# Allocating limited COVID-19 vaccines to non-seniors minimizes deaths/infections across *all* age groups

Jonathan Lee

17 November 2020


## Abstract

With the anticipation of a COVID-19 vaccine, the government and public health officials face a new challenge: how to prioritize available doses. To address this, we utilize an age-stratified SIRV model to estimate the effectiveness of a vaccine, and to evaluate the impact of different prioritization strategies on cumulative infections and deaths. We find that a vaccine should be prioritized for people under the age of 65 to minimize both cumulative incidence and deaths.


In December 2019, a novel strand of Coronavirus (SARS-CoV-2) was identified in Wuhan, Hubei Province, China [1]. On March 11, the World Health Organization (WHO) declared the COVID-19 outbreak a pandemic [2]. As of November, there have been 51.9 million cases and 1.28 million deaths reported worldwide [3]. To date, a variety of interventions have been implemented to slow the spread of the virus including travel restrictions, shelter in place orders, and remote work and schooling. These tactics are temporary solutions, but for life to return to normal, the world desperately needs a vaccine.

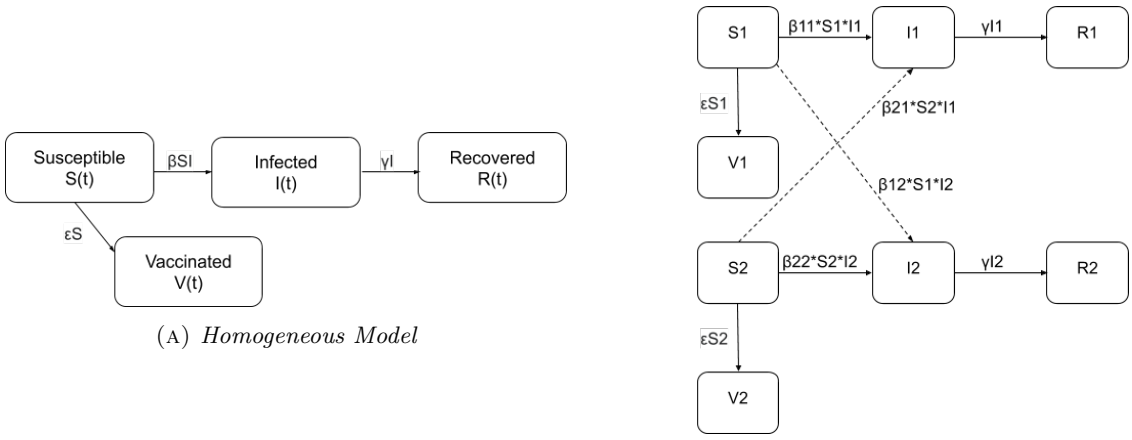
As many candidate vaccines head to clinical trials, it appears that scientists may be narrowing in on conferring immunity to SARS-CoV-2. A first generation of COVID-19 vaccines are expected to gain approval as soon as the end of 2020 or early 2021. However, once a vaccine is approved to treat coronavirus, the enormous demand will likely outweigh the limited initial supply. As a result, mathematical models have become an important tool to derive an optimal vaccine prioritization strategy. This paper seeks to model how distributing vaccines differently across pre-65 and post-65 age groups will influence the trajectory of a future COVID-19 wave in the United States.

This study builds on recent work by Kat Rock et al on epidemiological models that introduce a vaccinated class to account for the effects of immunization. They formulate a model and develop a suitable equation for the dynamics of an SIR-type infection with vaccination based on the level of successful immunization. We also draw on the most recent projections published by Pfizer that anticipate up to 50 million vaccine doses can be produced in 2020 [4]. We combine this research to create an age stratified SIRV model with a maximum of 50 million available vaccines, and use the CDC's provisional COVID-19 deaths data to estimate our beta coefficients. We then analyze how different vaccine allocations would affect a future outbreak, and conclude that while Pfizer's 50 million doses are insufficient to induce nationwide herd immunity, allocating them to nonsenior citizens best minimizes infections and deaths.

# 1 Methods

We describe our differential equations model, data collection process, and parameter estimation via contact matrices and *fmincon*. Notably, section 1.2 describes an original procedure for estimating "active infectious cases" per day from the CDC's death data, and vice versa. This is pertinent because (virtually) no dataset directly reports "active infectious cases" per day while differentiating by age group.

## 1.1 Age-Stratified SIRV




(A) *Homogeneous Model*

(B) *Heterogeneous Model*

FIGURE 1.1.1: *Compartment Diagrams: Homogeneous vs Age-Stratified SIRV*

The model contains several features.

1. Two age strata. Unlike the homogeneous SIRV model, we distinguish between a "younger group" (age $< 65$) and an "older group" (age $\geq 65$). This division coincides with the active versus retired senior population, corresponding to distinct patterns of social interaction and thus transmission.
2. Compartments. The U.S. population of 328.2 million is split into four compartments: susceptible ("S"), infected ("I"), recovered ("R"), and vaccinated ("V"). Birth/death dynamics are ignored, so people cannot be added to the susceptible group, and can only be removed by being infected or receiving a vaccine. We also assume the vaccine is 100% effective, so vaccination confers total immunity. Once an infected individual has recovered, we assume they also develop full immunity.
3. Homogeneously mixed populations. We assume all individuals in each age group are equally susceptible and equally infectious if they become infected.
4. Vaccinations. We assume that in this future outbreak, all 50 million vaccines are ready to distribute.

