An Analysis of the 2020 DANMASK COVID-19 Study

Aushna Saleem

Professor Christian Beneš

January 23, 2022

Honors Thesis

Table of Contents

Section 1: Historical Context and Overview

Beginning in 2019, respiratory syndrome coronavirus-2 (SARS-CoV-2) has been responsible for a worldwide pandemic that has persisted into 2022. First reported in Wuhan, China, SARS-CoV-2 quickly traveled across international borders, transmissible from humans to other humans (9). While genetic sequencing has shown similarity between the disease and another variant of a coronavirus affecting bats, as well as pangolins, the origin of SARS-CoV-2 remains unknown (9). The infectious disease has introduced massive political and scientific turmoil, as society combats high levels of casualties, a shortage of medical officials, as well as lack of resources.

As of January 2022, the number of COVID-19 related deaths in the United States of America has been totalled to be more than 836,236 (5). The number of total reported COVID-19 cases in the United States of America has been totalled to be 60,164,525 as of January 2022 as well (6). Additionally, the national average of new daily COVID-19 cases has surpassed 800,000 as of January 2022 - more than at any other time since the emergence of COVID-19 (6). This recent surge in cases is due to a variant of the original strain of COVID-19, called the Omicron variant (14). The disease has been found to have several variants which have been said to cause more infection and spread faster than other variants (14).

As research continues to pursue the best means of mitigating the disease and its effects, mRNA vaccines have been introduced and distributed to prevent infection. Presently, Pfizer/BioNTech Comirnaty, AstraZeneca, Johnson and Johnson, Moderna, Sinopharm, Sinovac-Coronavac, and Bharat Biotech have developed vaccines that have been approved by the World Health Organization for emergency use (4). Of these vaccines, three have been approved for use in the United States of America and been used widely: the Pfizer/BioNTech, Moderna, and Johnson and Johnson vaccines (14). These vaccines provide strong protection

against serious illness from the disease, and research shows signs of these vaccines also being effective in preventing infected persons from passing the virus onto others (4). Both Moderna and Pfizer vaccines specifically are licensed and approved by the World Health Organization for distribution to those 12 years and older, while all of the other vaccines are approved for those 18 years and older (4). The Pfizer vaccine specifically has also been approved for distribution to children between 5 and 11 years old (11). These vaccines are still being studied through clinical trials to determine how long immunity lasts, but current data shows evidence of it lasting at least 6 months (4). This immunity can vary depending on the existence of other medical conditions (4). As of December 2021, all vaccinated individuals aged 16 and older are encouraged to get a booster shot from Pfizer or Moderna in order to sustain immunity, especially due to the emergence of new variants of the disease (11).

In a pursuit to reduce COVID-19 infection, many governments have established mask mandates, as well as other protective measures. Although the mandates have changed as the number of cases of COVID-19 change, there has been a general suggestion by the Centers for Disease Control and Prevention (CDC) in most public spaces for masks, including for individuals who are vaccinated. Masks are often mandatory in many indoor spaces, such as public transit and airplanes (19). These mandates often become more flexible as the number of cases of COVID-19 decreases significantly, or the number of vaccinated individuals increases significantly. With the implementation of these mask mandates, many studies have been conducted to research the effectiveness of protective face masks in preventing COVID-19, as well as other respiratory viruses.

Section 2: Background of DANMASK 19 Study

The World Health Organization acknowledges a lack of evidence over whether a mask protects against infection of SARS-CoV-2 (2). However, various observational studies are said to conclude that mask use could be linked to reduced risk for SARS, MERS, and SARS-CoV-2 (other respiratory diseases) (2). As a result, a variety of research has been conducted and published on SARS-CoV-2, as well as on the public guidelines for safety. To target what was considered a controversy on mask wearing benefits, a study was published in March 2021 to examine the efficacy of masks in public spaces in Denmark.

This study has been called "DANMASK 19" and was conducted through a randomized controlled trial to determine whether a recommended mask mandate outside of the home and in public reduces wearers' risk for SARS-CoV-2 infection in a community where mask wearing is uncommon and unrecommended. The authors of the study reported finding that masks were not effective in communities similar to the one studied. These findings led to widespread controversy with scientists and statisticians protesting the outcomes. This thesis aims to understand the outcomes of the DANMASK 19 study, replicate the statistical aspects of the study, and provide a model for the benefits of mask-wearing in a community similar to Denmark.

Section 3: Methods from DANMASK 19 Study

The DANMASK 19 study was conducted from April 3, 2020 to June 2, 2020, in a Danish community where mask wearing was uncommon. This community was restricted such that cafes and restaurants were closed until May 18, 2020. Participants in the study were selected based on media advertisements seeking adults 18 years or older.

To attain a power of study such that they would be able to find a meaningful difference between the test groups with probability of 80%, the authors of the DANMASK 19 study aimed to include from 4636 to 6000[1] participants. This was based on a calculation where the rate of infection of COVID-19 was estimated to be 2%, taking into account a 20% chance of participants failing to follow up with the researchers. Thus, the aim was to include 4636 - 6000 participants to limit the risk of Type II error - that is, that of inaccurately finding no difference between the mask and control group. A total of 17,258 Danish citizens responded to the advertisements, and 6024 fulfilled the eligibility criteria (respondents with positive results of COVID-19 antibody tests or prior COVID-19 infection were excluded from the study).

There were 2 groups as part of the study. The first was so called the mask group, and was the experimental group, and the second group was so called the control group, as the results of this group would be compared to the experimental group. The experimental group was advised to wear masks when outside of their home, to change their mask if outside for more than eight hours, and received 50 of the same surgical face masks for use through the experiment. The control group was not advised in this manner. Both groups however, received weekly follow-up emails encouraging participants to follow Danish COVID-19 safety guidelines.

Thus, the DANMASK 19 study was using a clear independent variable and dependent variable to compare the groups. The independent variable is the variable that the researchers are

[1] See #1 in Appendix for detailed replication of this calculation

in control of, and is manipulated between the two groups. In this case, the independent variable was whether or not masks were recommended to the participants of the group. The recommendation was at the researchers' control, as they chose to advise the mask group, and not the control group. The dependent variable, in contrast, is the value that is then measured by the researchers, and is considered to be the result of the manipulation of the independent variable. In this case, the dependent variable was development of COVID-19 infection. This infection was going to be measured as a result of the mask recommendation versus no recommendation.

3030 participants were placed into the mask recommended experimental group, while 2994 were in the control group. Due to the randomized nature of the study, different numbers of participants were in each group. While it is unclear what randomization process the researchers used, a randomized process often entails the use of a random number generator such that each participant is given a number randomly generated and is selected to be placed into either the mask or control group. Randomization is one of the necessities for good experimental design in order to prevent bias in how groups are formed for treatments.

4862 participants completed the study. Completion of study in this case would mean following up to the researchers as often as requested, filling out the allotted surveys, and not testing positive for COVID-19 antibodies at the beginning of the study.

Data was collected through follow up surveys. Four surveys throughout the course of the experiment were administered by email to collect information on SARS-CoV-2 antibody test results, adherence to the mask recommendation, development of SARS-CoV-2 symptoms, SARS-CoV-2 diagnosis by PCR test, as well as COVID-19 exposure.

The main dependent variable of the study was development of SARS-CoV-2 infection, through PCR test or hospital based diagnosis, or through presence of SARS-CoV-2 antibodies based on an antibody test result. This was referred to as the "primary outcome" throughout the

study (2). Data was also collected on whether or not participants developed other respiratory viruses throughout the study - this was referred to as the “secondary outcome” throughout the study (2).

Section 4: Reporting the Results of the DANMASK 19 Study

At the conclusion of the study, it was found that "46% of participants wore the mask as recommended, 47% predominantly as recommended, and 7% not as recommended" (2). 42 participants of the mask recommended group developed COVID-19, or antibodies - 1.8% of the entire group. 53 participants of the control group developed infection or antibodies - 2.1% of the entire group.

Based on these results, an intention-to-treat analysis was conducted. This is an important method used in randomized trials. Randomized trials, such as this one, are meant to protect against bias as much as possible. However, bias can be introduced throughout the study, despite the researchers' attempts to minimize this, and can disrupt the analysis of the study. In this case, the number of participants not following protocol, as well as some beginning the study without reporting the presence of COVID-19 antibodies could impact the analysis. Therefore, an intention-to-treat analysis can be undertaken in these cases, where researchers attempt to preserve as much as possible of the original methods and groups in the study (13). In this study, this would mean continuing to treat all of the participants as part of the original groups they were assigned, regardless of whether or not they followed the mask recommendation.

As part of this analysis, a between-group difference between the mask and control group was estimated, as well as an odds ratio. A between-group difference is the difference between the probabilities of each group developing the primary outcome, in this case, the mask and control group. An odds ratio is a measure of association between an exposure and an outcome (16). It is used to compare the relative odds of the occurrence of an outcome given exposure to the variable of interest (16). Precisely, it is the probability of developing the outcome given being exposed to the variable of interest, divided by the probability of developing the outcome without the exposure (16). In this case, the odds ratio would represent the probability of developing

COVID-19 while wearing a mask divided by the probability of developing COVID-19 without a mask. A confidence interval is often constructed along with an odds ratio and between group difference. This interval is used to measure the precision of the odds ratio and between group difference, and provide an interval within which we can say with confidence that the true value of the odds ratio and between group difference lies. A large confidence interval would mean there is low precision in the estimate calculated for the odds ratio and between group difference, while a small interval represents more precision in the estimation of the values (16).

Based on the intention-to-treat analysis, the authors reported that the between-group difference was -0.3%, with 95% confidence that the true value of the difference was between -1.2 and 0.4. The odds ratio was found to be 0.82, with 95% confidence that the true value of the ratio is between 0.54 and 1.23.

This analysis was then repeated with multiple imputation for missing data. This would be to make up for the loss of the participants that did not complete the study, which could impact the results. Missing data can impact the precision as well as the power of the study (15). Multiple imputation is a computational method used to address this issue by creating several data sets with plausible outcomes and then combining the results from the outcomes to substitute missing data (15). This allows researchers to draw conclusions of greater power than if they simply disregarded the missing data. In this analysis, the odds ratio was found to be 0.81, with 95% confidence that the true value of the ratio is between 0.53 and 1.23.

A per-protocol analysis was also conducted. A per-protocol analysis is where the researchers take the results of the study into consideration when conducting statistical tests (18). In this case, this would mean that once the researchers found that there was nonadherence to the protocol of the study, they chose to exclude those participants' results from the statistical analysis. Excluding the participants of the study who did not follow the recommendation to wear

masks, infection occurred in 40 participants, which was 1.8% of the mask group. Based on this value, the between group difference was found to be -0.4%, with 95% confidence that the true value is between -1.2 and 0.5. The odds ratio was found to be 0.84 based on this refinement, with 95% confidence that the true value is between 0.55 and 1.26.

A preplanned sensitivity analysis was also undertaken. A sensitivity analysis is a method of determining whether approaching the data from different ways will affect the outcome of the analysis. It is a method to determine "the extent to which results are affected by changes in methods, models, values of unmeasured variables, or assumptions" (18). In this sensitivity analysis, those with positive antibody results one month into the study but not reported at baseline were excluded. Based on this, 18 participants were excluded. 33 masked participants developed infection based on this analysis - 1.4% of the mask group. Based on this, the between-group difference is -0.4%, with 95% confidence that the true value lies between -1.1 and 0.4. The odds ratio was 0.77, with 95% confidence that the true value is between 0.49 and 1.22.

Two other analyses were done. These analyses were referred to as post hoc analyses. Post hoc analyses are conducted after the data has been seen, and are secondary to other statistical analyses. This would mean that these analyses were conducted after the results of the previous analyses were seen.

In the first of these post hoc analyses, only participants wearing masks exactly as per advisement were included in the mask group. This would exclude a large portion of the group, as only 46% of the mask group would be represented in this case. Of those participants, 22 participants developed COVID-19 or antibodies, which was 2.0% of the mask group. Based on this, the between-group difference was -0.2%, with 95% confidence that the true value of the

difference lies between -1.3 and 0.9; and the odds ratio was 0.93, with 95% confidence that the true value of the ratio lies between 0.56 and 1.54.

The second post hoc analysis excluded participants who did not provide antibody results at baseline. Based on this, 33 participants of the masked group developed infection - 1.7% of the group. Likewise, 44 participants of the control group developed infection - 2.1% of the group. The between-group difference of this analysis was -0.4%, with 95% confidence that the true value lies between -1.4 and 0.4, with an odds ratio of 0.80, with 95% confidence that the true value lies within 0.51 and 1.27.

