# Modeling Impacts of Vaccine Uptake: A Simulation-Based Extension of the Reed-Frost Model

Georgia Institute of Technology

ISYE 6644: Simulation and Modeling
*Author: Zachary Lewis*

**ABSTRACT**

Research into flu pandemics and the impact of infectivity, immunization, and other factors have been widely studied since the early 20th century. In this paper, we seek to replicate the foundational Reed-Frost model using simulation-based techniques with extensions for vaccine uptake to better understand the dynamics of a flu pandemic in an elementary classroom setting. Results provide strong evidence for the benefits of increasing vaccine uptake on all metrics of pandemic severity, with 50% uptake decreasing expected length by 70% and 80% uptake decreasing expected length by 80%.

## 1. BACKGROUND AND DISCUSSION OF PROBLEM

### 1.1 Overview

The mechanisms through which disease spreads and the behavior of pandemics are rooted in microbiology and other fields which have greatly increased our understanding of diseases and their health impacts while improving our ability to mitigate their spread. Nonetheless, it is possible to view a pandemic through a more mechanical lens - as a system composed of entities, with outcomes and interactions modeled by relatively simple statistical distributions. In fact, one of the first major simulation-based approaches to epidemiology relied on a box with colored marbles and a wooden trough [1]. This paper explores additional simulation-based techniques which build upon foundational methodologies and incorporate vaccine uptake to model impact on expected pandemic length and peak in a population.

To this end, we explore the dynamics of a flu pandemic in an elementary classroom with 60 healthy students, and 1 student (known as Tommy) infected with the flu. Tommy and any other subsequently infected students are contagious for 3 consecutive days, with an infection probability $\mathbf{p} = 0.01$ per interaction. All students and days are independent, resulting in $\mathbf{k}$ i.i.d. Bern($\mathbf{p}$) trails per infected student where $\mathbf{k}$ is the number of susceptible students.

Key questions we will be explore include:

1. How does the # of infected students on day 1 distribute?
2. What is the expected number of infected students on day 1, day 2, and beyond?
3. How long is the pandemic expected to last?
4. What is the impact of an expected immunization rate of 50%?

### 1.2 Literature Review

The flu infects 8% of the U.S. population on average each year with impacts varying based on strain specific characteristics and other factors [2]. The flu is one of the deadliest respiratory diseases globally, and has sparked worldwide pandemics in 1918, 1957, and 1968 [3], resulting in considerable research into the modeling of flu pandemics. Approaches include the simple SIR model [3], stochastic, agent-based models [4], and chain binomial models [5].

Chain binomial models are effective for modeling secondary attack risk and are useful in comparing simulated cases to observed in various disease scenarios [6]. One of the earliest stochastic models, the Reed-Frost model formulated in 1928, is a chain binomial which models a pandemic's development over multiple generations [7]. The Reed-Frost model is a relatively simple iterative model with two key assumptions:

- Each infected individual independently infects each susceptible individual in the population with some probability **p** [7].
- Individuals that are infected by generation **t** constitute generation **t+1** and generation **t** is removed from the process [7].

These assumptions are useful as they align closely with the assumptions of our classroom setting, and thus we will leverage the Reed-Frost model as a baseline for our modeling approach.

## 2. METHODOLOGY

Our approach leverages two related methodologies to simulate the classroom pandemic and impact of vaccine uptake. We leverage an adapted Reed-Frost model which allows for multiple generations (days) of infectivity per student to set a baseline model for our pandemic. Subsequently, we implement an approach which simulates the classroom pandemic using outcomes generated from a Binomial distribution in combination with a tracking dictionary of individual student health statuses. We compare the two approaches using a Chi-Squared test and Kolmogorov-Smirnov test to validate that our simulation-based approach replicates the Reed-Frost model. Assuming a satisfactory outcome from the Goodness-of-Fit tests, the simulation-based approach then incorporates vaccination uptake to model the impact of various rates as an extension of the Reed-Frost model. All models are implemented in Python.

