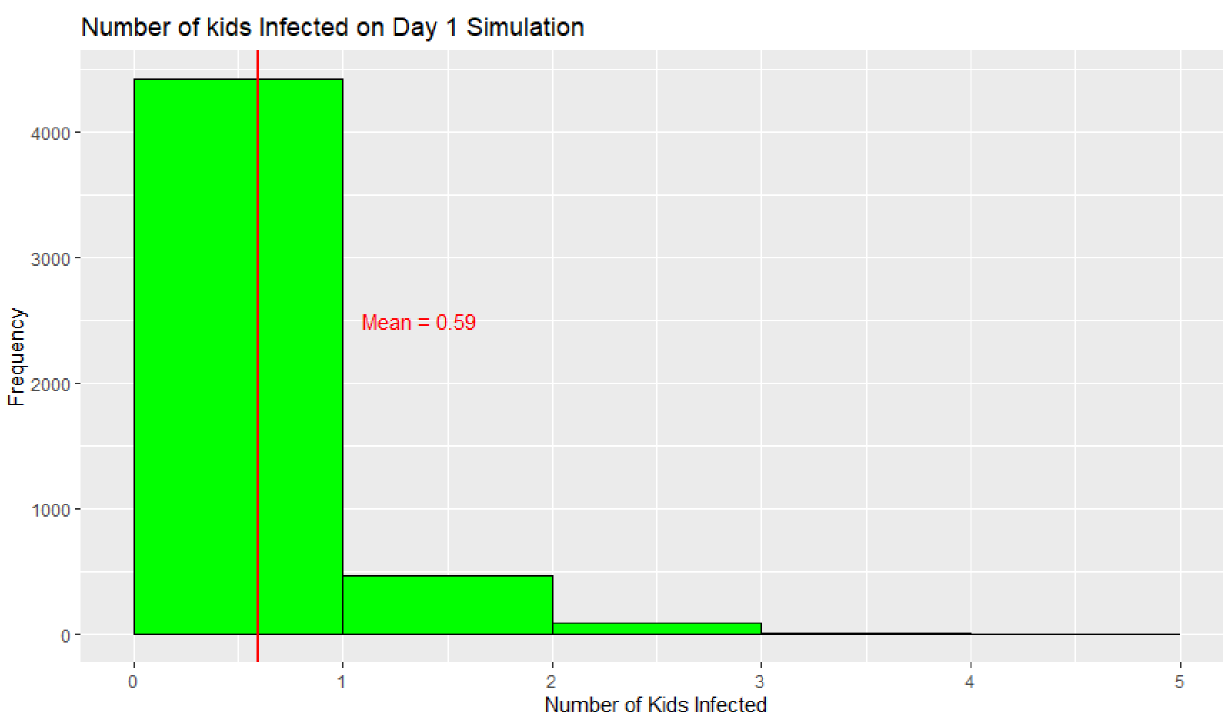

Figure 2 (above): Plot of Binomial (60, 0.01) distribution with 5,000 trials

Figure 2 also highlights the mean value of 0.59, marked by a red vertical line, indicating that more trials would more closely approximate the expected value of 0.6. Nonetheless, this still reinforces the expectation that, on average, Tommy will infect **0.6 people on Day 1**.

(c) What is the expected number of kids that are infected by Day 2?

Assuming that Tommy is included as an infected, the formula for the expected number of kids infected by Day 2 is as follows:

$$E(x_2) = Tommy + E(x_1) + E(x_2^{Tommy}) + E(x_2^{Kids})$$

Where:

- $E(x_2)$ = Expected number of students infected by Day 2
- $Tommy$ = The initial flu spreader
- $E(x_1)$ = Expected number of students infected on Day 1
- $E(x_2^{Tommy})$ = Expected number of students infected on Day 2 by Tommy
- $E(x_2^{Kids})$ = Expected number of students infected on Day 2 by kids that were infected on Day 1

The four components of this formula can be calculated:

$Tommy = 1$

$$E(x_1) = n \times p = 60 \times 0.01 = 0.6$$

$$E(x_2^{Tommy}) = [n - E(x_1)] \times p = [60 - 0.6] \times 0.01 = 0.594$$

$$E(x_2^{Kids}) = E(x_1) \times [n - E(x_1) - E(x_2^{Tommy})] \times p = 0.6 \times [60 - 0.6 - 0.594] \times 0.01$$

$$= 0.6 \times 58.806 \times 0.01 = 0.353$$

Therefore:

$$E(x_2) = 1 + 0.6 + 0.594 + 0.353 = 2.547$$

This suggests that, on average, we expect **2.547 infections by Day 2** with the inclusion of Tommy, and this number is partly attributed to the possibility that infectees (who were previously infected by Tommy) will also have a chance of infecting other susceptibles on the next day.