Section 5: Setting Up the Hypotheses of the DANMASK 19 Study

In order to understand these results and their importance, we will use the information provided in the DANMASK 19 study to attempt to replicate the processes the authors used in the intention to treat analysis. To compute each of the between-group differences and odds ratios, the following set of hypotheses was tested in each analysis, with different conditions:

- H_0: There is no difference between the rate of COVID-19 infection in the group recommended to wear masks and the control group with no such recommendation.
- H_1: There exists a difference between the rate of COVID-19 infection in the group recommended to wear masks and the control group with no such recommendation.

The proportion of participants who developed the primary outcome in the mask group was computed, and the proportion of those who developed the primary outcome in the control group was computed as well. These ratios are estimators for the probability of developing the primary outcome for the masked group, p_Y, and the probability of developing the primary outcome for the control group, p_X, and are denoted by $\overline{Y}$ and $\overline{X}$, respectively. With this notation, the hypothesis is, formally, as follows.

- H_0: $p_X = p_Y$
- H_1: $p_X \neq p_Y$

where

p_Y = P(developing COVID-19 in the mask group)

p_X = P(developing COVID-19 in the control group)

Other important variables and parameters in the experiment are listed and used throughout the analysis, as follows.

n = total number of participants in the control group

m = total number of participants in mask group

X = number of participants of control group who developed primary outcome (1)

Y = number of participants of mask group who developed primary outcome

$\overline{X}$, an estimator of $p_X = \frac{X}{n}$

$\overline{Y}$, an estimator of $p_Y = \frac{Y}{m}$

Now that we have established these definitions, we can utilize them to define the between-group difference and its 95% confidence interval, as well as the odds ratio and its 95% confidence interval.

The between-group difference is the difference between the probability of developing the primary outcome in the mask group and control group, and is represented as follows.

$$\text{Between-group difference} = p_Y - p_X$$

Based on this definition, since $\overline{Y}$ is an estimator for p_Y, and $\overline{X}$ is an estimator for p_X, as defined in (1), then an estimator for the between group difference is $\overline{Y} - \overline{X}$.

The 95% confidence interval of the between-group difference is based on the following approximate equality[2].

$$P\left(-1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} + (\overline{Y} - \overline{X}) \leq p_Y - p_X \right.$$

$$\left. \leq 1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} + (\overline{Y} - \overline{X})\right) \approx 5\%$$

[2] See detailed explanation on between group difference confidence interval construction in #4 of Appendix

Next, the odds ratio represents the odds of developing COVID-19 given that one is part of the mask group, compared to the odds of developing COVID-19 given that one is part of the control group. The estimate for the odds ratio is based on the following approximate equalities, the first defining the odds ratio, and the second defining the estimator for the odds ratio[3].

$$\text{Odds Ratio } (OR) \ = \frac{p_Y(1-p_X)}{p_X(1-p_Y)}$$

$$\text{Estimator for } OR \ = e^{\hat{\beta}}$$

This estimator for the odds ratio is based on $\hat{\beta}$, which is defined as follows, using the estimators we defined for p_X and p_Y in (1).

$$\hat{\beta} = \ ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})$$

This estimator follows an approximate normal distribution. This is due to the fact that the distribution of the odds ratio is skewed, as it cannot be negative but can be any positive value. However, the natural log of this distribution can take any positive or negative value and yields, approximately, a normal distribution. This is represented as follows.

$$\hat{\beta} = \ ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}) \ \overset{approx.}{\sim} N(ln(OR) \ , \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}}) \ ,$$

and in this case,

n_1= number of participants of the mask recommended group developed COVID-19

n_2= total number of participants in mask group

n_3= number of participants of the control group who developed COVID-19

n_4= total number of participants in control group

[3] See detailed explanation for how this estimator was found in #6 of Appendix

The 95% confidence interval for the odds ratio is based on the following approximate equality[4].

$$P\left(e^{-1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln\left(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}\right)} \leq OR \leq e^{1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln\left(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}\right)} \right) \approx 5\%$$

We can now use these distributional facts to replicate the processes undertaken in the intention-to-treat analysis done in the study.

[4] See #6 in Appendix for detailed explanation on how the confidence interval for the odds ratio was constructed

Section 6: Replicating the Intention to Treat Analysis

A study flow diagram was included in the DANMASK study that delineated the number of participants in each group that completed the study: 2470 participants of those assigned to the control group completed the study, while 2392 participants of those assigned to the mask group completed the study. Also described in the diagram was how many participants in each group developed the primary outcome. 42 participants, or 1.8%, of the mask group developed COVID-19 or antibodies. 53 participants, or 2.1%, of the control group developed infection or antibodies. We can use these values to substitute for the variables that we defined in Section 5 to find the between group difference, the odds ratio, and their corresponding 95% confidence intervals.

We can define n = total number of participants in the control group, m = total number of participants in the mask group, x = number of individuals in the control group who developed COVID-19, y = number of individuals in the mask group who developed COVID-19.

$$n = 2470$$
$$m = 2392 \quad (1)$$
$$x = 53$$
$$y = 42$$

Based on these values, we can determine $\bar{x} = \frac{x}{n}$, and $\bar{y} = \frac{y}{m}$.

$$\bar{x} = \frac{53}{2470} = 2.145749\% \approx 2.1\%$$
$$\bar{y} = \frac{42}{2392} = 1.755853 \approx 1.8\% \quad (2)$$

Therefore, the estimated between group difference is $\bar{y} - \bar{x} = 1.8\% - 2.1\% = -0.3\%$

Based on this estimate for the between group difference, we can compute the 95% confidence interval which contains the true value of the between group difference.

The 95% confidence interval was defined by the following approximate equality in Section 5.

$$P(-\ 1.96\ *\sqrt{\frac{\overline{Y}(1-\overline{Y})}{m}+\frac{\overline{X}(1-\overline{X})}{n}}\ +(\overline{Y}-\ \overline{X})\leq\ p_Y-p_X$$

$$\leq 1.96\ *\sqrt{\frac{\overline{Y}(1-\overline{Y})}{m}+\frac{\overline{X}(1-\overline{X})}{n}}\ +(\overline{Y}-\ \overline{X}))\approx\ 5\%$$

We can substitute the values that we found for m, n, $\overline{x}$ and $\overline{y}$ from (1) and (2) into this confidence interval and solve for the upper and lower bounds of this interval. The lower bound is the left hand side of the inequality, while the upper bound is the right hand side of the equality. Substituting the values from (1) and (2), we find

$$P(-\ 1.96\ *\sqrt{\frac{\frac{42}{2392}(1-\frac{42}{2392})}{2392}+\frac{\frac{53}{2470}(1-\frac{53}{2470})}{2470}}\ +(\frac{42}{2392}-\ \frac{53}{2470})\leq$$

$$p_X-p_Y\leq 1.96\ *\sqrt{\frac{\frac{42}{2392}(1-\frac{42}{2392})}{2392}+\frac{\frac{53}{2470}(1-\frac{53}{2470})}{2470}}\ +(\frac{42}{2392}-\ \frac{53}{2470}))\approx\ 5\%$$

Now we will calculate the lower bound and upper bound of the equality[5].

Lower bound = $-\ 1.96\ *\sqrt{\frac{\frac{42}{2392}(1-\frac{42}{2392})}{2392}+\frac{\frac{53}{2470}(1-\frac{53}{2470})}{2470}}\ +(\frac{42}{2392}-\ \frac{53}{2470})\approx$ -1.2%

Upper bound = $1.96\ *\sqrt{\frac{\frac{42}{2392}(1-\frac{42}{2392})}{2392}+\frac{\frac{53}{2470}(1-\frac{53}{2470})}{2470}}\ +(\frac{42}{2392}-\ \frac{53}{2470})\approx 0.4\%$

The 95% confidence interval of the between group difference is then [-1.2, 0.4]. This is the same interval as the one found in the DANMASK 19 study.

Thus, we can say with 95% confidence that the true between-groups difference is contained in the interval [-1.2, 0.4]. It is important to note that 0 is contained within this confidence interval, which means that we can say with 95% confidence that there is a possibility that the true between group difference between the mask and control group may be zero, and there may be no significant difference between the two groups. In the context of the study, this

[5] See #5 in Appendix for code used to find the upper and lower bound

means that we can say with 95% confidence that there may be no difference in the occurrence of the primary outcome (developing COVID-19 or its antibodies) between the group recommended to wear masks and the control group.

We can understand this between group difference further by replicating the hypothesis test conducted and finding the p-value found in the study. This p-value would provide the probability of developing as extreme of a between group difference found in the intention-to-treat analysis or more, and would further address the question of whether there is a difference between the mask group and control group in this intention to treat analysis. The test was done at the 5% significance level, so we will be rejecting our null hypothesis if the value of p is less than or equal to 0.05.

As reported in the results of the DANMASK study, the calculated p-value was p=0.38. We can test the hypotheses presented in the beginning of this section by using the following principle.

$$\frac{\frac{Y}{m}-\frac{X}{n}-(p_Y-p_X)}{\sqrt{\frac{p_Y(1-p_Y)}{m}+\frac{p_X(1-p_X)}{n}}} \overset{approx.}{\sim} N(0,1)$$

Since we are testing H_0: $p_X = p_Y$, then under H_0, using p as a common parameter for p_X and p_Y, we can say that $\frac{\frac{Y}{m}-\frac{X}{n}}{\sqrt{\frac{p\,(1-p\,)}{m}+\frac{p\,(1-p\,)}{n}}} \overset{approx.}{\sim} N(0,1)$.

We can estimate the denominator of this fraction as $\frac{1}{m+n}\sqrt{(Y+X)(m+n-(Y+X))(\frac{1}{m}+\frac{1}{n})}$ by approximating p as $\frac{X+Y}{n+m}$ [6].

Using this denominator, we can then compute the p-value as follows.

[6] See #7 in Appendix for details on how this denominator was estimated

P(Something as extreme or more than what we observed would happen under the assumption H_0)

$$= 2 * P(\frac{Y}{m} - \frac{X}{n} < \frac{42}{2392} - \frac{53}{2470} \mid H_0)$$

$$= 2 * P(\frac{\frac{Y}{m} - \frac{X}{n}}{\frac{1}{m+n}\sqrt{(Y+X)(m+n-(Y+X))(\frac{1}{m}+\frac{1}{n})}} < \frac{\frac{42}{2392} - \frac{53}{2470}}{\frac{1}{m+n}\sqrt{(Y+X)(m+n-(Y+X))(\frac{1}{m}+\frac{1}{n})}} \mid H_0)$$

Substituting n, m, x, and y as defined in (1),

$$= 2 * P(\frac{\frac{Y}{m} - \frac{X}{n}}{\frac{1}{m+n}\sqrt{(Y+X)(m+n-(Y+X))(\frac{1}{m}+\frac{1}{n})}} < \frac{\frac{42}{2392} - \frac{53}{2470}}{\frac{1}{2392+2470}\sqrt{(42+53)(2392+2470-(42+53))(\frac{1}{2392}+\frac{1}{2470})}} \mid H_0)$$

We can then approximate the p-value as,

$$2 * P(Z \leq -0.9819765) \approx 0.3261114$$

Thus, the p-value that we found when using the data given in the study is approximately 0.33. This is slightly less than the value of p that the authors of the DANMASK study reported, which was p=0.38. This difference may be due to the authors using a different method to calculate the p-value, or due to using different counts or estimations during the calculation. However, in both cases, there is no reason to reject the null hypothesis. In our case, we found that the probability of a result as extreme as the one found in this study is approximately 0.33. Since our p-value is greater than our level of significance, 0.05, we do not reject the null hypothesis. This means that we cannot reject the possibility of there being no difference in the probability of developing the primary outcome between the mask and control groups in this study.

We can now replicate the process used to estimate the odds ratio and its 95% confidence interval in the intention-to-treat analysis.

As we defined in Section 5, an estimator for the odds ratio is $e^{\hat{\beta}}$, where $\hat{\beta} = ln(\frac{\bar{Y}(1-\bar{X})}{\bar{X}(1-\bar{Y})})$.