### 2.1 Adapted Reed-Frost Model

Our Adapted Reed-Frost Model (ARFM) follows the original formulation [1] below, where $\boldsymbol{I_t}$ represents cases at time **t**, $\boldsymbol{S_t}$ represents susceptible individuals at time **t**, and $\boldsymbol{\mathcal{B}(x)}$ is a Bernoulli random variable with probability **x**.

$$\boldsymbol{I_{t+1}} = \sum_{\boldsymbol{k=0}}^{\boldsymbol{S_t}} \boldsymbol{\mathcal{B}(1-(1-p)^{I_t})}$$

$$\boldsymbol{S_{t+1}} = \boldsymbol{S_t} - \boldsymbol{I_{t+1}}$$

The key adaptation of our implementation involves implementing tracking the # of infected at time **t** ($\boldsymbol{I_t}$) and # of days infected ($\boldsymbol{D_t}$), as well as a parameter (**d**) for days contagious. We update $\boldsymbol{I_{t+1}}$ to include:

$$\sum_{x=0}^{t} I_{t-x} \; where \; D_{t-x} < d$$

allowing for multiple generations of infectivity. When parameter **d** (days contagious) = 1, the ARFM is equivalent to the original Reed-Frost model. In order to return **E[X]** at time **t** where **X** = total infected or **X** = active contagious, the adapted model is run with 1000 replications and **E[X]** and associated confidence intervals are obtained from generated results.

### 2.2 Simulation Based Tracking Model

The Simulation Based Tracking Model (SBTM) is a rules-based model which relies on a central tracking dictionary of individual cases and outcomes generated from a Bernoulli distribution to model a pandemic over time. The dictionary is constructed of **n** entries with each key set to patient_x where x is an assigned integer, and each value contains a dictionary with the following attributes:

1. Infected (Boolean, default 0)
2. Days Infected (Integer, default 0)
3. Vaccinated (Bernoulli Random Variable)

Vaccinated values are randomly generated from the Bernoulli distribution based on an input parameter for vaccination uptake rates.

The model generates outcomes at time **t** using the following process:

1. For every contagious individual present (Infected == 1), every susceptible individual (Infected == 0) is exposed and infected with probability **p,** with the outcome generated from a Bernoulli distribution.
2. If the individual is vaccinated, the infection is nullified.
3. If the individual is not vaccinated, their Infected value is set to 1 and their Days Infected count is set to "incubating" (count becomes 1 on the subsequent day).
4. The # of infected individuals is reported by summing up Infected variables in the dictionary.
5. The # of contagious individuals is reported by summing up Infected variables in the dictionary where Days Infected < **d**, where **d** = parameter for # days contagious.

The simulation is then run for a number of days and replications, with results for **E[X]** and associated confidence intervals obtained in the same manner as for the ARFM. When vaccination rates for the SBTM are set to 0, the model is analogous to the ARFM.

## 3. RESULTS

Our main analyses involve comparing simulated results to analytical methodologies, comparing results from the ARFM and SBTM approaches, and an exploration of the impacts of vaccination rates on pandemic duration. These steps validate the foundations of our models with basic probability theory, confirm that we have accurately simulated the Reed-Frost Model baseline, and extend the model's behavior to incorporate vaccine uptake respectively.

### 3.1 Simulation 1: Basic Probability

Given that our classroom setting involves independent i.i.d. Bern(**p**) trials for each infected student **k**, we would expect the # of students infected by our patient 0 (Tommy), to follow a Binomial distribution where n = 60 and p = 0.01. Under this assumption, the expected number of students infected on Day 1 would be $np = 0.6$, and the probability that 0 students are infected would be:

$$\mathcal{P}_0 = \binom{60}{0}.01^0(1-.01)^{60} = 0.547 \textbf{ (1)}$$

Given that each student is contagious for 3 days and each day is independent, the probability that patient 0 infects no other students after 3 days would be:

$$\mathcal{P}_{No\ Infections} = (0.547)^3 = 0.164 \textbf{ (2)}$$

The expected number of new cases on Day 2 of the pandemic is dependent on cases from Day 1 and is given by:

$$\sum_{x=0} \mathcal{P}\binom{60}{x}0.01(60-x)(1+x) \sim 0.94 \textbf{ (3)}$$

where $\mathcal{P}\binom{60}{x}$ gives the probability of x cases on Day 1, and the remaining term adjusts the expected value $np$ based on the outcome of Day 1. Thus, the total number of expected cases on Day 2 is approx. 2.54 = Patient 0 + E[X] Day 1 + E[X] Day 2.