The following differential equations describe the dynamics of individuals in each of the compartments of Figure 1.1.2:

$$\dot{S}_1 = -\beta_{11}\frac{I_1}{N_1}S_1 - \beta_{12}\frac{I_2}{N_2}S_1 - \epsilon S_1 \qquad \dot{S}_2 = -\beta_{22}\frac{I_2}{N_2}S_2 - \beta_{21}\frac{I_1}{N_1}S_2 - \epsilon S_2$$

$$\dot{I}_1 = \beta_{11}\frac{I_1}{N_1}S_1 + \beta_{12}\frac{I_2}{N_2}S_1 - \gamma I_1 \qquad \dot{I}_2 = \beta_{22}\frac{I_2}{N_2}S_2 + \beta_{21}\frac{I_1}{N_1}S_2 - \gamma I_2$$

$$\dot{R}_1 = \gamma I_1 \qquad \dot{R}_2 = \gamma I_2$$

$$\dot{V}_1 = \epsilon S_1 \qquad \dot{V}_2 = \epsilon S_2$$

Four final comments on the model:

1. In the Results section, we conduct simulations where $V_1 + V_2$ cannot exceed 50 million, reflecting the U.S.' limited vaccine supply in 2021. In that case, our model becomes a *piecewise* set of functions, where once $V_1 + V_2$ equal 50 million, $\epsilon$ is set to zero, reflecting no more people can be vaccinated.
2. $\epsilon$ is held constant at 1/100 across both groups because both senior and nonsenior citizens have expressed similar willingness to be vaccinated in polls, and ethnicity and political affiliation are likely

| Variables | Description |
|---|---|
| subscript $1, 2$ | denotes pop. 1 (age $< 65$), pop. 2 (age $> 65$) |
| $\beta_{ij}$ | (freq. of contact) $\times$ ($Pr$[susceptible $i$, infected $j$ contact $\rightarrow$ infection]) |
| $N_i$ | subpopulation size: $N_i = S_i + I_i + R_i + V_i$ |
| $S_i, I_i, R_i, V_i$ | # of individuals in each compartment |
| $\gamma$ | recovery rate: inverse recovery time |
| $\epsilon$ | fraction of $S_i$ vaccinated in a day |

TABLE 1.1.1: *Descriptions of parameters in SIRV model*

the main determinants of vaccine enthusiasm [5]. Population sizes are also assumed to be fixed, where $N_1, N_2$ partition the entire U.S. population. Our model is valid under this assumption because $\dot{S}_i + \dot{I}_i + \dot{R}_i + \dot{V}_i = 0$ for $i = 1, 2$.

3. All simulations use RK4 with stepsize $h = 0.01$ to approximate solutions.
4. Initial conditions are set to $R_1 = 20545867$, $R_2 = 2589185$, $S_1 = N_1 * 0.999$, and $S_2 = N_2 * 0.999$. The first two values represent how many people have already recovered from COVID-19. (We assume they remain immune in a future outbreak.) The latter two values represent the catalysts for our hypothetical pandemic. (These calculations for recovered individuals per age group are simple subtractions from WorldOMeter data, and are documented in Appendix A.)

A nondimensionalized version of this system is derived in Appendix B.

## 1.2 Data Collection/Manipulation

We estimated parameters from the CDC's provisional death counts by sex, age, and week in America [6]. This dataset identifies new deaths per week for various age groups, but does not specify the number of infections. (In fact, no public dataset reports active cases vs time per age group.) We addressed this by developing a procedure to estimate active infectious cases (per day) from daily death counts:

**Procedure 1.2.1** (deaths $\rightarrow$ active cases per day)

1. Estimate the infection fatality rates (IFR) for the pre-65 and post-65 populations. We used Dartmouth Professor Andrew Levin's estimates [7]. Levin's study estimates the IFR for 10-year age groups; taking averages (weighted by strata population) provides estimates for nonseniors and seniors.
2. Construct a new column consisting of "new infections" per day, estimated via infections $= \frac{\text{deaths}}{\text{IFR}}$. This formula is valid because IFR is the ratio of deaths to infected individuals, so (deaths)/(IFR) = deaths $\cdot$ (infections)/(deaths) = infections.
3. Construct a column consisting the number of *active infectious cases* at each time. (Note that this is distinct from new cases per day.) To estimate the entries, first compute a column of cumulative cases after each week. Then, subtract the values of every $i^{th}$ row entry from the $i + 14^{th}$ row entry. This approximation is valid because the average infectious period is approximately two weeks.

In the Results section, we use the *inverse* of this method to infer new deaths from active cases per day. (This is needed to estimate the death counts from different vaccine strategies, since the SIRV model does not include a deaths compartment.)

**Procedure 1.2.2** (active cases $\rightarrow$ deaths per day)

1. Solve the recurrence for values of $c_j(t)$ when $j = 1, 2$:

$$\forall i \in [1, 14] : c_j(i) = I_j(i)$$
$$\forall i > 14 : c_j(i) = I_j(i) + c_j(i - 14)$$

where $j$ corresponds to the population ($j = 1$ is ($< 65$) pop., $j = 2$ is ($> 65$) pop.), $I_j(t)$ is the number of "active infectious individuals" in population $j$, and $c_j(t)$ is the *cumulative infection* count after each day. This is the inverse of step 2 in Procedure 1.2.1.