We can find the value of the estimate for the odds ratio in the intention-to-treat analysis using the values of $\bar{x}$ and $\bar{y}$ from (2) that we defined before.

$$e^{\beta_e} = \frac{(\frac{42}{2392})(1-\frac{53}{2470})}{(\frac{53}{2470})(1-\frac{42}{2392})} \approx 0.82 \quad (3)$$

Using this estimate for the odds ratio, we can now compute the 95% confidence interval which contains the actual value of the odds ratio. We can recall that the number of individuals in the mask group who developed COVID-19 was 42, and the total number of people in the mask group was 2392. Likewise, there were 53 participants in the control group who developed COVID-19, from a total of 2470 control participants. Then, n_1=number of participants who developed primary outcome in mask group, n_2= number of participants in mask group, n_3= number of participants who developed primary outcome in control group, and n_4= number of participants in the control group.

$$n_1 = 42 \qquad n_2 = 2392 \qquad n_3 = 53 \qquad n_4 = 2470 \quad (4)$$

We can use these values, as well as the values for $\bar{x}$ and $\bar{y}$ from (2), in the 95% confidence interval for the odds ratio, as defined in Section 5.

$$P\left(e^{-1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}} \leq OR \leq e^{1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}} \right) \approx 5\%$$

The upper bound of this confidence interval is the right hand side of the inequality, or $e^{1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}}$. The lower bound of this confidence interval is the left hand side of this inequality, or $e^{-1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}}$. We can substitute the variables that we have defined, (2), (3), and (4), to find the upper and lower bounds of the confidence interval as follows.

Lower bound of 95% CI:

$$e^{\left[\ln\left(\frac{\left(\frac{42}{2392}\right)\left(1-\frac{53}{2470}\right)}{\left(\frac{53}{2470}\right)\left(1-\frac{42}{2392}\right)}\right) - 1.96\ \mathrm{sqrt}\left(\frac{1}{42}+\frac{1}{2392}+\frac{1}{53}+\frac{1}{2470}\right)\right]} \approx 0.5415593$$

Upper bound of 95% CI:

$$e^{\left[\ln\left(\frac{\left(\frac{42}{2392}\right)\left(1-\frac{53}{2470}\right)}{\left(\frac{53}{2470}\right)\left(1-\frac{42}{2392}\right)}\right) + 1.96\ \mathrm{sqrt}\left(\frac{1}{42}+\frac{1}{2392}+\frac{1}{53}+\frac{1}{2470}\right)\right]} \approx 1.226644$$

The 95% confidence interval of the odds ratio is then approximately [0.54, 1.23]. This is quite similar to the value found in the study, which was [0.53, 1.23].

Thus, we can say with 95% confidence that the true odds ratio of the intention-to-treat analysis is contained within the interval [0.54, 1.23]. In the context of this study, we can say with 95% confidence that the true value of the odds of the primary outcome occurring in the masked group compared to the control group is contained within the interval [0.54, 1.23]. Since 1 is contained in this interval, it is meaningful to point out that we can say with 95% confidence that there is a possibility that both groups, based on this study, are at the same risk for developing COVID-19 or its antibodies.

Section 7: Interpreting the DANMASK 19 Study Analyses

The same methods of computation were used in the other analyses that the authors of the DANMASK 19 study conducted, and the results are reported in Section 4. We will interpret all of the results of the different analyses at once instead of consecutive replications of the same methods. We will be doing this because the results and conclusions for each analysis are very similar. The only analysis we will not be interpreting is the second analysis. In the second analysis, which included multiple imputation for missing data, we were unable to replicate the methods of the study since the number of individuals added or excluded as missing data from the original participant pool was not given in the study. Thus, this analysis was disregarded.

In all of the other analyses however, as reported in Section 4, it is important to point out that the confidence interval for the between group difference, regardless of how the analysis was conducted, contains 0. This is significant, since 0 being contained in the confidence interval, as we discussed in Section 6, provides evidence that there is no significant difference in the primary outcome between the groups regardless of how the groups are isolated. According to the DANMASK 19 researchers, this could mean there is no real difference in the development of COVID-19 or its antibodies for the mask group and the control group.

Likewise, when interpreting the odds ratios found in the study, it is important to point out that the odds ratios, in each analysis, varied from 0.77 to 0.93. In this case, an odds ratio close to 1 shows that the probability of developing COVID-19 between the mask and control group is close to the same. Additionally, all of the odds ratio confidence intervals, in each analysis, contain 1. So, regardless of which subsets of the mask and control group are taken, we can say with 95% confidence that there is evidence that the odds of developing the primary outcome does not differ between the mask and control group. This would provide evidence that mask group participants are as easily infected as the control group participants.

These two points are important, and were used by the authors of the DANMASK 19 to conclude that “a recommendation to wear a surgical mask did not reduce, at conventional levels of significance, incidence of SARS-COV-2 infection compared to mask recommendation” (2). The authors claim that no statistically different significance in infection was observed, but the confidence intervals do show that there is a possibility of a “46% reduction to 23% increase in infection amongst mask wearers" (2). Although, the authors of the DANMASK 19 study did state that these conclusions should not be used to say that there is no benefit of everyone in a community wearing masks, as during the period of the study, masks were not recommended in community spaces, or outside of hospitals and settings. This meant that a majority of the individuals that the participants were exposed to, regardless of mask or control group, were not wearing masks. Thus, there is a discrepancy in this study and its applicability to the general populace outside of Denmark, such as in the United States and United Kingdom, where there are active mask mandates, and most people are encouraged to wear masks in public spaces.

There are also gaps in the study itself. Having to conduct post hoc analyses due to there being noncompliance with the directions of the study, and having to simulate data, as in the case of multiple imputation, to make up for loss of data, does affect the reliability of conclusions. Likewise, when attempting to reproduce the statistical methods of the other analyses, there were several instances where the authors of the text did not explain how many participants were excluded from analysis, such as in the case of the multiple imputation analysis. The conclusions of this study do point towards there being no significant difference in COVID-19 infection of masked versus non-masked individuals within communities that do not have strict guidelines for mask wearing in public spaces. However, the results of this study were widely critiqued, with many scientific experts re-interpreting the results of the study and the conclusions.

Section 8: Criticism of the DANMASK 19 Study

It is important to point out that the DANMASK 19 study described in Section 2 was controversial, and many scientists and statisticians disagreed with the conclusions drawn by the authors. Several experts critiqued the study in open letters, exposing the gaps in the experimental design and analysis.

Thomas R. Frieden, an American infectious disease and public health physician, and former director of the CDC, penned such a letter (8). Frieden is a fervent supporter of face masks reducing the risk of COVID-19 infection, and begins his letter on this note. Dr. Frieden points out that regardless of the results, the specificity of the DANMASK 19 study results in it having low statistical power and generalizability, meaning that the results are not applicable to a greater population. For one, since the study was limited only to a small population within Denmark, a community with a relatively low transmission rate of COVID-19, this community was a poor representative for most other nations, especially the United Kingdom and United States of America, where the transmission rates are much higher. While the authors of the DANMASK 19 study did not claim to model a community with a higher transmission rate, the conclusions of the study were general - claiming that masks lack effectiveness in communities without strict mask guidelines. According to Frieden et al, this conclusion is considerably limited when considering larger countries with higher transmission rates.

Additionally, Frieden says that the DANMASK 19 study failed to consider the differences between occupations of participants, and how much time they routinely spent out of their homes. This presents an issue, as some participants may have had more exposure to COVID-19 than others, and there was no method of control for this in either group. This lack of control meant that there was no way of assessing how masks benefit individuals in certain conditions more than others, limiting the conclusions of the study. Many participants who wore

their masks as recommended may have been in high risk environments, and were protected against developing COVID-19. The lack of nuance in the experiment set-up impacts its usefulness, as some individuals are more at risk for infection than others, and this is a factor that should have been accounted for.

Dr. Frieden also makes note of a limitation within the experimental design of the study. The authors of the study used self reported antibody test results or PCR COVID-19 test results to determine the primary outcome, which was the dependent variable: developing COVID-19 or its antibodies. Of the 95 participants in the study who developed COVID-19, 84% made this report through an antibody test result. Dr. Frieden points out that antibody tests are a faulty method of collecting COVID-19 results, as the tests can have false positives. Specifically, the manufacturers of the antibody test used in the study reported approximately a 98.5% effectiveness of the test, meaning there is about a 1.5% probability of a false positive on the antibody test. While this seems to be a low probability, Dr. Frieden states that "Bayes' law implies that all of the antibody-positive results in both intervention and control groups could have been false positives." (8). He states this as a possibility as there was a low prevalence of infection throughout the study. Since there was such a small proportion of participants who reported a positive COVID-19 result, many of whom did so with an antibody test, there is a possibility that if the prevalence of infection is lower, which Frieden did say was the case in Denmark, then all of those positive results may have been due to false positives[7]. If the prevalence of infection in Denmark was 1.5% and not 2% as reported by the authors of the study, then there is a possibility that the 95 positive antibody tests were all false positives. This shows that using antibody tests to quantify the primary outcome was flawed, and could have led to

[7] See #8 in Appendix for a detailed explanation for this statement.

flawed conclusions. Using antibody tests as a means of collecting data may have generated false positive results, skewing the data and results.

Similar concerns were voiced by Noah Haber, Sarah E. Weiten, and Emily R. Smith, in the Danish Medical Journal (10). Haber and Weiten are affiliated with the Meta-Research Innovation Center at Stanford University, while Smith is affiliated with the Milken Institute School of Public Health at The George Washington University. The authors say that the design of the study limits its conclusions, making this study a poor representative for global mask standards, like Frieden. They point out a severe underpowering of the study due to a flawed assumption that masks halve the risk of infection by COVID-19. The authors of the DANMASK 19 study cited that between 4636 - 6000 participants would be needed to accurately power the study, assuming that there is an incidence of SARS-CoV-2 infection of at least 2% and that wearing a face mask halves the risk for infection. However, Haber, Weiten, and Smith point out that assuming that face masks halve the risk of infection is unreasonably high, since this is a combination of mask compliance (which was low in the community of the study) and the protectiveness of masks.

To understand Haber et al's point, we computed what the sample size of the study should be if masks do not reduce infection by a half, but instead less - a quarter. In doing this, we found a discerning difference in sample size. If the difference between risk of infection between mask wearers and non-mask wearers is not 0.01, as would be the case if mask wearing halve the risk of infection, and is instead 0.005, the sample size should be at least 24,615 to achieve a power of 0.8 - much higher than the original sample size of the study.[8] Sample size plays a role in accurately powering a study, and using a sample size too small due to flawed assumptions would indeed underpower the study.

[8] See #2 in Appendix for this calculation

Likewise, Haber, Weiten, and Smith state that since there was such low transmission of COVID-19 in Denmark during this time, the proposed 2% infection incidence over the course of the study is inflated. Haber et al say that the combination of the inflated transmission rate and mask risk reduction exacerbates the underpowering of the DANMASK 19 study. We found that if the transmission rate of COVID-19 was indeed lower in Denmark, not 2%, but 1%, and the difference in risk of infection between mask wearers and non wearers was 0.005, not 0.01, then the original sample size necessary to obtain 80% power in the study would lead to a much lower power in this case - about 50.4%. [9] This shows that the sample size used in this study puts its conclusions at serious risk if the transmission rate of COVID-19 was lower and mask wearing does not halve risk of infection.

Haber et al bring up another flaw with the collection of results as well. Haber, Weiten, and Smith point out that the participants of the study were only followed for 30 days. Based on the time it takes for COVID-19 to develop and show symptoms, this is too short of a period. The authors say that while approximately half of those infected will report symptoms and test positive for COVID-19 within 5 days of infection, others will not develop symptoms or test positive for two weeks or more than after exposure. This would mean that 30 days is too short of a period to cover a large majority of the infections that could result throughout the study.

Another flaw they discuss in the experimental design is the distribution of masks. Haber et al claim that providing the participants of the study in the dependent group with surgical face masks, while not providing the control group with masks does not measure if masks are effective. Instead, since some of the members of the mask group failed to wear masks as recommended, this study more so investigates if providing masks induces mask wearing. Haber et al claim that the design of the study does not answer the proposed question of the study. They

[9] See #3 in Appendix for explanation on how this value for power was found

also point out that the study only focuses on the protective impact of masks - whether masks protect the wearer. The study does not acknowledge how wearing a mask impacts transmission of COVID-19 from one infected person to the next, which would impact how effective masks are. The authors end their critique by starting, “this study poses a serious risk of mistranslation, in part due to misleading statements about what the study actually measures in the protocol paper and trial registration” (10).

Kamran Abbasi, a physician and executive director of theBMJ, a scientific journal, had a different take upon analyzing the study (1). Abbasi claims that results of the study are being misinterpreted. Abbasi says that the results of the study are “inconclusive rather than negative”, meaning that the study was not able to make any conclusion about the effectiveness of masks (1). He even says that the results are more likely to show a benefit to mask wearers than no benefit, with the same logic as Haber et al, saying that the authors did not examine how masks affected transmission of COVID-19 from a mask wearer. In saying this, Abbasi means that if the authors of the DANMASK 19 study had considered how masks protected mask wearers from the virus, they were likely to have found a benefit.