To put this number to the test, the next step will try to approximate the expected value based on 5,000 trials.

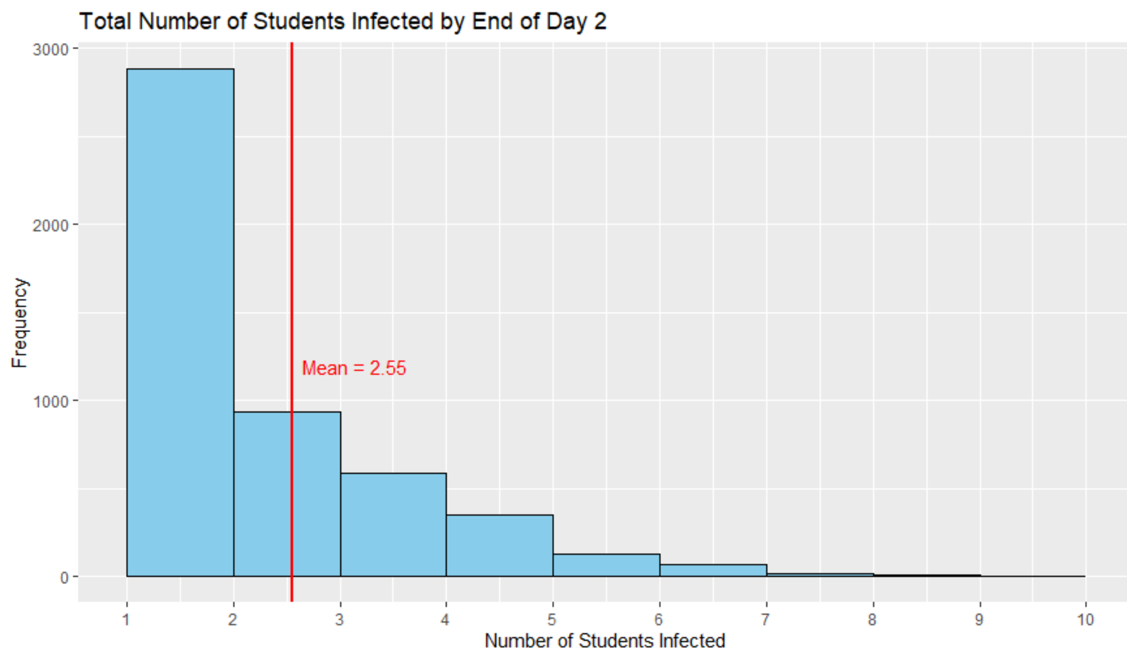

Figure 3 (above): Distribution of the number of students infected by Day 2 based on 5,000 trials. When including Tommy as an infected, the number will always be greater than 1 in any of the trial runs.

Figure 3 illustrates that the mean is 2.55 infected students, which is also extremely close to the theoretical estimate of 2.547. Compared to day 1, it is far more likely to have a major outbreak involving three or more infected children, or roughly 5% of the children's population in this scenario. Thus, this highlights the potentially dangerous day-to-day consequences of inaction regarding the spread of the flu in schools.

(d) Simulate the number of kids that are infected on Days 1, 2… What are the (estimated) expected numbers of kids that are infected by Day i, i = 1, 2…? Provide a histogram detailing how long the "epidemic" will last.