2. Solve for *new* infections per day by consecutively subtracting cumulative counts per day. The number of new infections (in population $j$) on day $t$ is simply $c_j(t) - c_j(t-1)$.
3. Compute the IFR for ages $0-65$ and ages 65+. (We took weighted averages of Levin's estimates for each age strata between $0-65$, then took for the remaining strata.) Multiply each age group's "new infections per day" by its respective IFR estimate. This yields (infected)(IFR) = (infected)(deaths)/(infections) = deaths per day.

These modified datasets, and the code that generates them, can be found in Appendix A.

## 1.3 Parameter Estimation

We estimated the following parameters for our age-stratified SIRV model: $\gamma = 1/10$, $\beta_{11} = 0.135$, $\beta_{12} = 0.01$, $\beta_{21} = 0.04$, $\beta_{22} = 0.02$. Our $\gamma$ estimate is based on studies that the mean recovery time is around 10 days [8]. We reached our estimates for the $\beta$ coefficients via two steps.

*First*, we created a contact matrix to derive an inequality relating each $\beta$ coefficient.

1. In the $2 \times 2$ contact matrix, $\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$, each entry $C_{ij}$ describes the rate at which an infected individual from age group $j$ contacts susceptible individuals from age group $i$.
2. We started from Prem et al.'s database of contact matrices [9], and took weighted averages of entries (based on relative population size) to reduce Prem et al.'s $8 \times 8$ contact matrices for U.S. contacts into our desired $2 \times 2$ matrix. Specifically, we found that

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} 1.0539 & 0.0822 \\ 0.3298 & 0.3868 \end{bmatrix}$$

3. Researchers have concluded that COVID-19 infection probability is constant across age groups. This means the WAIFW matrix of $\beta$ coefficients is proportional to the contact matrix [10,11].
4. Since each $\beta_{ij}$ coefficient is the product of contact frequency and infection probability, it will be directly proportional to its corresponding $C_{ij}$ value in the contact matrix. Thus, we used the relative values of contact matrix entries to conclude that $\beta_{11} > \beta_{21} > \beta_{22} > \beta_{12}$.

*Second*, we used Matlab's optimization toolbox to reconcile these contact rate findings with the CDC data.

1. Our constraint on the $\beta$ coefficients required us to disregard our initial "best fit" estimate from Matlab's *fmincon* algorithm, which fit the CDC data best but exhibited $\beta_{22}, \beta_{12} > \beta_{11}, \beta_{21}$. When specifying $\beta_{11} > \beta_{21} > \beta_{22} > \beta_{12}$, *fmincon* insisted on returning

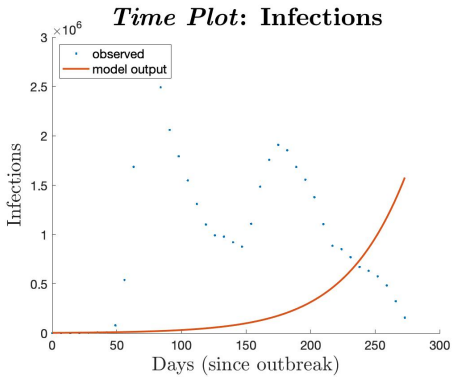



FIGURE 1.3.1: *fmincon's questionable "best fit"*

which, as Figure 1.3.1 highlights, only captures the *beginning* of an epidemic.

2. Instead, we preferred the parameter set $\gamma = 1/14$, $\beta_{11} = 0.135$, $\beta_{12} = 0.01$, $\beta_{21} = 0.04$, $\beta_{22} = 0.02$, which produced the shape

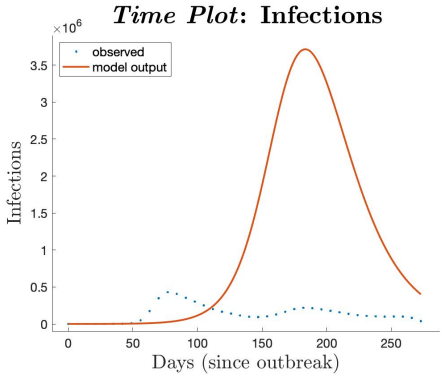



FIGURE 1.3.2: *shape-preserving overestimate*

which is skewed and an obvious overestimate, but captures a full wave of an epidemic while obeying our constraint, $\beta_{11} > \beta_{21} > \beta_{22} > \beta_{12}$.

3. Although this parameter set is a suboptimal fit to the CDC data, we suspect the discrepancies are less troublesome than they appear. One, the CDC data is known to be a large *underestimate*, because it only reports confirmed cases. Thus, it is likely that their death/infection data is "off" by some scalar multiple. Two, the CDC dataset contains two peaks (corresponding to two mini-waves of coronavirus), but our SIRV model can only describe one.

The specific datasets/code used to create the contact matrix and optimize parameters are described in Appendix A.

## 1.4 Quantifying Herd Immunity

Herd immunity is defined as the immune population (recovered plus vaccinated) passing a critical threshold, which sharply reduces COVID-19 spread and renders the nation invulnerable to future outbreaks. Mathematically, this threshold equals $1 - \frac{1}{R_0}$, where $R_0$ is the expected number of infections

created by one infected individual.

When comparing immunity in our simulations, we will use the popular estimate for the herd immunity threshold (70%), instead of $1 - \frac{1}{R_0}$ computed from our parameters.