The New York Times also published an article containing the opinions of various experts on the Denmark study (12). Susan Ellenberg, a biostatistician at the University of Pennsylvania Perelman School of Medicine, had the same critique as Abbasi - the results of the DANMASK 19 study, although not statistically significant, point towards a benefit towards mask wearers, if any result. Dr. Elizabeth Halloran, a statistician at Fred Hutchinson Cancer Research Center in Seattle, made a similar point as Dr. Frieden, noting that the authors of the study neglected to focus on how much exposure to COVID-19 each of the participants had through the course of the study. She cited individuals working in high risk spaces, such as hospitals, as those who depend on masks due to the nature of their occupation, and claimed they would support masks.

Thus, the DANMASK 19 study has faced harsh backlash since its publication, with physicians and statisticians alike making a variety of criticism. Freiden et al, Haber et al, and Halloran cite that the authors of the study had flawed experimental design, with lack of control over COVID-19 exposure, and a flawed method of collecting results. These errors increase the variance and standard error within the study, reducing the power of the results (9). They make a main criticism that the study's methodology was flawed. Abbasi and Ellenberg bring up concerns of the study's results being misinterpreted, making misleading claims.

In fact, there was so much critique before the study's publication, that some journals refused to publish the article, for fear of the controversial results (1). The results of the study are therefore limited, due to the methods and lack of adherence to protocol within the study. This opens the door to an analysis of the spread of COVID-19 in a community where masks are mandatory in public spaces for everyone. The analysis presented in the next section in particular, would be another approach to investigate and report on the benefits of mask-wearing.

Section 9: An SIR Model for COVID-19

Noting that the DANMASK 19 study was limited in its applicability, we can attempt to replicate the conditions of this study to consider a new model for the spread of COVID-19 in Denmark. The DANMASK 19 study did not consider the possibility of everyone in the community wearing a mask. Addressing this issue will be the goal of considering a model for the spread of COVID-19 in Denmark. We will be using the so called SIR model to model how a population the size of Denmark wearing a mask would impact the spread of COVID-19 through the population.

The SIR model is an epidemic model, and should therefore be an appropriate model for COVID-19. The model has been used to describe how the number of infected people evolves as disease spreads. The model divides a population into 3 categories: susceptible - able to be infected by the disease; infected - currently have the disease; and removed - have been infected and cannot be infected again. The model assumes that no one enters or leaves the community, the population size remains constant, and there is no contact outside the community. While these assumptions may not seem to be realistic, we can make them for the population we are studying, given the quarantine advisory in Denmark during the course of the DANMASK 19 study. As stated in the DANMASK 19 study, during the course of this experiment, quarantining individuals with COVID-19 was recommended in Denmark, and social distancing was encouraged by the Danish government. Additionally, at the time of the study, only a small proportion of the population was allowed to enter or leave Denmark. Thus, the assumptions necessary for the SIR model were being met, for the most part, in this study.

We will now give a detailed description of the SIR model using parameters inspired by COVID-19 situation in Denmark at the time of the DANMASK 19 study.

Firstly, there are 3 categories that all members of the population being studied are placed in.

- The first category is those members of the population that are Susceptible to COVID-19, denoted by (S) - (able to catch COVID-19)
- The second category is those members of the population that are Infected by COVID-19, denoted by (I) - (currently has the COVID-19 and can spread COVID-19)
- The third category is those members of the population that are considered Removed from the system denoted by (R) - (they have already had COVID-19 and will not get it again, and this includes the possibility of death).

Next, it is important to state that in the SIR model, initially, every person is either susceptible (S) or infected (I). We also assume that once someone gets COVID-19 through the course of this simulation, they cannot get COVID-19 again. This is somewhat realistic considering COVID-19 antibodies last and can produce immunity for a period of time, currently estimated at about 3 to 61 months (3).

Based on these definitions, the SIR model considers 3 variables:

- S(n): number of people in the population susceptible after period n
- I(n): number of people infected after period n
- R(n): number of people removed after period n

These functions represent the counts of the number of people in each group as a function of time. We will quantify the number of people in each group every week in our analysis, so n is measured in weeks.

To model the evolution of these functions for COVID-19, we must determine an average length of disease. Over this period of time, a person is deemed infected and can infect others. In

the case of COVID-19, a person is considered infectious for 10 days after first developing symptoms, which in weekly units is 10/7 weeks.

We will also be considering a set of functions that measure fractions of the total population in each of the 3 categories. In this case, N represents the total number of individuals in the system. The set of functions is as follows.

$$s(n)=S(n)/N$$

$$i(n)=I(n)/N$$

$$r(n)=R(n)/N$$

It is important to point out that $s(n)+i(n)+r(n)=1$ at all times.

An important assumption we will make is that each infected individual has a fixed proportion, **b**, of susceptible individuals per week that will be infected upon interacting with the infected individual. This proportion is the same for each infected individual. On average, each infected individual generates **b*S(n)** new infected individuals per week. We will also assume that a fixed fraction **k** of the infected group will recover during any given week. In our SIR models for COVID-19, we will be manipulating the value of **b** to study how the number of infected individuals in Denmark would change over time if individuals were more or less infectious. We will also assume that wearing a mask makes individuals less infectious. The value of **k** will remain the same throughout the study, since it is infection dependent, meaning the value does not vary since we are only considering the original strain of COVID-19. If it takes 10/7 weeks for individuals to recover, then approximately 7/10 individuals of the infected group will recover each week and be added to the removed group. Thus, **k** will be equal to 7/10 during the course of our SIR models. Regarding this, the number of removed people in the population, over time, denoted by n, measured in weekly units, would be represented by the following equation.

$$R(n+1) = R(n) + k\, I(n) \qquad (1)$$

There are two effects contributing to the evolution of the number of infected people over time. The number of infected individuals is decreased by the number of people removed each week: **k* I(n)**. The number of infected individuals is increased each week by the number of susceptible people who come into contact with infected people and catch the disease: **b*S(n)*I(n)**. Based on this, we can define an equation for the number of infected individuals over time, denoted by n, measured in weeks.

$$I(n+1) = I(n) + b*S(n)*I(n) - k\ I(n) \quad (2)$$

The number of susceptible people per week decreases only by the number of people who become infected. The number of susceptible people at every time n, measured in weeks, can be represented by the following equation based on this.

$$S(n+1) = S(n) - b*S(n)*I(n) \quad (3)$$

Now we have built a system of equations in (1), (2), and (3), representing the number of removed, infected, and susceptible people per week that can be manipulated. We can substitute the value of **k** in to write the system of equations as follows.

$$R(n+1) = R(n) + (7/10)\ I(n) \quad (4)$$

$$I(n+1) = I(n) + b*S(n)*I(n) - (7/10)\ I(n) \quad (5)$$

$$S(n+1) = S(n) - b*S(n)*I(n) \quad (6)$$

We will be using this system of equations to examine 3 different cases, with different values of b, to model the evolution of COVID-19 in Denmark for different infectivity rates. In each case, we will be running simulations in order to model the long term behavior of the system, and understand how a differing infectivity due to masks would affect a population the size of Denmark.

Case 1:

The first case we will examine will be the one presented in the DANMASK 19 study. In the study, it was assumed that the number of infected people in Denmark remained constant. This would mean that Δi was 0. We can use that information to determine the initial value of b in the context of the study, using equation (5) since the value of infected individuals should be the same between the start time and one week into the study if the change in infection rate is zero. To do ths, we must first determine the number of infected, removed, and susceptible individuals at the beginning of the study.

The DANMASK 19 study was conducted beginning in April of 2020. On April 1, 2020 the number of COVID-19 cases in Denmark was 282. One week before the study began, March 25, the number of infected individuals was 151. The total number of COVID-19 cases up to March 25 was 1485. Thus, at the start of the study, the number of removed cases would be R(0)=1485 - 282 = 1203, based on equation (4). The population of Denmark at this time was 5.831 million. If the number of infected individuals remains constant, then I(0) = I(1) = 151. The number of susceptible individuals at the start of the study would then be the population of Denmark minus the removed and infected at the start of the study. We can determine this value using equation (6).

$$S(0) = (5.831 * 10^6) - 151 - 1203 = 5829646$$

Now that we have the number of removed, infected and susceptible people during the start time of the study, we can list them.

$$R(0) = 1203$$
$$I(0) = I(1) = 151 \quad (7)$$
$$S(0) = 5829646$$

We can use these values to find the value of b in this case. We will substitute the number of infected and susceptible people at time 0, from (7), the beginning of the DANMASK study, into equation (5) and determine the value of b during the course of the study.

$$I(1) = 151 = I(0) - (7/10)*I(0) + b*S(0)*I(0)$$

$$= 151 = 151 - (7/10)*(151)+b*(5829646)*(151)$$

$$b = 1.200759 * 10^{-7} \quad (8)$$

With this value of b, we can now simulate the SIR model for the case represented in the study, applied to a population the size of Denmark. We can simulate how many removed, susceptible, and infected individuals there are weekly in this population, until a long term pattern of the system is shown. There are several possible long-term patterns that can be observed. They are numbered below.

1. The number of infected individuals decreases to zero, effectively eliminating the virus.
2. The number of susceptible individuals falls to zero, meaning everyone in the population has been exposed to the disease.
3. The number of removed individuals reaches the total number of people in the population, meaning that the disease has taken over the population.

The following system of equations was used to model the evolution of the system, substituting (8) for b, into equations (4), (5), and (6).

$$R(n+1) = R(n) + (7/10)\, I(n)$$

$$I(n+1) = I(n) + (1.200759 * 10^{-7})*S(n)*I(n) - (7/10)\, I(n) \quad (9)$$

$$S(n+1) = S(n) - (1.200759 * 10^{-7})*S(n)*I(n)$$

The starting values of the system are listed again below, as stated in (7).

$$R(0) = 1203,\ I(0) = I(1) = 151,\ S(0) = 5829646$$

To understand its evolution and long term behavior, the system in (9) was simulated for a number of weeks, and graphs of R(n), S(n), and I(n) were created and are provided[10].

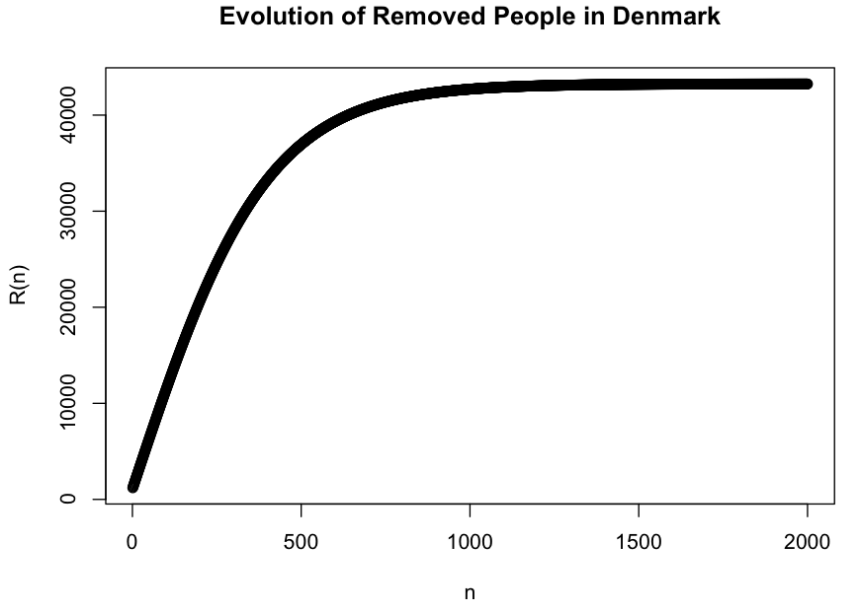

Figure 1: A graph of the number of removed people weekly in the system. The x-axis represents n, time, in weeks, and the y-axis represents, R(n), the number of removed people in Denmark. By the 1000th week, the number of removed people has stabilized at 40000

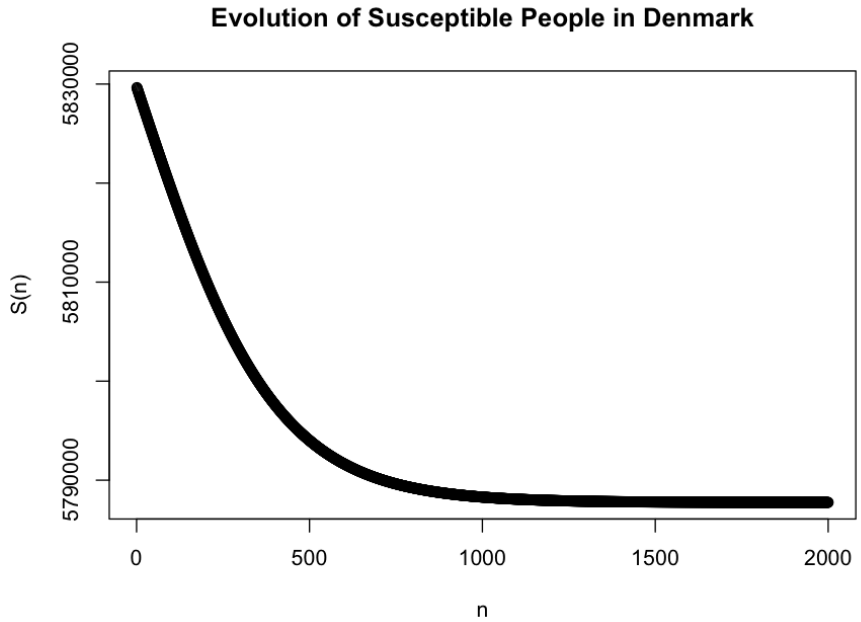

Figure 2: An representation of the number of susceptible people weekly in the system. The x-axis represents n, the number of weeks, and the y-axis, S(n), represents the number of susceptible people in the system. By the 1000th week, the number of susceptible people has stabilized at 5790000.