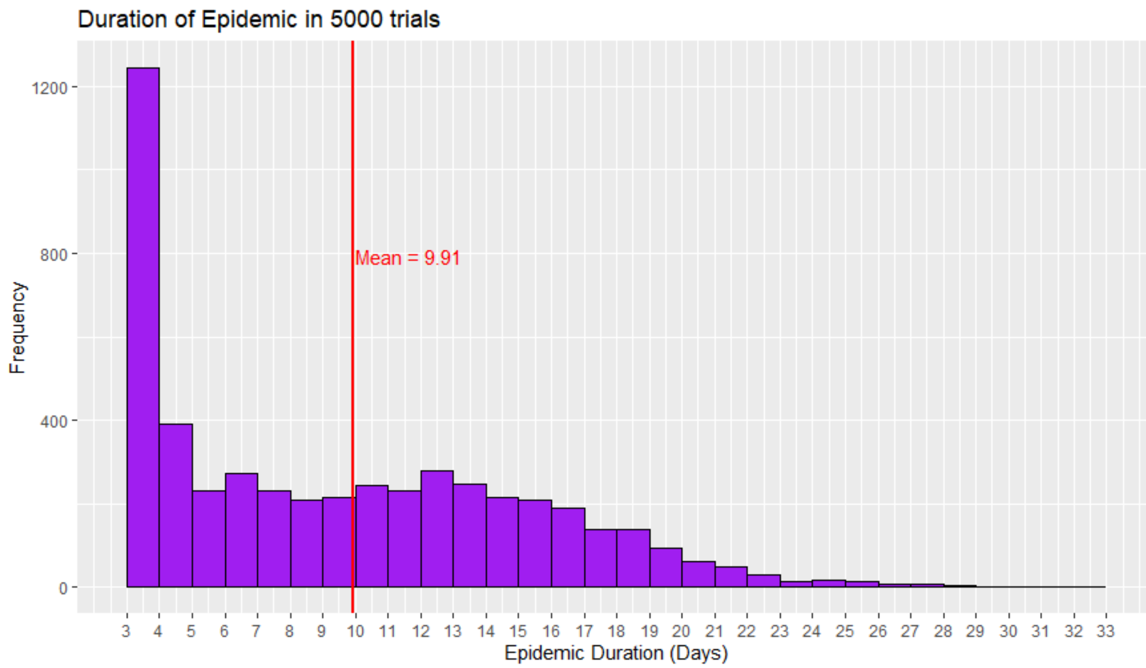

Figure 4 (above): Histogram showcasing the simulation of the duration of the flu epidemic

There are further considerations for this simulation:

- The minimum epidemic duration is 3 days, which is equivalent to Tommy failing to infect other kids in all 3 days.
- A kid that is infected on any given day, D_i, becomes infectious starting on the next day, and can infect susceptible kids between D_{i+1} to D_{i+3}.
    - e.g., Billy is infected by Tommy on day 1; thus, Billy can infect others from day 2 to day 4, but is no longer infectious on day 5 and beyond.
- A kid who has been infected will be removed from the susceptible pool and cannot be reinfected.
- Simulation length is capped at 60 days.

Five-Number Summary of Epidemic Duration

Statistic	Value
Minimum	3.00
1st Quartile (Q1)	5.00
Median	9.00
Mean	9.91
3rd Quartile (Q3)	14.00
Maximum	33.00

Table 1 (above): Five-number summary of the epidemic duration simulation

Table 1 shows that, in a simulation of 5,000 trials, the mean epidemic duration was approximately 10 days. This is also a right-skewed distribution as the mean is greater than the median of 9 days. In a very rare case, the epidemic lasted for a maximum of 33 days, but the most frequent duration was a minimum of 3 days. Furthermore, 25% of outbreaks ended in 5 days, which suggests that in these cases, the last infected kid was infected on the second day. Early prevention and intervention should be a key focus for schools, which in turn will help reduce the likelihood of long-lasting epidemics.

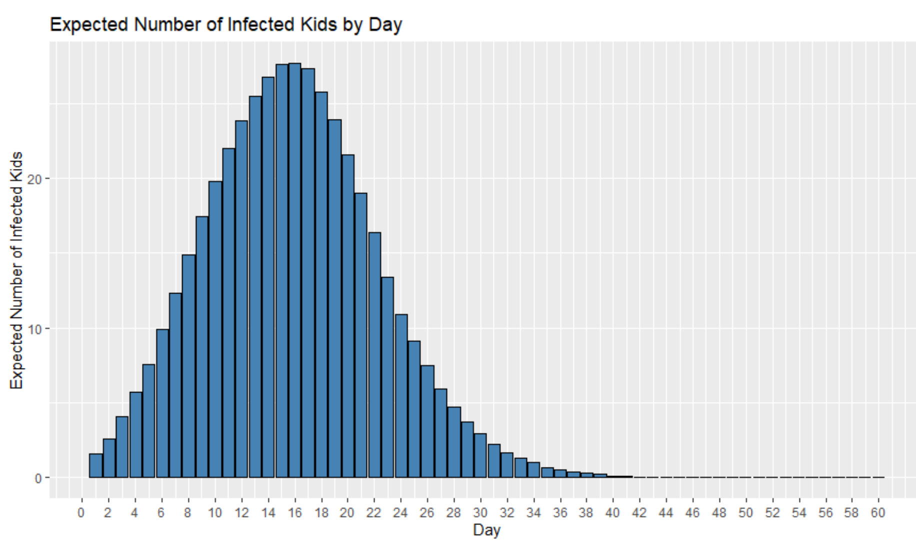

Figure 5 (Above): Expected Number of Infected Kids with no Immunization

Figure 5 articulates the expected number of infected kids throughout the epidemic, with the mode of around 27 kids being infected on day 16. The gradual decrease in the expected number of infected kids after day 16 illustrates the effects of post-infection immunity (no longer susceptible after being infected for 3 days), resulting in no expected infected kids after 7 weeks.