1. Parameter variation. Achieving herd immunity means the U.S. has become (provisionally) invulnerable to all future outbreaks, not just a single wave. However, $\beta$ coefficients of infectiousness can vary from wave to wave, and since $R_0$ is a function of $\beta$ coefficients and $\gamma$, passing the herd immunity threshold in one wave *does not* guarantee immunity to the next wave.
2. The 70% immunity threshold is based on the expectation that $R_0$ can become as high as 3.33 nationwide. Once the U.S. crosses this threshold, COVID-19 will certainly die out [12].

In each infection vs time figure, we plot two herd immunity thresholds - the one computed from our data, and the 70% benchmark - but will use the latter value when discussing progress towards herd immunity.

# 2 Results

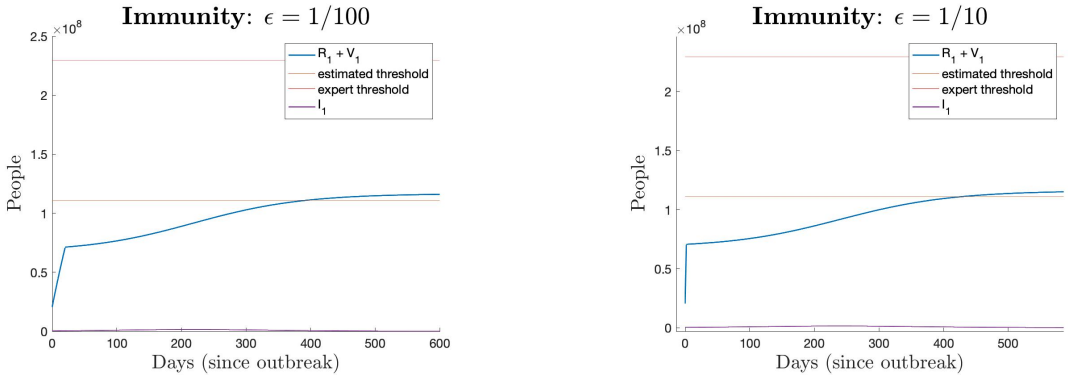
This section will:

1. Show that an age-based vaccine analysis is needed to avoid overestimating progress towards herd immunity,
2. Depict how Pfizer's limited 50 million vaccines would dampen the outbreak but fail to confer herd immunity,
3. Show that given a limited vaccine supply, the optimal strategy for reducing both infections and deaths in *both age groups* is catering to the pre-65 population.

## 2.1 The bottleneck is vaccine supply, not vaccination rate

All simulations in this section will use $\epsilon = 1/100$ for several reasons:

1. The aim of the paper is to determine *optimal vaccine* distribution, not estimate people's enthusiasm about the new vaccine. In practice, $\epsilon$ could be higher, lower, or even non-constant, but as long as older and younger people possess a similar willingness to be vaccinated [5], our qualitative conclusion - that vaccines should be prioritized for non-seniors - will remain unchanged.
2. The bottleneck on vaccine effectiveness (and its ability to confer herd immunity) is our limited supply of 50 million vaccines, *not* the vaccination rate!




(A) $\epsilon = 1/100$ (B) $\epsilon = 1/10$

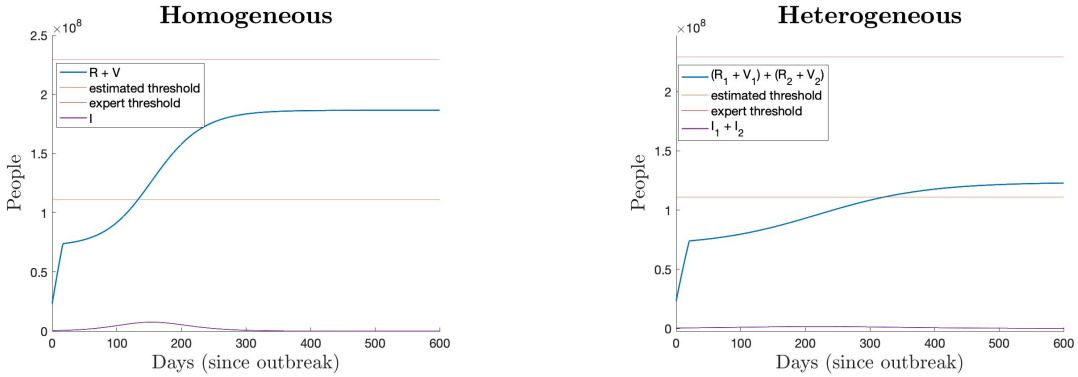
FIGURE 2.1.1: *Minimal effect from altering $\epsilon$, given limited vaccine supply*

Figure 2.1.1 depicts infected cases and our progress towards herd immunity (recovered plus vaccinated) for $\epsilon = \frac{1}{100}$ versus $\epsilon = \frac{1}{10}$. Both values of $\epsilon$ produce nearly identical outcomes despite differing by an order of magnitude! This is because in both cases, the vaccine is quickly depleted.

3. We will henceforth fix $\epsilon = 1/100$, and analyze how changing the percentage of vaccines available to each age group affects the outbreak's trajectory.

## 2.2 Homogeneous analysis misidentifies herd immunity progress

Before analyzing different vaccine distributions, we show why an age-based analysis is needed to gauge the effect of vaccines on immunity.