[10] See #9 in Appendix for code used in R to create graphs

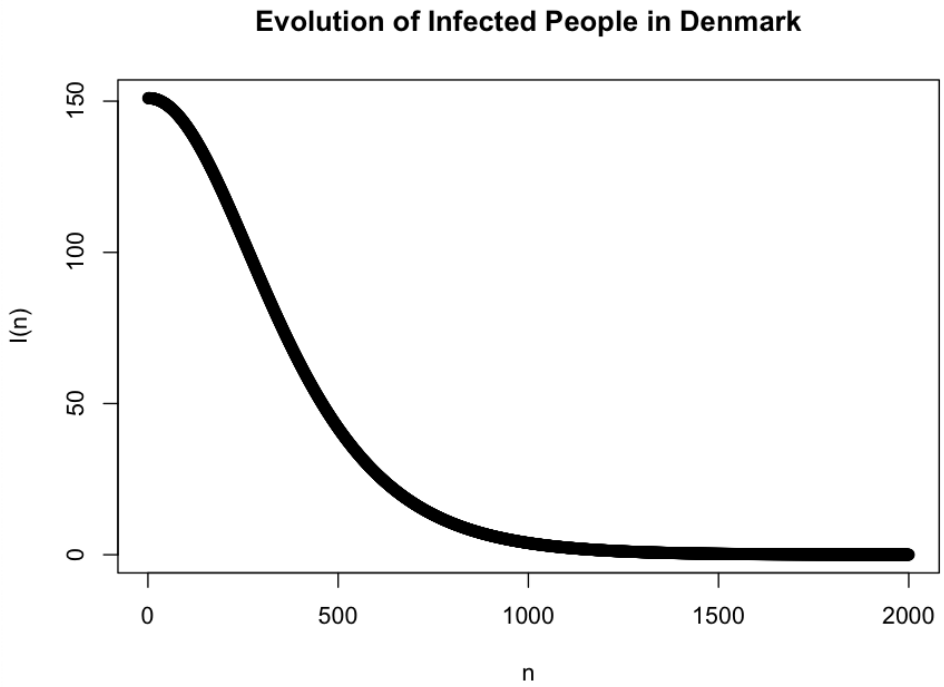

Figure 3: Graphical representation of the evolution of the number of infected people weekly in the system. The x-axis represents n, the number of weeks, and the y-axis represents I(n), the number of infected people in the system. By the 1200th week, the number of infected people has stabilized, and gone to zero.

From Figures 1, 2, and 3, we can see that the number of removed, susceptible, and infected individuals plateau by the 2000th week. Studying the graphs of the removed, infected, and susceptible individuals for every week, we also find that after the 1409th week, there are no infected individuals left after this week. There will be no more infections after that time step. This means that it would take a number of years before COVID-19 infection dies out of the population in the case presented by the authors of the DANMASK 19 study. It is also important to point out that the number of infected individuals continuously drops in the system, and never exceeds the initial number of infected people at the beginning of the study. This means that as time passes, the number of infections decreases, starting from the very first week that infection starts in Denmark. The disease also spreads slowly, taking many weeks to show the long term behavior of the system, but it slowly does continue to infect the population.

Case 2:

We can now model what would happen to a population the size of Denmark if masks were worn and enforced by the entire community. The authors of the DANMASK 19 study claimed that masks halve the risk of infection within the community. We can represent this in this system by manipulating the value of b. If everyone in the community were to wear a mask, we can change the value of b found in (8) to be half of what it was in Case 1, by computing (b/2). This calculation for b in Case 2 is thus, as follows.

$$b = (1.200759*10^{-7} \ / 2) = 6.003795* \ 10^{-8} \qquad (10)$$

The value of k remains the same in this case, k = 7/10, as regardless of wearing a mask, COVID-19 remains infectious for 10 days after symptoms develop. We will keep the values of R(0), S(0), and I(0) from (7) as the same, as b has no effect on these values. We can then model this system. The value of b found in (10), and k can now be used to manipulate our system for this case. We will substitute b, R(0), S(0), and I(0) into equations (4), (5), and (6) to model the system.

$$R(n+1) = R(n) + (7/10)\ I(n)$$

$$I(n+1) = I(n) + (6.003795 \ * \ 10^{-8})*S(n)*I(n) - (7/10)\ I(n) \qquad (11)$$

$$S(n+1) = S(n) - (6.003795 \ * \ 10^{-8})*S(n)*I(n)$$

The system in (11) was simulated for a number of weeks to observe its long term behavior as well. The graphical representations of the evolution of the system were created and are included as follows[11].

[11] A similar code to the one used in #9 of Appendix for Case 1 was used.

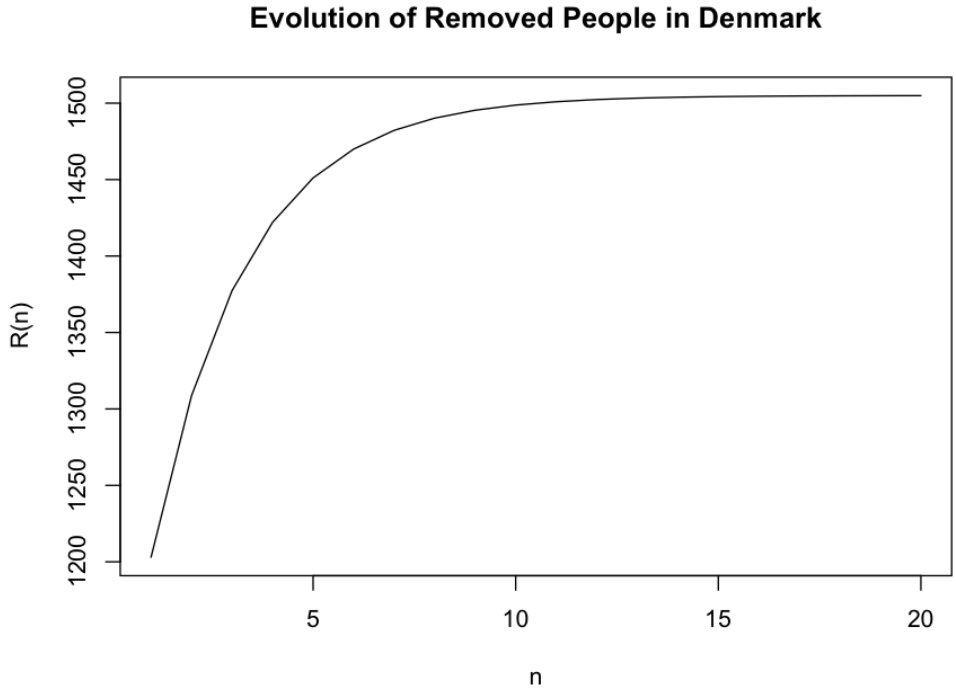

Figure 4: Representation of the number of removed individuals weekly in the system. The x-axis represents n, the number of weeks, and the y-axis, R(n), represents the number of removed individuals. The number of removed individuals stabilized by the 12th week at 1500 people.

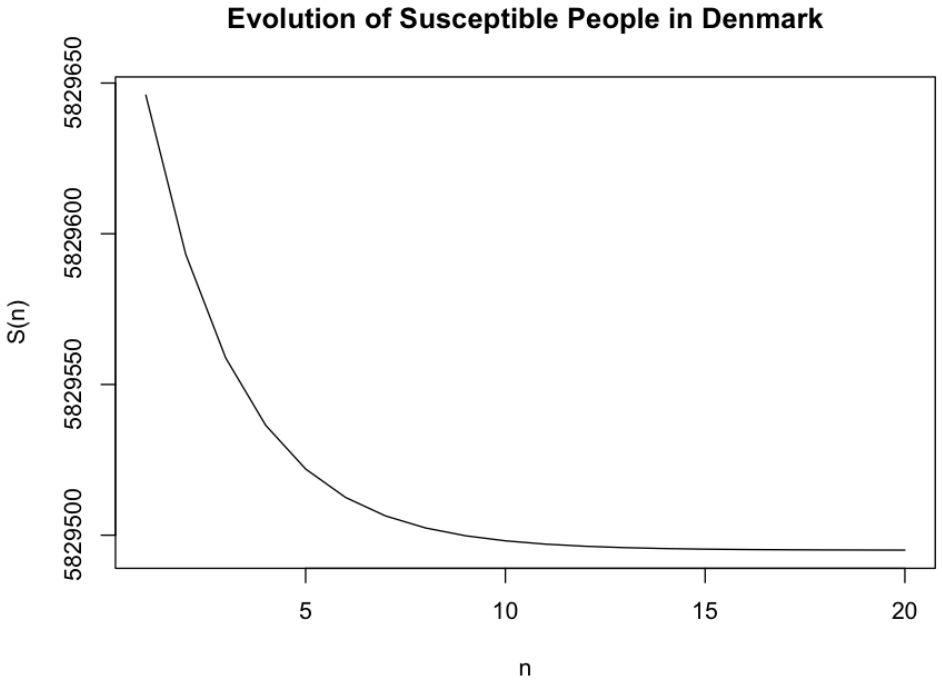

Figure 5: Representation of the number of susceptible people weekly in the system. The number of weeks n, is represented on the x-axis, and the number of susceptible people, S(n), is represented on the y-axis. The number of susceptible people stabilized at 5829500 people by the 12th week.

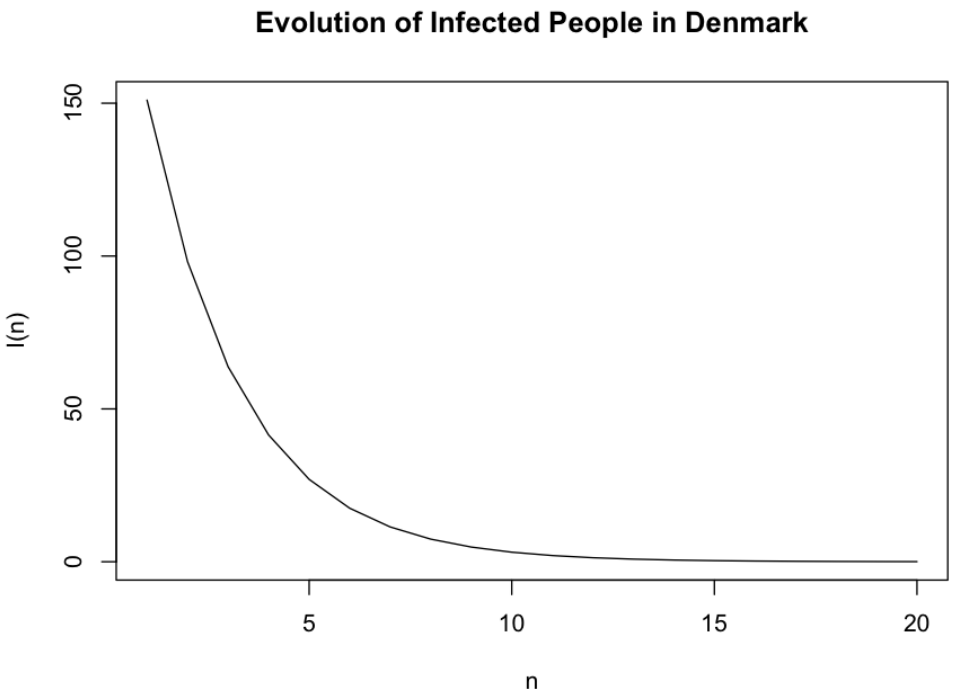

Figure 6: Representation of the number of infected people weekly in the system. The number of weeks, n, are represented on the x-axis. The number of infected people I(n), is represented on the y-axis. Number of infected people stabilized by the 15th week.