(e) What if each kid has a 50-50 chance of already being immunized (and the immunization works perfectly)?

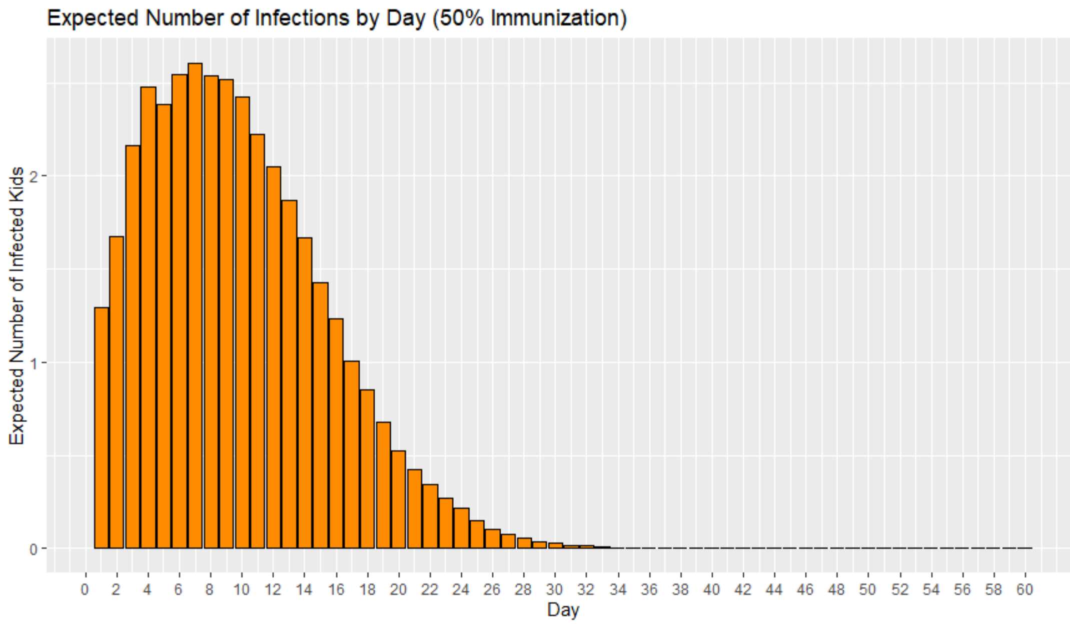

Figure 6 (Above): Expected Number of Infected Kids with 50% Chance of Immunization

Lastly, Figure 6 illustrates the positive effects of immunization against influenza. When each kid has a 50% chance of being immune to the flu, the expected number of infected kids will reach its peak in about 8 days and will gradually decline over the next few weeks. More importantly, the expected number of infected kids at day 8 is approximately 2.7 kids, which is significantly lower compared to Figure 5, where there is no immunization.

Conclusion

This project investigated a scenario in which 61 children navigated a flu outbreak. Several key findings emerged. First, the earliest days of the epidemic are the most critical—if Tommy fails to infect others within the first three days, the epidemic usually ends quickly. However, once Tommy succeeds in infecting another, the number of infected can grow at an increased rate, highlighting how even small probabilities of transmission can lead to significant spread over time.

Second, the introduction of a 50% chance of immunization drastically reduced both the duration and magnitude of the epidemic. This implies the real-world importance of vaccination programs, health education, and proactive health measures in schools.

For future work, this simulation could be expanded in several ways: by introducing different environments, as in the case of the literature review (same classrooms versus different grade/age students), or by varying the infection probability over time to reflect behavioral changes. Additionally, incorporating real-world data—such as historical absentee records or vaccination rates—could enhance the model's realism and predictive value.

References

Endo (遠藤彰), A., Uchida (内田満夫), M., Liu (刘扬), Y., Atkins, K. E., Kucharski, A. J., Funk, S., Abbas, K., van Zandvoort, K., Bosse, N. I., Waterlow, N. R., Tully, D. C., Meakin, S. R., Quaife, M., Russell, T. W., Jit, M., Foss, A. M., Rosello, A., Quilty, B. J., Prem, K., … Flasche, S. (2022). Simulating respiratory disease transmission within and between classrooms to assess pandemic management strategies at Schools. *Proceedings of the National Academy of Sciences, 119*(37). https://doi.org/10.1073/pnas.2203019119