(A) *Homogeneous: $\beta = 0.15$*

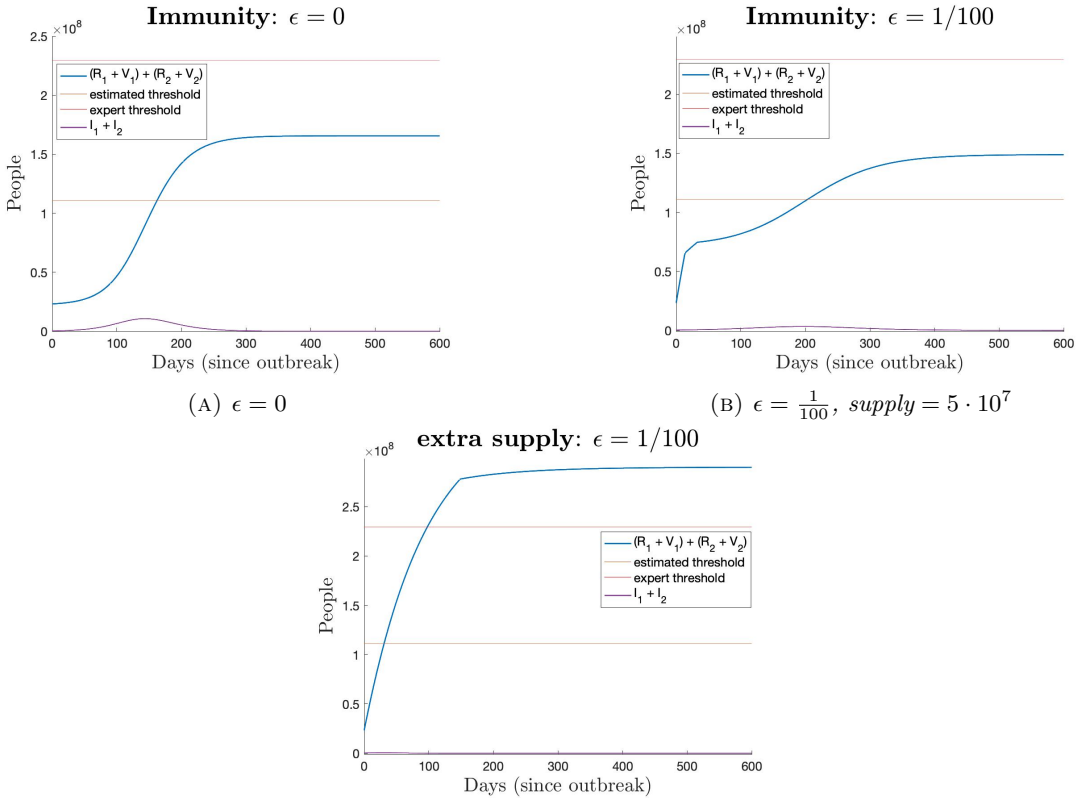
(B) *Heterogeneous: parameter estimates from Sec. 1.3*

FIGURE 2.2.1: *Homogeneous SIRV ($\beta = 0.15$) vs Age-Stratified SIRV*

Failing to distinguish between age strata overestimates the U.S.' progress towards herd immunity. Figure 2.2.1 compares infections and progress towards herd immunity predicted by a homogeneous SIRV model (specified in Figure 1.1.1) versus an age-stratified model.

1. Both models are run with $\gamma = 1/10$, $\epsilon = 1/100$, and a 50 million vaccine cap. The homogeneous model's $\beta = 0.15$ is the average of $\beta_{11} + \beta_{21}$ and $\beta_{12} + \beta_{22}$, weighted by population sizes. This $\beta$ estimate is consistent with other non-stratified SIR studies on COVID-19 [13].
2. The homogeneous model over-predicts infections per day past $t > 20$ days, and claims 7 million more people will have developed immunity 1 year into the outbreak. This would place the U.S. immune population at 80% of the herd immunity threshold. In contrast, the age-stratified model predicts that vaccines are more successful in containing the epidemic, and that the U.S. would only lie at 55% of the herd immunity threshold.
3. Thus, strategies ignoring age heterogeneity will underestimate the number of additional vaccines needed to reach herd immunity. After considering age dynamics, the heterogeneous model anticipates many more vaccines will be needed.

## 2.3 Limited vaccine supply can stop one wave, but not the pandemic




(A) $\epsilon = 0$

(B) $\epsilon = \frac{1}{100}$, $supply = 5 \cdot 10^7$

(C) $\epsilon = \frac{1}{100}$, $supply = 3 \cdot 10^8$

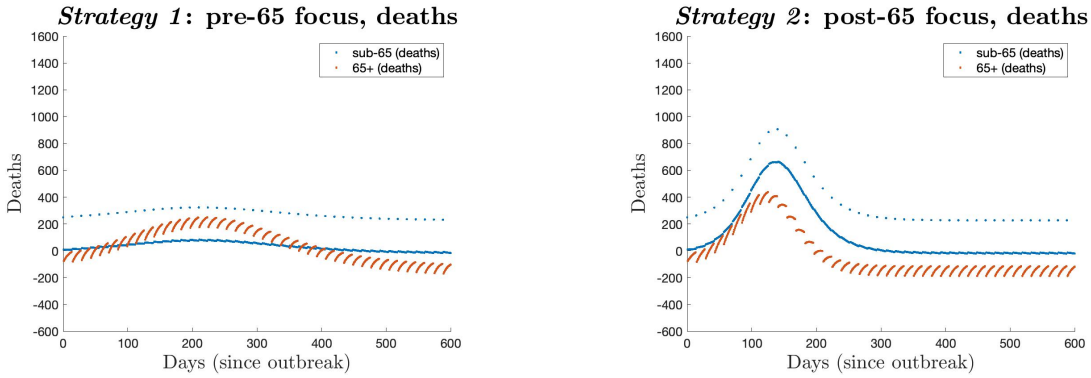
FIGURE 2.3.1: *Effects of zero, limited, and expanded vaccine supply*