Examining Figures 4, 5, and 6, there is a dramatic difference in how this system evolves as compared to Case 1. In this case, by the 15th week, the number of infected individuals drops to zero and there are no more infections after that time amongst the population. The course of this infection is much faster than in Case 1. Based on this model, if everyone in a population the size of Denmark adhered to a mask mandate, COVID-19 would last for a mere number of months, as compared to years in the first case. Additionally, by the end of the infection course, less people would have been infected in this case. By the end of the infection course in Case 1, about 40000 people had been removed from the population through infection. Comparatively, in this case, less than half of that number of people have been removed. The authors of the DANMASK 19 study claimed that masks half infectivity, but show no benefit to users in their sample. However, using their claims, this SIR model shows that if the value of b was half of

what it was in the original case, meaning everyone in a population the size of Denmark were to wear a mask, not only would COVID-19 only last within the community for a number of weeks, less people would be infected.

Case 3:

We will also study the SIR model in the case where b is larger than the value of b from (8) that we found in Case 1. This is to examine how the population would react to a larger number of individuals being infected weekly by each infected individual. To do this, we will represent b as 10% larger than the value of b in (8) that we found originally in Case 1, meaning that every person is 10% more infective than originally represented in the DANMASK 19 study.

$$b = (1.200759 * 10^{-7}) * (1.1) \quad (12)$$

As in the previous case, the values of k, R(0), S(0), and I(0) from (7) remain the same, as the value of b has no effect on these values. We can then model this system as follows, by substituting the value of b from (12) into equations (4), (5), and 6.

$$R(n+1) = R(n) + (7/10)\, I(n)$$

$$I(n+1) = I(n) + ((1.200759 * 10^{-7}) * (1.1)) * S(n) * I(n) - (7/10)\, I(n) \quad (13)$$

$$S(n+1) = S(n) - ((1.200759 * 10^{-7}) * (1.1)) * S(n) * I(n)$$

The graphical representation of how the number of removed, infected, and susceptible individuals change weekly to model the long term behavior of the system from (13) was created and is as follows[12].

[12] A similar code to the one used in #9 of Appendix for Case 1 was used.

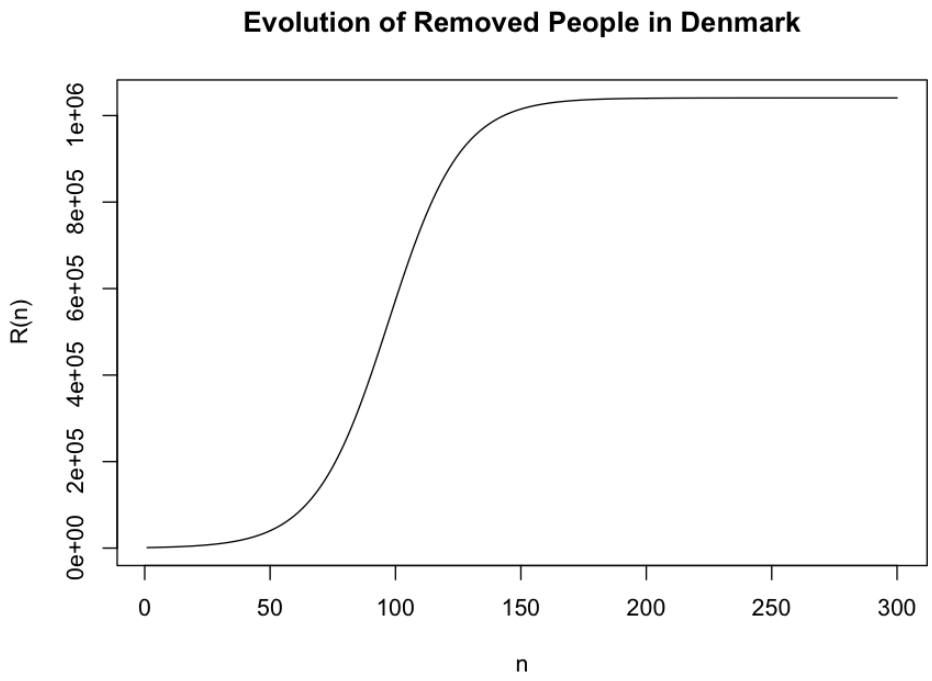

Figure 7: Representation of the progress of removed people in the system. The number of weeks is represented by n, on the x-axis. The number of removed individuals is represented on the y-axis. Number of removed people stabilizes at 10^6 people by the 150th week.

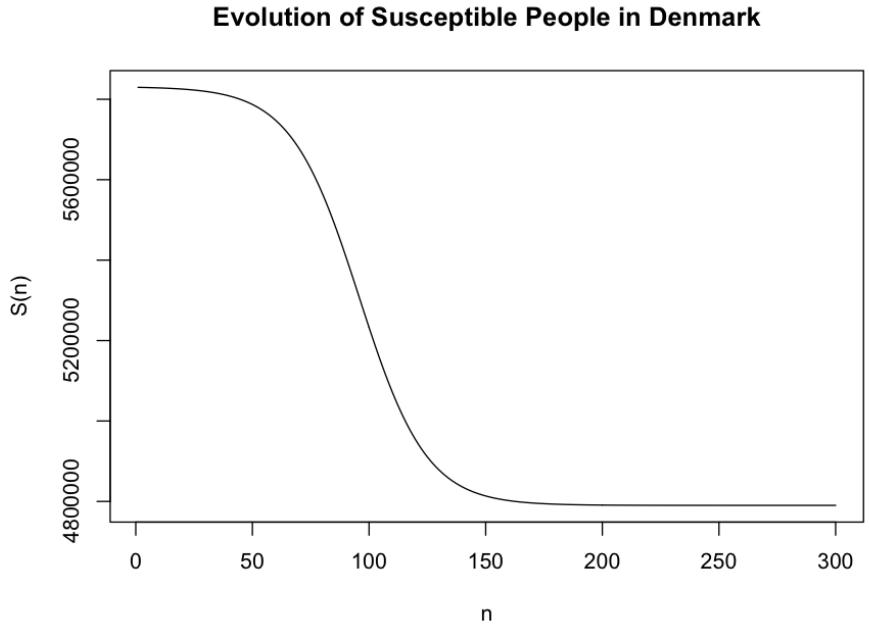

Figure 8: Representation of the number of susceptible people weekly in the system. The x-axis represents n, the number of weeks, and the y-axis represents, S(n), the number of susceptible people. Number of susceptible people stabilizes at 4800000 by the 200th week.

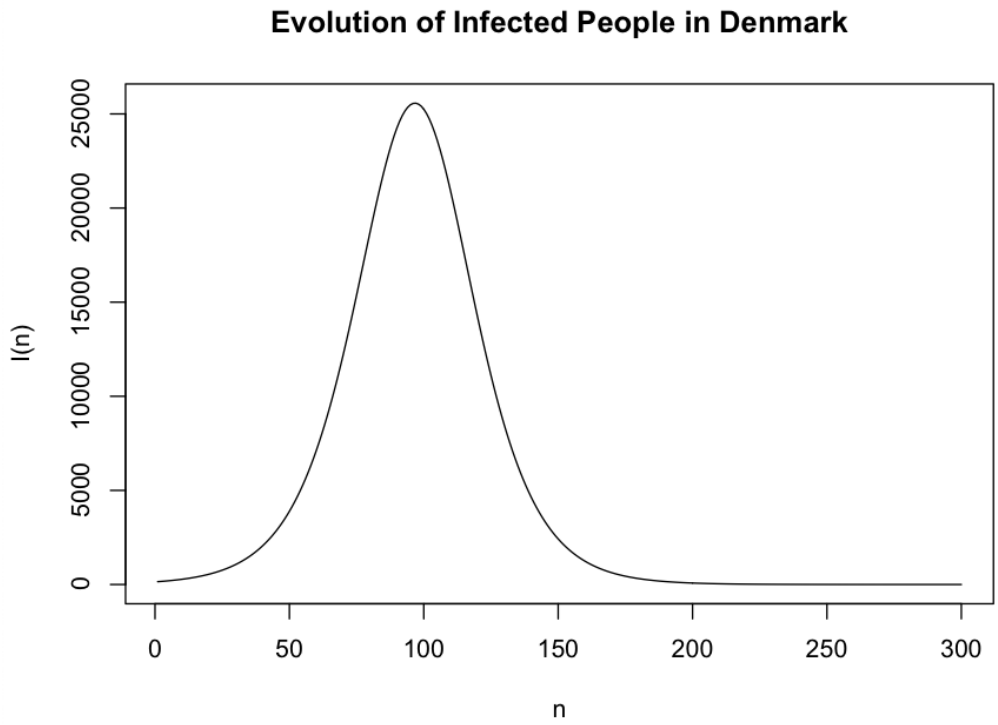

Figure 9: Representation of the number of infected people in the system. The x-axis represents, n, the number of weeks, and the y-axis represents I(n), the number of infected people in the system. Number of infected people stabilizes at 0 by the 200th week.

Considering Figures 7, 8, and 9, there is a difference in the long term behavior of this system as compared to Case 1. In this case, the number of infected individuals is reduced to zero by the 273rd week. This is a faster evolution than in Case 1, where b was 10% less than it is here, meaning that more individuals are infected faster when b is greater, as in this case. It took approximately 1200 weeks in Case 1 for the number of infected individuals to reach zero. This means that COVID-19 would run its course faster in Case 3, where one person can infect more individuals. It is interesting to note in this case that the number of infected people in a single week is higher than in Case 1. In Case 1, the number of infected people drops every week consistently. However, in this case, until the 100th week, the number of infected people increases weekly, until the number of infected people drops weekly after that. This means that more individuals are infected in this case than in Case 1, however, infection stops faster here. In this

case, when the infection has run its course on the population, 1029646 total people have been infected. In Case 1, however, 39646 total people were infected once the infection had run its course. The smallest number of individuals are infected in Case 2, where only 146 people have been infected once the infection has run its course. The most people total were infected in Case 3 by the end of the pandemic. So, although we changed b, the proportion of susceptible people weekly than an infected person can infect, by only a small amount here as compared to in Case 1, by only 10%, this small change changed the course of the pandemic, making the number of weeks that infection lasts in the community shorter, but the number of people infected in this time greater.

Section 10: Discussion and Conclusion

The goal of the analysis undertaken in this thesis was to understand the assumptions, limitations, and flaws of the DANMASK-19 study that allowed it to make the controversial conclusions that it did. The authors of the DANMASK 19 study claimed that masks were not shown to be effective in a community like the one in Denmark that they had studied. However, there were various issues within the design of the study and its assumptions that made these conclusions controversial.

For one, the authors of the study assumed that the COVID-19 infection rate in their community was 2%, which may have been inflated as per the fact that at the time of the study, Denmark was a community with a very low transmission rate of COVID-19. Additionally, the authors of the study claimed that wearing face masks reduces infection by COVID-19 by 50%. This is an inflated statistic as well. A combination of both of these issues meant that the sample size of the study was likely too small to have accurately represented the Denmark population, leaving the study underpowered, as discussed in Section 8. Considering these factors, we found that the sample size used in the study likely gave the study 50% power as opposed to the intended 80%. We also found that if masks reduce risk of infection by a quarter as opposed to a half, the sample size should have been much greater in order to accurately power the study.

The authors also failed to consider the protective impacts of masks for mask wearers in the results of their study. Additionally, using a majority of antibody tests to garner positive test results could have meant that many of positive results, if not all of them, were false positives, impacting the validity of the results.

Understanding these issues with the DANMASK 19 study, we sought to present a better model of the Denmark population and COVID-19 pandemic to show that if everyone in a

community similar to Denmark wore and enforced masks, then the outcome of the study would have been different. We created 3 versions of the SIR model to do this.

In generating the SIR model, a plausible model for the course of disease, the goal was to present a model for COVID-19 in Denmark to determine whether small changes in the assumptions made in the study would lead to big differences in the evolution of COVID-19 in Denmark. In particular, we showed that under certain assumptions, if the rate of infection was lowered, achieved by a mandatory mask mandate for all civilians, the course of the disease would be altered dramatically, and similarly if the infection rate was increased. Instead of simply using 6000 participants as the researchers in the DANMASK 19 study did, we decided to use a population the size of the entire population of Denmark, and consider the situation where the entire population enforced a mask mandate. We also used COVID-19 statistics (number of infected individuals) in Denmark during the dates of the study as the initial conditions of our analysis. Using this data, we simulated the number of infected, removed, and susceptible individuals in weekly intervals to determine the long-term behavior of COVID-19 in the simulation.