Analytical solutions for modeling pandemic dynamics become increasingly complex as variables are introduced as evidenced by the small example above. Thus, we aim to replicate analytical results using simulation which allows greater flexibility and customization in our analyses. The behavior of the Binomial is replicated by repeated samples from software packages such as NumPy, and long-term behavior approximates the true expected value. Results of simulations with 5000 iterations are shown in Figures 1 and 2, with results in line with expectations from basic probability techniques.

**Fig 1.**

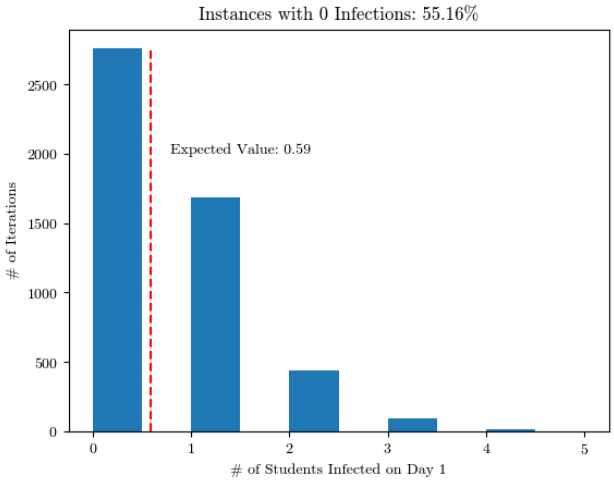
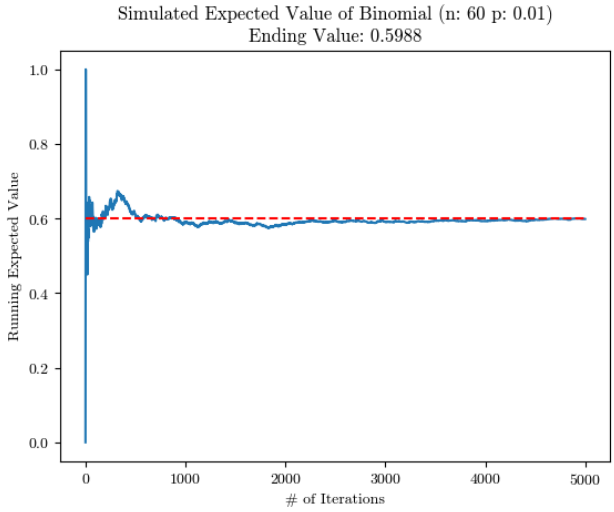



**Fig 2.**

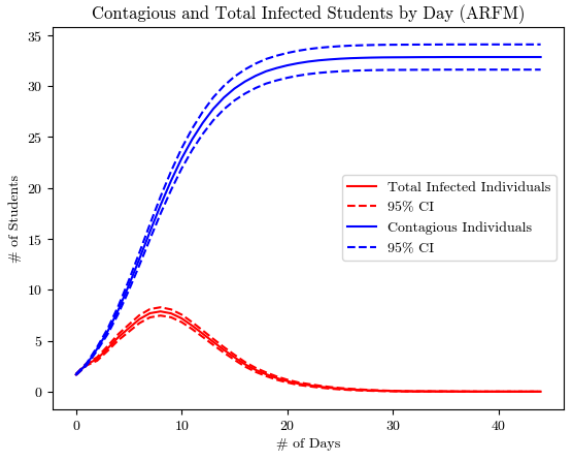



The same fundamental simulation processes underpin approaches in subsequent sections, with the expectation that results will align with results demonstrated in (1)(2) and (3).

### 3.2 Simulation 2: ARFM Results

Implementation of the ARFM using 1000 replications results in the pandemic curve and associated 95% CIs found in Figure 3 and provides the baseline for comparing against our SBTM method. The pandemic peaks at day 8 and infects approx. 30 out of 60 students.