A limited vaccine supply reduces disease infectiousness, but fails to induce herd immunity or shorten the length of the outbreak. Figure 2.3.1 contrasts infections/immunity under three scenarios: without vaccines, with a limited supply (50 million), and with an expanded supply (300 million doses).

1. Each simulation assumes 30% of vaccines are made exclusively available to seniors.
2. While the limited supply (50 million) initially reduces infections, a infection wave emerges once the vaccine supply is depleted, peaking at day 200. 1 year into the outbreak, the U.S. will only be at 55% of the herd immunity threshold.
3. Expanding the U.S. vaccine supply is needed to shorten the pandemic and induce herd immunity. With an additional 250 million vaccines, the infection curve becomes flat, and herd immunity can be achieved in less than 150 days.

## 2.4 Catering to the non-senior population protects all age groups

We find that allocating the limited vaccine supply to nonseniors (rather than seniors) is more effective in reducing deaths and infectious spread in both age groups.

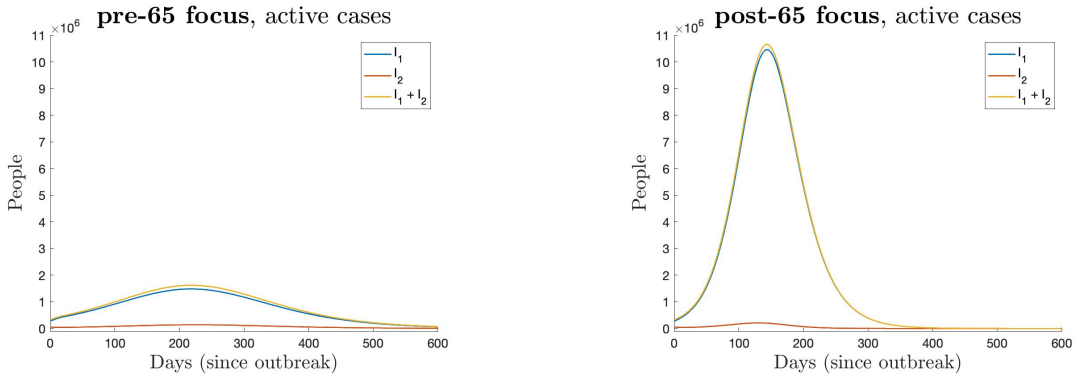



(A) *all vaccines allocated to pre-65* (B) *all vaccines allocated to post-65*

FIGURE 2.4.1: *Deaths: pre- vs post-65 vaccine allocations*

Prioritizing vaccine allocation for nonseniors, rather than seniors, substantially reduces daily deaths at every point of the outbreak. Figure 2.4.1 compares the deaths resulting from the two vaccine strategies: (1) allocating all 50 million vaccines to the pre-65 population, and (2) allocating all vaccines to the post-65 population.

1. Strategy (1) constantly suppresses daily pre-65 deaths to below 100 and post-65 deaths to below 200. In contrast, strategy (2) allows nearly 350 pre-65 deaths and 600 post-65 deaths at the outbreak's peak (days 110-140).
2. The two sets of blue dots reflect uncertainty from estimating daily death counts from active infections in Procedure 1.2.2. The upper set of dots represent overestimates, and the dense curve of lower dots represent underestimates. (The pre-65 death counts likely lie somewhere in between these two dots, though closer to the lower dots.)
3. This result is counterintuitive: since 8 out of 10 reported COVID-19 deaths have occurred among adults aged 65 years and older [14], it seems that a death-minimizing vaccine strategy should target the elderly. However, this ignores that those pre-65 are disproportionate spreaders of the disease. Since $\beta_{11} > \beta_{21} > \beta_{12} > \beta_{22}$ (and $\beta_{11}$ is a order of magnitude larger than the other coefficients), the "superspreaders" in the pre-65 population are the largest contributors to infections in *both* age groups.




(A) *all vaccines allocated to pre-65* (B) $\epsilon = \frac{1}{100}$, *supply* $= 3 \cdot 10^8$

FIGURE 2.4.2: *Active cases: pre- vs post-65 vaccine allocations*

Similarly, focusing on nonseniors is vital to reduce the number of infections. Figure 2.4.2 compares the

infections resulting from strategy (1) - allocating all 50 million vaccines to the pre-65 population - to strategy (2), which allocates all vaccines to the post-65 population.

1. Strategy (1) is superior at reducing infections in both age groups, at every point of the outbreak.
2. Strategy (2) allows up to 10.5 million total active cases on day 150. In contrast, strategy (1) never allows more than 2 million total active cases at any point.