There was a clear difference in the evolution of the SIR model if mask wearing was not enforced in the population, compared to if mask wearing was enforced in the population. In Case 1 of Section 9, using the conditions of the DANMASK study, where mask wearing was not enforced in the community, we found that COVID-19 would persist in the community for more than 1000 weeks, until the number of infected individuals dropped to zero. However, if everyone in the community were to wear masks, there would be a substantial difference in the course of disease in the population. In Case 2 of section 9, we modeled what the course of COVID-19 would look like if everyone in a population similar to Denmark were to wear masks, and found that the number of infected individuals would drop to zero within just 15 weeks. While this is an

optimistic estimate, this shows that if mask wearing does reduce infection by 50%, it is incredibly effective in preventing infection when used by everyone in a community. Likewise, if mask wearing is enforced in the community, there would be fewer infected individuals in total.

We also used different methods of analyzing the SIR model to model the course of COVID-19 in Denmark. In Case 3 of Section 9, we manipulated the system to model what the course of the pandemic would look like if the proportion of susceptible people an infected person could possibly infect within a week was greater than we initially estimated in Case 1. We found that if this were the case, although we only made the proportion 10% greater than in Case 1, there was a drastic change in the course of disease. The number of infected individuals drop to zero faster, meaning the disease runs its course sooner, but a greater number of individuals total are infected in the long term.

Additionally, we found that regardless of the number of people weekly an infected individual could infect, in none of the cases did all individuals get infected. This is an argument in favor of mask wearing. In the case that the authors presented (Case 1), where only a small percentage of the population wears masks, and in the case where everyone in the population wears masks (Case 2), there were always some individuals who were not infected through the course of the disease. Mask wearing may have been the factor that prevented those people in the population from getting infected.

Considering all of these models, we can conclude that in a community where mask wearing is enforced and common, there can be considerable benefits for the community in terms of COVID-19 infection and spread. Firstly, fewer individuals will be infected while the disease runs its course. Also, the course of COVID-19 would be much shorter. While these are assumptions in an ideal case where every person wears their mask as recommended, this is an apt rebuttal to the claims made in the DANMASK 19 study.

Finding a benefit of mask wearing in this analysis is also a relevant and important result considering that in the United States of America, as well as other countries, face coverings are still recommended or mandatory in many public spaces. The SIR models in this analysis show evidence of a benefit to a population who adheres to a mask mandate. Infection will seize faster, and fewer people will be infected through the course of the disease. Likewise, with the emergence of new variants, such as Omicron, which is said to spread faster and more easily than other variants of COVID-19, an investigation of this kind is interesting to analyze how to control infection spread (14). Understanding the benefits of masks, and the flaws in studies that promote otherwise is important to preventing the spread of COVID-19, especially in the case where there is the possibility of infection falling to zero, as shown in our models. Thus, while targeting some of the issues in the DANMASK 19 study, we modeled COVID-19 under different infection rates in a population the size of Denmark and were able to demonstrate the potential benefits of mask wearing to users and communities.

Section 11: Appendix

1. Replicating the sample size calculation from the DANMASK-19 Study

The goal of the study was to produce a power of at least 0.8. This was assuming that wearing masks halves the risk of COVID-19 infection from 0.02 to 0.01. In this case, p_Y represents the probability of developing COVID-19 whilst wearing a mask, and p_X represents the probability of developing COVID-19 not wearing a mask (control group). This would mean that we are computing the power under the alternative hypothesis that the difference between p_X and p_Y would be 0.01. We are assuming here that m, the size of the mask group and n, the size of the control group are equal. If N represents the number of participants in the study as a whole, N = (m+n), so that N/2 = m =n.

Now that we have established these definitions and assumptions, we can determine the critical region, working at the $\alpha = 0.05$ confidence level. Firstly, under the null hypothesis H_0, assuming that $p_X = p_Y$ = p, we know that,

$$\frac{(\frac{X-Y}{N})-(p_X-p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n}+\frac{p_Y(1-p_Y)}{m}}} = \frac{(\frac{X-Y}{N})-(p_X-p_Y)}{\sqrt{\frac{2p_X(1-p_X)}{N}+\frac{2p_Y(1-p_Y)}{N}}} = \frac{(\frac{X-Y}{N})-(p_X-p_Y)}{2\sqrt{\frac{p(1-p)}{N}}}$$

So since under the null hypothesis, $p_X = p_Y$ = p, the critical region for

$$D = \frac{\frac{X-Y}{N}}{2\sqrt{\frac{p(1-p)}{N}}}$$

at the 5% level of significance is [-1.96, 1.96][c].

Based on this, to find the sample size, we can define the probability of a Type II error, β, in this study. A Type II error is the probability that we fail to reject the null hypothesis, given that it is false. We can determine the probability of Type II error by using the power desired in the study, since power = 1-β. Since the authors of this study aimed to have at least 80% power in

their study, the probability of a Type II error in this case is at most 0.2. The probability of Type II error is equal to the following.

$$P(-1.96 - \frac{0.01}{2\sqrt{\frac{p(1-p)}{N}}} \leq \frac{\frac{X-Y}{n} - 0.01}{2\sqrt{\frac{p(1-p)}{N}}} \leq 1.96 - \frac{0.01}{2\sqrt{\frac{p(1-p)}{N}}})$$

Since the probability of Type II error occurring is at most 0.02, this can be approximated by,

$$P(-1.96 - \frac{0.01}{2\sqrt{\frac{p(1-p)}{N}}} \leq Z \leq 1.96 - \frac{0.01}{2\sqrt{\frac{p(1-p)}{N}}}) \leq 0.2$$

In this case, we will assume that p=0.02, since the authors stated that at the time of the study, there was a 2% prevalence of infection by COVID-19 in Denmark.

$$P(-1.96 - \frac{0.01}{2\sqrt{\frac{0.02(1-0.02)}{N}}} \leq Z \leq 1.96 - \frac{0.01}{2\sqrt{\frac{0.02(1-0.02)}{N}}}) \leq 0.2$$

Equivalently,

$$P(-1.96 - 0.03571428571\sqrt{N} \leq Z \leq 1.96 - 0.03571428571\sqrt{N}) \leq 0.2$$

To solve for a sample size N that will satisfy these conditions, we will use a guess and check method, where we check different values of N to determine which value of N will satisfy the conditions and produce a probability ≤ 0.2. Doing this, the smallest value of N that satisfies these conditions is 6153. This is slightly different from the authors' sample size value, 4636.

However, the authors of the study aimed to include 20% more than the calculated value of N, in case participants failed to follow up. Based on that, the optimal number of participants would be 1.2*N, which in their case was about 6000, and based on our calculations is (1.2)*(6153), which is approximately 7384 participants.

2. Determining the sample size if masks do not lower transmission by COVID-19 by a half

If masks do not reduce risk of infection from COVID-19 by a half, but instead less, the sample size necessary to accurately power the study would change. To find this sample size, we can still use the same test statistic as in Appendix #1:

$$D = \frac{\frac{X-Y}{N}}{2\sqrt{\frac{p(1-p)}{N}}}$$

However, in this case, instead of $\frac{X-Y}{N}$ being 0.01, as would be the difference between risk of infection between mask wearers and non wearers if masks reduced the risk of infection by a half, we will use 0.005. This represents what would be the case if masks reduced risk of infection by a quarter, as opposed to a half. Based on this, we can now estimate the sample size in this case by replicating the same calculation as before in #1 of the Appendix.

The probability of Type II error is equal to the following.

$$P(-1.96 - \frac{0.005}{2\sqrt{\frac{p(1-p)}{N}}} \leq \frac{\frac{X-Y}{n} - 0.005}{2\sqrt{\frac{p(1-p)}{N}}} \leq 1.96 - \frac{0.005}{2\sqrt{\frac{p(1-p)}{N}}})$$

The probability of Type II error occurring is the same as before - at most 0.2, and can be approximated by,

$$P(-1.96 - \frac{0.005}{2\sqrt{\frac{p(1-p)}{N}}} \leq Z \leq 1.96 - \frac{0.005}{2\sqrt{\frac{p(1-p)}{N}}}) \leq 0.2$$

(assuming that p = 0.02)

$$P(-1.96 - \frac{0.005}{2\sqrt{\frac{0.02(1-0.02)}{N}}} \leq Z \leq 1.96 - \frac{0.005}{2\sqrt{\frac{0.02(1-0.02)}{N}}}) \leq 0.2$$

That method yields 24615 as the value of N - the sample size that satisfies the conditions. This is far greater than the original sample size value found in #1 of the Appendix, 6153, as well as the sample size desired by the authors of the study, 4636-6000.

3. Determining the power of the study if infection rate is lower

We determined in #1 of the Appendix that a sample size of N=6153 would provide the DANMASK 19 study with a power of 0.8 as desired by the authors of the study. However, Haber et al point out that the infection rate in Denmark may have been lower than the 2% reported, and masks may not half the risk of infection, but instead less. Considering this, we can determine what the power of the study would be in the case that the infection rate in Denmark was not 0.02, but 0.01, and the difference between masked and unmasked infection rates is 0.005, as opposed to 0.01. To do this, we can first determine the probability of a Type II error, β, in this case. Power is then the result of 1-β. We will carry out the following calculations to find the power of the study.

Based on the value of N=6153 that we found, and the calculation carried out in #1 of the Appendix, we should reject the null hypothesis, H_0, if and only if,

$$Z \geq -1.96 - \frac{0.005}{2\sqrt{\frac{p(1-p)}{N}}} \text{ or } Z \leq 1.96 - \frac{0.005}{2\sqrt{\frac{p(1-p)}{N}}}$$

Assuming that p = 0.01 and N=6153, this would mean if and only if

$$Z \geq -1.96 - \frac{0.005}{2\sqrt{\frac{0.01(1-0.01)}{6153}}} \text{ or } Z \leq 1.96 - \frac{0.005}{2\sqrt{\frac{0.01(1-0.01)}{6153}}}$$

Equivalently, this is represented as,

$$Z \geq -3.930905806 \text{ or } Z \leq -0.010905806$$

Thus, P(Type II error) = $P(-3.930905806 \leq Z \leq -0.010905806) \approx 0.496$.

This means that $\beta \approx 0.496$ and the power of the study in this case would be $1-\beta \approx 0.504$. This means that in the case that the prevalence of infection in Denmark is lower, and masks reduce risk of infection by less than a half, the sample size originally needed to provide the study with a power of 0.8 would not be sufficient. The original sample size would only provide the study with a power of about 0.5, much less than desired.

4. Constructing the 95% confidence interval for a between-group difference

Based on the two-sample binomial test, we know that

$$\frac{\frac{X}{n}-\frac{Y}{m}-(p_X-p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n}+\frac{p_Y(1-p_Y)}{m}}} \overset{approx.}{\sim} N(0,1)$$

In this, we can estimate p_X using $\overline{X} = \frac{X}{n}$ and p_Y using $\overline{Y} = \frac{Y}{m}$. Substituting these,

$$\frac{\overline{X}-\overline{Y}-(p_X-p_Y)}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}+\frac{\overline{Y}(1-\overline{Y})}{m}}} \overset{approx.}{\sim} N(0,1)$$

Thus the 95% confidence interval for a between-group difference can be constructed as follows.

$$P(-1.96 \leq \frac{\overline{Y}-\overline{X}-(p_Y-p_X)}{\sqrt{\frac{\overline{Y}(1-\overline{Y})}{m}+\frac{\overline{X}(1-\overline{X})}{n}}} \leq 1.96) \approx 5\%$$

$$P(-1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} \leq \overline{Y} - \overline{X} - (p_Y - p_Y)$$

$$\leq 1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}}) \approx 5\%$$

$$P(-1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} - (\overline{Y} - \overline{X}) \leq -(p_Y - p_X)$$

$$\leq 1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} - (\overline{Y} - \overline{X})) \approx 5\%$$

$$P(-1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} + (\overline{Y} - \overline{X}) \leq p_Y - p_X$$

$$\leq 1.96 * \sqrt{\frac{\overline{Y}(1-\overline{Y})}{m} + \frac{\overline{X}(1-\overline{X})}{n}} + (\overline{Y} - \overline{X})) \approx 5\%$$

5. Code used to determine the 95% confidence interval of between group difference in Section 6

To find the 95% confidence interval of the between group difference, the values m, n, $\overline{x}, \overline{y}$ were substituted into the formula found in Appendix #4,

$$P(-\ 1.96\ *\sqrt{\frac{\overline{Y}(1-\overline{Y})}{m}+\frac{\overline{X}(1-\overline{X})}{n}}\ +(\overline{Y}\ -\ \overline{X})\ \leq\ p_Y - p_X$$

$$\leq\ 1.96\ *\sqrt{\frac{\overline{Y}(1-\overline{Y})}{m}+\frac{\overline{X}(1-\overline{X})}{n}}\ +(\overline{Y}\ -\ \overline{X}))\ \approx\ 5\%$$

The code in R was as follows, where ybar=$\overline{y}$, and xbar=$\overline{x}$. The upper bound is the first line, and the lower bound is the second line.

```
> (-1.96)*sqrt((((ybar)*(1-ybar))/(m)) + (((xbar)*(1-xbar))/(n)))+(ybar - xbar)
> (1.96)*sqrt((((ybar)*(1-ybar))/(m)) + (((xbar)*(1-xbar))/(n)))+(ybar - xbar)
```

The confidence interval was found to be [-1.2, 0.4].