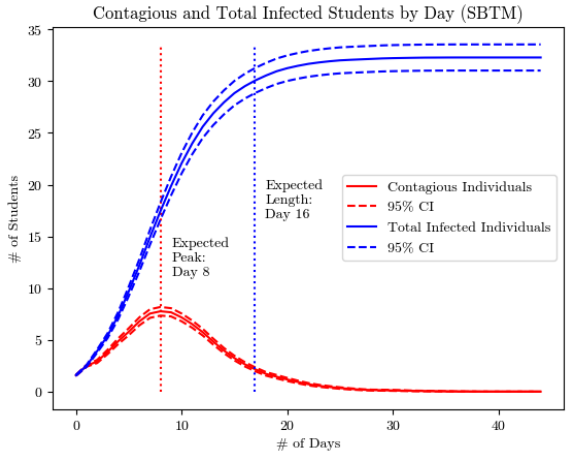
**Fig 3.**




### 3.3 Simulation 3: SBTM Results

Implementation of the SBTM using 1000 replications results in the pandemic curve and associated 95% CIs found in Figure 4, with an expected peak at day 8 and infects approx. 30 out of 60 students.

**Fig 4.**




### 3.4 Simulation 4: Model Comparison

Visual inspection of the models' pandemic curves indicates similarity, which is confirmed by statistical tests performed on the curves for Contagious Individuals. Figures 5 and 6 demonstrate high p-values for Chi-Squared and Kolmogorov-Smirnov tests for cases where the contagious period is 3 days and 1 day respectively, indicating an exceedingly low

probability that the models produce significantly different results.

**Fig 5.**

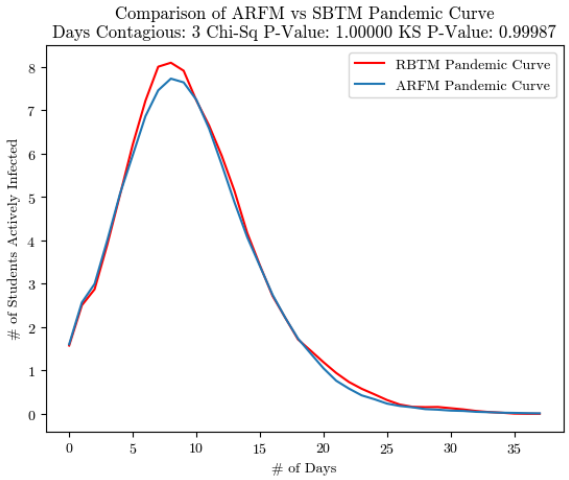



**Fig 6.**

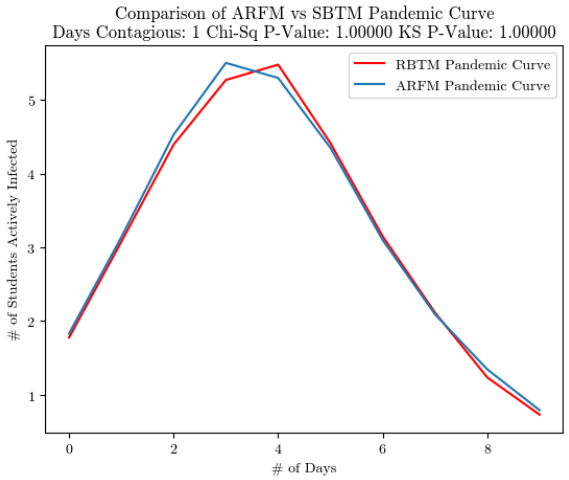



Figure 7 provides additional validation that our models provide good approximations and are well suited for modeling more complex extensions by comparing against analytical results achieved in (2) and (3).

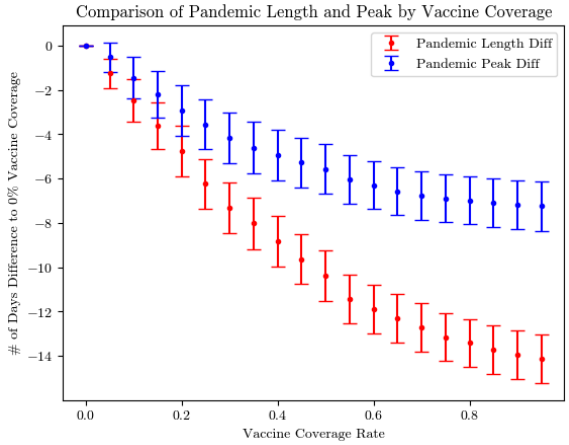
**Fig 7.**

| | **Analytical Models** | **ARFM** | **SBTM** |
|---|---|---|---|
| Expected Cases on Day 1 | 1.6 | 1.62 | 1.58 |
| Expected Cases on Day 2 | 2.54 | 2.56 | 2.49 |
| % with No Infections (10k Replications) | 16.40% | 16.22% | 16.26% |