# 3 Discussion

This study demonstrated the use of an age-stratified model to compare vaccine prioritization strategies for SARS-CoV-2. Dividing the population into two age groups (pre- and post-65) reveals that a heterogeneous population demands a higher vaccination rate to achieve herd immunity. The U.S. population is heterogeneous, not homogeneous, and thus better modeled with an age-stratified model.

Pharmaceutical companies predict that we can expect 50 million vaccines ready for distribution by the end of 2020. Given a limited initial supply, the number of vaccines made available to each age group will dramatically affect the infectiousness and total death count of a future outbreak. 50 million vaccines is unfortunately not enough to reach herd immunity and end the pandemic. The United States will need about 300 million vaccines to achieve herd immunity, which Pfizer hopes to produce by late 2021.

In the meantime, with the initial supply of vaccines, the best strategy for both minimizing infections and minimizing deaths is to reserve all vaccines for the younger population. This strategy will protect the elderly by immunizing younger people who most often spread the virus to their elders.

This study is subject to a number of limitations. First, we have assumed the vaccine is 100% effective. To date, it is unknown how effective the first vaccine will be, and effectiveness can vary from patient to patient. Further, we have only analyzed two age strata. Dividing the population into more groups, each with a smaller range, would provide more precise estimates about who to vaccinate. Lastly, we consider variation in disease severity only by age. Other factors correlate with disease outcomes, such as healthcare access, rural versus urban location, socioeconomic status, sex, and race. These are areas that further research could expand upon.

# References


[1] About COVID-19. (2020, September 1). Retrieved November 15, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/cdcresponse/about-COVID-19.html

[2] Branswell, H., & Joseph, A. (2020, March 11). WHO declares the coronavirus outbreak a pandemic. Retrieved November 15, 2020, from https://www.statnews.com/2020/03/11/who-declares-the-coronavirus-outbreak-a-pandemic/

[3] Coronavirus Cases. (2020, November 15). Retrieved November 15, 2020, from https://www.worldometers.info/coronavirus/

[4] Pfizer and BioNTech Announce Vaccine Candidate Against COVID-19 Achieved Success in First Interim Analysis from Phase 3 Study. (2020, November 9). Retrieved November 15, 2020, from https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against

[5] Mullen O'Keefe, S. (2020, November 03). One in Three Americans Would Not Get COVID-19 Vaccine. Retrieved November 15, 2020, from https://news.gallup.com/poll/317018/one-three-americans-not-covid-vaccine.aspx

[6] Provisional COVID-19 Death Counts by Sex, Age, and Week. (2020, November 12). Retrieved November 15, 2020, from https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-W/vsak-wrfu/data

[7] Levin, A., Hanage, W., Owusu-Boaitey, N., Cochran, K., Walsh, S., & Meyerowitz-Katz, G. (2020, October 31). Assessing the Age Specificity of Infection Fatality Rates for COVID-19: Systematic Review, Meta-Analysis, and Public Policy Implications. Retrieved November 15, 2020, from https://www.medrxiv.org/content/10.1101/2020.07.23.20160895v7.full

[8] Duration of Isolation and Precautions for Adults with COVID-19. (2020, October 19). Retrieved November 15, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html

[9] Prem, K., Cook, A., & Jit, M. (2017, September 12). Projecting social contact matrices in 152 countries using contact surveys and demographic data. Retrieved November 15, 2020, from https://journals.plos.org/ploscompbiol/article?id=10.1371

[10] Omori, R., Matsuyama, R., & Nakata, Y. (2020, October 06). The age distribution of mortality from novel coronavirus disease (COVID-19) suggests no large difference of susceptibility by age. Retrieved November 15, 2020, from https://www.nature.com/articles/s41598-020-73777-8

[11] Walker, M. (2020, June 25). CDC: Age Nothing But a Number for COVID-19 Risks. Retrieved November 15, 2020, from https://www.medpagetoday.com/infectiousdisease/covid19/87272

[12] Aschwanden, C. (2020, October 21). The false promise of herd immunity for COVID-19. Retrieved November 15, 2020, from https://www.nature.com/articles/d41586-020-02948-4

[13] Cooper, I., Mondal, A., & Antonopoulos, C. (2020, June 28). A SIR model assumption for the spread of COVID-19 in different communities. Retrieved November 15, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7321055/

[14] Older Adults and COVID-19. (2020, September 11). Retrieved November 15, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html

# Appendix A: Scripts and Datasets

All scripts have been added to the GitHub repository.

# Appendix B: Nondimensionalization

## 3.1 Model Rescaling

*First*, we define the following rescalings:

- $s_1 = \frac{S_1}{N_1}$, and $s_2 = \frac{S_2}{N_2}$.
- $i_1 = \frac{I_1}{N_1}$, and $i_2 = \frac{I_2}{N_2}$.
- $r_1 = \frac{R_1}{N_1}$, and $r_2 = \frac{R_2}{N_2}$.
- $v_1 = \frac{V_1}{N_1}$, and $v_2 = \frac{V_2}{N_2}$.