6. Constructing an odds ratio and its 95% confidence interval

It states in the study that odds ratios and confidence limits were calculated using logistic regression. Based on this,

$$log(\frac{P_X}{1-P_X})\ =\ \alpha\ +\ \beta x$$

where *x* is as follows: 1 for individuals from the control group, and 0 for individuals from the mask group.

We can then find the maximum likelihood estimator (MLE) for β. Maximum likelihood estimators allow us to estimate parameters of an assumed probability distribution, given observed data. As opposed to other methods of finding estimators, in the maximum likelihood method, we maximize the likelihood function after constructing it.

We can find the MLE as follows.

$$log(\frac{P_1}{1-P_1}) = \alpha + \beta$$

$$log(\frac{P_0}{1-P_0}) = \alpha$$

$$\beta = log(\frac{P_1}{1-P_1}) - log(\frac{P_0}{1-P_0})$$

$$\beta = log(\frac{P_1(1-P_0)}{P_0(1-P_1)})$$

In our case specifically, this would mean that

$$\beta = log(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})$$

and the estimator for the odds ratio would be $e^{\hat{\beta}}$.

To construct a 95% confidence interval for the odds ratio:

So we know that the Odds Ratio $(OR) = \frac{p_Y(1-p_X)}{p_X(1-p_Y)}$, and we found an estimator for this above,

$$\text{Estimator for } OR = e^{\hat{\beta}}$$

$$\text{Where } \hat{\beta} = ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}) \overset{approx.}{\sim} N(ln(OR), \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}}),$$

and in this case,

n_1= number of participants of the mask recommended group developed COVID-19

n_2= total number of participants in mask group

n_3= number of participants of the control group who developed COVID-19

n_4= total number of participants in control group

To find the 95% confidence interval for the odds ratios, we can do the following.

$$P(-\ 1.96 \le \frac{\beta - ln(OR)}{S} \le 1.96)\ \approx\ 5\%;$$

$$\text{Where } S\ =\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}}$$

$$\text{Therefore, } P(-\ 1.96 \le \frac{ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})}) - ln(OR)}{S} \le 1.96) \approx 5\%$$

Now, to isolate OR, we do the following.

$$P(-\ 1.96\ *\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}} \le ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})\ -\ ln(OR)\ \le$$

$$1.96\ *\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}})\ \approx\ 5\%$$

$$P(-\ 1.96\ *\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}} - ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})\ \le\ -\ ln(OR)\ \le$$

$$1.96\ *\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}}\ -\ ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})\ \approx\ 5\%$$

$$P(-\ 1.96\ *\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}}\ +\ ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})\ \le\ ln(OR)\ \le$$

$$1.96\ *\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4}}\ +\ ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})\ \approx\ 5\%$$

$$P(\ e^{-1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})} \le OR \le\ e^{1.96*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}+ln(\frac{\bar{Y}(1-\bar{\bar{X}})}{\bar{\bar{X}}(1-\bar{Y})})}\ )\ \approx\ 5\%$$

7. Using p = $\frac{X+Y}{n+m}$ to estimate $\sqrt{\frac{p\ (1-p\ )}{m} + \frac{p\ (1-p\ )}{n}}$

We can estimate $\sqrt{\frac{p\ (1-p\ )}{m} + \frac{p\ (1-p\ )}{n}}$ where p = $\frac{X+Y}{n+m}$, as

$\frac{1}{m+n}\sqrt{(Y\ +\ X)(m\ +\ n\ -\ (Y\ +\ X))(\frac{1}{m} + \frac{1}{n})}$ by doing the following.

$$\sqrt{\frac{p\ (1-p\ )}{m} + \frac{p\ (1-p\ )}{n}} = \sqrt{\frac{(\frac{X+Y}{n+m})(1-\frac{X+Y}{n+m})}{m} + \frac{(\frac{X+Y}{n+m})(1-\frac{X+Y}{n+m})}{n}}$$

$$= \sqrt{\frac{(\frac{X+Y}{n+m})(\frac{n+m}{n+m}-\frac{X+Y}{n+m})}{m} + \frac{(\frac{X+Y}{n+m})(\frac{n+m}{n+m}-\frac{X+Y}{n+m})}{n}}$$

$$= \sqrt{\frac{(Y+X)(m+n-(Y+X))}{m} * (\frac{1}{m+n})^2 + (\frac{(Y+X)(m+n-(Y+X))}{n} * (\frac{1}{m+n})^2}$$

$$= \sqrt{(\frac{1}{m+n})^2 * (\frac{(Y+X)(m+n-(Y+X))}{m} + \frac{(Y+X)(m+n-(Y+X))}{n})}$$

$$= \frac{1}{m+n}\sqrt{(Y+X)(m+n-(Y+X))(\frac{1}{m}+\frac{1}{n})}$$

8. The Explanation for Bayes' Law, as referenced in Section 6

Bayes' law describes the probability of an event occurring, while taking into account the knowledge of information that may affect the probability. Bayes Law is as follows.

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$$

Dr. Frieden cited Bayes' Law to show that all of the antibody-positive results in both intervention and control groups could have been false positives. Frieden described a 1.5% probability of a false positive on an antibody test, and the authors of the study stated that there was a 2% prevalence of infection.

In terms of Bayes' Law, this can be described as follows.

P(none of the participants were actually positive | 95 participants tested positive) =

$$\frac{P(95\ participants\ test\ positive\ |\ no\ participants\ actually\ positive) * P(no\ participants\ positive)}{P(95\ participants\ test\ positive)}$$

Of this equation, Frieden utilized only two of the probabilities to make his claim. Namely, he referenced P(95 participants test positive | no participants actually positive) and P(95 participants test positive) in order to conclude that all of the positive results in the study could have been false positives.

We can calculate both of these probabilities as follows.

- P(95 participants test positive | no participants actually positive)

  represents the probability of 95 of the positives being false positives, and using the 1.5% probability of a false positive, equals

  $= \binom{4862}{95} * (0.015)^{95} * (1 - 0.015)^{4862-95} = 0.001866635$

- P(95 participants test positive)

  represents the probability that 95 participants of the study test positive, and using the 2% prevalence of infection, equals

  $= \binom{4862}{95} * (0.02)^{95} * (1 - 0.02)^{4862-95} = 0.04021902$

Based on these calculations, there is approximately a 0.02% probability that 95 participants test positive given that no participants actually test positive, and approximately a 4% probability that 95 participants in the study test positive. Now, if we consider the probability of 95 participants testing positive, it is clear that if the rate of infection changes, this probability will change as well. Frieden states in his critique that the rate of infection in Denmark may have been lower than cited in the study. If this were the case, then the probability of 95 participants testing positive would decrease. Specifically, if the rate of infection were to be 1.5%, the probability of infection in 95 participants would decrease such that it would be the same as the probability of 95 participants testing positive given that none of the participants actually developed COVID-19.

- P(95 participants test positive) =

  $=\binom{4862}{95} * (0.015)^{95} * (1 - 0.015)^{4862-95} = 0.001866636$

This was Frieden's point - that since the rate of infection in the community was lower than reported, in the case that the rate of infection was 1.5%, the probability of 95 participants testing positive, and all 95 of those positive tests being false positives, could be the same.

9. Code used to generate graphs for SIR model in Case 1, Section 9

To generate graphs to demonstrate the evolution of the SIR model for Case 1, the SIR system was simulated in R. The code used to do this was as follows.

In R:

```
> N = 2000

> k=7/10

> b = 1.200759e-07

> A = array(rep(0,3*N),c(N,3))

> A[1,]=c(1203, 5829646, 151)

> for (i in 2:N)

A[i,]=c(A[i-1,1]+k*A[i-1,3],A[i-1,2]-b*A[i-1,2]*A[i-1,3],A[i-1,3]*(1-k+b*A[i-1,2]))

> A
```

References

1. Abbasi, K. (2021, November 26). *The curious case of the Danish mask study*. thebmj. Retrieved December 14, 2021, from https://www.bmj.com/content/371/bmj.m4586
2. Bundgaard, H. (2021, March). *Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers*. Annals of Internal Medicine. Retrieved December 14, 2021, from https://www.acpjournals.org/doi/10.7326/m20-6817
3. Cassata, C., Chesak, J., & Schimelpfening, N. (2021, December 8). *How Long Does Immunity Last After COVID-19? What We Know*. Healthline. Retrieved January 22, 2022, from https://www.healthline.com/health-news/how-long-does-immunity-last-after-covid-19-what-we-know
4. *Coronavirus disease (COVID-19): Vaccines*. (2021, November 29). WHO | World Health Organization. Retrieved December 20, 2021, from https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-vaccines
5. *Coronavirus History: How did coronavirus start?* (2021, August 15). WebMD. Retrieved December 14, 2021, from https://www.webmd.com/lung/coronavirus-history
6. *Coronavirus in the U.S.: Latest Map and Case Count*. (n.d.). The New York Times. Retrieved January 12, 2022, from https://www.nytimes.com/interactive/2021/us/covid-cases.html
7. Crist, C. (2021, November 22). *US COVID-19 Deaths in 2021 Surpass 2020 Total*. WebMD. Retrieved December 20, 2021, from https://www.webmd.com/lung/news/20211122/us-covid-deaths-2021-surpass-2020-total

8. Frieden, T. R., & Cash-Goldwasser, S. (2021, March). Of Masks and Methods. *Annals of Internal Medicine*. https://www.acpjournals.org/doi/10.7326/M20-7499
9. Gouvea, W. (n.d.). *Natural history of COVID-19 and current knowledge on treatment therapeutic options*. NCBI. Retrieved December 14, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7332915/
10. Haber, N. A., Weiten, S. E., & Smith, E. R. (n.d.). *Letter of concern regarding »Reduction in COVID-19 infection using surgical facial masks outside the healthcare system«*. Danish Medical Journal. Retrieved December 14, 2021, from https://ugeskriftet.dk/dmj/letter-concern-regarding-reduction-covid-19-infection-using-surgical-facial-masks-outside-healthcare
11. *Interim Clinical Considerations for Use of COVID-19 Vaccines*. (n.d.). CDC. Retrieved December 14, 2021, from https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html#background
12. Kolata, G. (2021, January 12). *A New Study Questions Whether Masks Protect Wearers. You Need to Wear Them Anyway. (Published 2020)*. The New York Times. Retrieved December 14, 2021, from https://www.nytimes.com/2020/11/18/health/coronavirus-masks-denmark.html?referringSource=articleShare
13. McCoy, E. (2017, September 18). *Understanding the Intention-to-treat Principle in Randomized Controlled Trials*. NCBI. Retrieved December 14, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5654877/
14. *Omicron Variant: What You Need to Know*. (n.d.). CDC. Retrieved December 14, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/variants/omicron-variant.html

15. Sterne, J. A. C., & White, I. R. (2009, June 29). *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls*. thebmj. Retrieved December 14, 2021, from https://www.bmj.com/content/338/bmj.b2393
16. Szumilas, M. (n.d.). *Explaining Odds Ratios*. NCBI. Retrieved December 14, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/
17. Tenny, S. (2021, May 30). *Odds Ratio - StatPearls*. NCBI. Retrieved December 14, 2021, from https://www.ncbi.nlm.nih.gov/books/NBK431098/
18. *A tutorial on sensitivity analyses in clinical trials: the what, why, when and how - BMC Medical Research Methodology*. (2013, July 16). BMC Medical Research Methodology. Retrieved December 14, 2021, from https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-92
19. *Your Guide to Masks*. (n.d.). CDC. Retrieved December 14, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/about-face-coverings.html