### 3.5 Simulation 5: Vaccine Extensions

Having established satisfactory similarity to analytical and Reed-Forest models, we are able to investigate impacts of vaccine uptake (which assume 100% efficacy) with our SBTM model. We compare the pandemic length and peak for various vaccine uptake rates using confidence intervals for differences derived from a paired t-test, accomplished using common random numbers. Figure 8 displays results for various uptake rates, with all rates (0.05+) showing a statistically significant improvement over no vaccination. Uptake rates of 50% lead to a 6 day decrease in expected peak (75% decrease) and an 11 day decrease in expected length (68.8% decrease)

**Fig 8.**




## 4. DISCUSSION

Simulation-based approaches approximate results derived from analytical techniques and have extensions to better understand the impact of vaccination uptake. Even relatively low uptake rates have significant benefits towards reducing pandemic length and severity, indicating positive impacts achieved by increases in immunization rates at any level.

Our approach provides a foundation which could be expanded to incorporate currently missing features such as vaccine efficacy which could improve the model's practicality. Of note are limitations of the current approaches – factors external to the classroom are not

accounted for, and students are assumed to interact with every classmate. Modifications to address these limitations could further improve model performance when compared against observed data.

## 5. CONCLUSION

Simulation-based modeling techniques allow us to effectively extend simple epidemiological models to incorporate more variables and external factors. These approaches approximate results found with analytical techniques and allow for increased customization. Our approach effectively modeled the dynamics of a flu pandemic in an elementary classroom, and the positive impacts that immunization have on pandemic length and severity. Results provide strong evidence for the benefits of increasing vaccine uptake on all metrics of pandemic severity, with 50% uptake decreasing expected length by 70% and 80% uptake decreasing expected length by 80%.

## 6. REFERENCES


[1] L. Engelmann, "A box, a trough and marbles: How the Reed-Frost epidemic theory shaped epidemiological reasoning in the 20th century," *History and Philosophy of the Life Sciences*, vol. 43, no. 3, Aug. 2021, doi: https://doi.org/10.1007/s40656-021-00445-z.

[2] J. I. Tokars, S. J. Olsen, and C. Reed, "Seasonal Incidence of Symptomatic Influenza in the United States," *Clinical Infectious Diseases*, vol. 66, no. 10, pp. 1511–1518, May 2018, doi: https://doi.org/10.1093/cid/cix1060.

[3] F. Brauer, C. Castillo-Chavez, and Z. Feng, "Models for Influenza," *Texts in Applied Mathematics*, pp. 311–350, 2019, doi: https://doi.org/10.1007/978-1-4939-9828-9_9.

[4] J. Whitman and C. Jayaprakash, "Stochastic modeling of influenza spread dynamics with recurrences," *PLOS ONE*, vol. 15, no. 4, p. e0231521, Apr. 2020, doi: https://doi.org/10.1371/journal.pone.0231521.

[5] Y. Sharker and E. Kenah, "Estimating and interpreting secondary attack risk: Binomial considered biased," *PLOS Computational Biology*, vol. 17, no. 1, p. e1008601, Jan. 2021, doi: https://doi.org/10.1371/journal.pcbi.1008601.

[6] S. AKHTAR, T. E. CARPENTER, and S. K. RATHI, "A chain-binomial model for intra-household spread of Mycobacterium tuberculosis in a low socio-economic setting in Pakistan," *Epidemiology and Infection*, vol. 135, no. 1, pp. 27–33, Jun. 2006, doi: https://doi.org/10.1017/s0950268806006364.

[7] M. Deijfen, "Epidemics and vaccination on weighted graphs," *Mathematical Biosciences*, vol. 232, no. 1, pp. 57–65, Jul. 2011, doi: https://doi.org/10.1016/j.mbs.2011.04.003.