Applying these to our model produces:
$\dot{s}_1 = \frac{1}{N_1}(-\beta_{11} i_1 s_1 N_1 - \beta_{12} i_2 s_1 N_1 - \epsilon s_1 N_1) = -\beta_{11} i_1 s_1 - \beta_{12} i_2 s_1 - \epsilon s_1.$
$\dot{i}_1 = \frac{1}{N_1}(\beta_{11} i_1 s_1 N_1 + \beta_{12} i_2 s_1 N_1 - \gamma i_1 N_1) = \beta_{11} i_1 s_1 + \beta_{12} i_2 s_1 - \gamma i_1.$
$\dot{r}_1 = \frac{1}{N_1}(\gamma i_1 N_1) = \gamma i_1.$
$\dot{v}_1 = \frac{1}{N_1}(\epsilon s_1 N_1) = \epsilon s_1.$

$\dot{s}_2 = \frac{1}{N_2}(-\beta_{22} i_2 s_2 N_2 - \beta_{21} i_1 s_2 N_2 - \epsilon s_2 N_2) = -\beta_{22} i_2 s_2 - \beta_{21} i_1 s_2 - \epsilon s_2.$
$\dot{i}_2 = \beta_{22} i_2 s_2 + \beta_{21} i_1 s_2 - \gamma i_2.$
$\dot{r}_2 = \gamma i_2.$
$\dot{v}_2 = \epsilon s_2.$

*Second*, we rescale <u>time</u> via $t = a\tau$, where $a := \frac{1}{\gamma}$, then define $\phi := \frac{\epsilon}{\gamma}$, $R_0^{11} := \frac{\beta_{11}}{\gamma}$, $R_0^{12} := \frac{\beta_{12}}{\gamma}$, $R_0^{21} := \frac{\beta_{21}}{\gamma}$, and $R_0^{22} := \frac{\beta_{22}}{\gamma}$.
$\frac{ds_1}{d\tau} = \frac{ds_1}{dt}\frac{dt}{d\tau} = -as_1(\beta_{11} i_1 + \beta_{12} i_2 + \epsilon) = -s_1(\frac{\beta_{11}}{\gamma} i_1 + \frac{\beta_{12}}{\gamma} i_2 + \frac{\epsilon}{\gamma}) = -s_1(R_0^{11} i_1 + R_0^{12} i_2 + \phi).$
$\frac{di_1}{d\tau} = \frac{di_1}{dt}\frac{dt}{d\tau} = a[s_1(\beta_{11} i_1 + \beta_{12} i_2) - \gamma i_1] = s_1(\frac{\beta_{11}}{\gamma} i_1 + \frac{\beta_{12}}{\gamma} i_2) - i_1 = s_1(R_0^{11} i_1 + R_0^{12} i_2) - i_1.$
$\frac{dr_1}{d\tau} = \frac{dr_1}{dt}\frac{dt}{d\tau} = a\gamma i_1 = i_1.$
$\frac{dv_1}{d\tau} = \frac{dv_1}{dt}\frac{dt}{d\tau} = a\epsilon s_1 = \frac{\epsilon}{\gamma} s_1 = \phi s_1.$

$\frac{ds_2}{d\tau} = -s_2(R_0^{22} i_2 + R_0^{21} i_1 + \phi).$
$\frac{di_2}{d\tau} = s_2(R_0^{22} i_2 + R_0^{21} i_1) - i_2.$
$\frac{dr_2}{d\tau} = i_2.$
$\frac{dv_2}{d\tau} = \phi s_2.$

So our rescaled model is

$$\begin{aligned}
\frac{ds_1}{d\tau} &= -s_1(R_0^{11}i_1 + R_0^{12}i_2 + \phi) \\
\frac{di_1}{d\tau} &= s_1(R_0^{11}i_1 + R_0^{12}i_2) - i_1 \\
\frac{dr_1}{d\tau} &= i_1 \\
\frac{dv_1}{d\tau} &= \phi s_1
\end{aligned}$$

$$\begin{aligned}
\frac{ds_2}{d\tau} &= -s_2(R_0^{22}i_2 + R_0^{21}i_1 + \phi) \\
\frac{di_2}{d\tau} &= s_2(R_0^{22}i_2 + R_0^{21}i_1) - i_2 \\
\frac{dr_2}{d\tau} &= i_2 \\
\frac{dv_2}{d\tau} &= \phi s_2
\end{aligned}$$

where:

- $s_1 = \frac{S_1}{N_1}$, and $s_2 = \frac{S_2}{N_2}$.
- $i_1 = \frac{I_1}{N_1}$, and $i_2 = \frac{I_2}{N_2}$.
- $r_1 = \frac{R_1}{N_1}$, and $r_2 = \frac{R_2}{N_2}$.
- $v_1 = \frac{V_1}{N_1}$, and $v_2 = \frac{V_2}{N_2}$.
- $t = \frac{1}{\gamma}\tau$, $\phi := \frac{\epsilon}{\gamma}$, $R_0^{11} := \frac{\beta_{11}}{\gamma}$, $R_0^{12} := \frac{\beta_{12}}{\gamma}$, $R_0^{21} := \frac{\beta_{21}}{\gamma}$, and $R_0^{22} := \frac{\beta_{22}}{\gamma}$.

## 3.2 Epidemiological Meaning of Parameters

$R_0^{ij}$ retains its epidemiological meaning: the average number of people in group $i$ that can be infected by a person from group $j$. We can then approximate subpopulation $R_0$'s via $R_{01} = \frac{1}{\gamma}(\beta_{11} + \beta_{21})$, which is the average infectious period for group 1 ($\frac{1}{\gamma}$) times the sum of rates at which an infected 1 infects other 1's and 2's. (Similarly, $R_{02} = \frac{1}{\gamma}(\beta_{12} + \beta_{22})$.)