Investigating the marginal and herd effects of COVID-19 vaccination for reducing case fatality rate: Evidence from the United States.

Tenglong Li¹*, Zilong Wang², Shuyue He¹ and Ying Chen¹

¹ Wisdom Lake Academy of Pharmacy, Xi’an Jiaotong-Liverpool University
² Department of Financial and Actuarial Mathematics, Xi’an Jiaotong-Liverpool University

* To whom correspondence may be addressed: Tenglong.Li@xjtlu.edu.cn

Abstract

Vaccination campaigns have been rolled out in most countries to increase the vaccination coverage and protect against case mortality during the ongoing pandemic. To evaluate the effectiveness of COVID-19 vaccination, it is vital to disentangle the herd effect from the marginal effect and parameterize them separately in a model. To demonstrate this, we study the relationship between the COVID-19 vaccination coverage and case fatality rate (CFR) based on a U.S. vaccination coverage at county level, with daily records from March 11th, 2021 to Jan 26th, 2022 for 3109 U.S. counties. Using segmented regression, we discovered three breakpoints of the vaccination coverage, at which the herd effects could potentially exist. Controlling for county heterogeneity, we found the size of the marginal effect was not constant but actually enlarged as the vaccination coverage increased, and only the herd effect at the first breakpoint was statistically significant, which implied indirect benefit of vaccination may exist at the early stage of a vaccination campaign. Our results have demonstrated that public health researchers should carefully differentiate and quantify the herd and marginal effects in analyzing vaccination data, to better inform vaccination campaign strategies as well as evaluate vaccination effectiveness.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Keywords: COVID-19 vaccine, herd effect, marginal effect, case fatality rate, segmented regression.

1 Introduction

The world has been living in the tunnel of COVID-19 pandemic since the outbreak in 2019, yet without a clear idea about its outlet. A great hope has been placed in COVID-19 vaccines to end the pandemic, as clinical trial results suggested COVID-19 vaccination can effectively prevent symptomatic infections especially severe symptoms, which protects against mortality associated with infections [1-3]. For this reason, public demand for COVID-19 vaccines was fervent and vaccination campaigns were initiated all over the world, for an early safe vaccine supply for populations at risk as well as a massive vaccine supply to match the public’s demand [4-5]. For an example, the Food and Drug Administration (FDA) issued emergency use authorizations (EUA) for Pfizer-BioNTech and Moderna COVID-19 vaccines in December 2020, which marked the beginning of the vaccination campaign in the U.S. COVID-19 vaccines were then first allocated for populations at risk, the elderly population (age 65+) and the frontline (mostly healthcare and education) workers. After president Biden announced that all Americans would be eligible for COVID-19 vaccines by May 1st, 2021, the vaccination campaign was further accelerated [6]. Booster doses of COVID-19 vaccines were introduced to restore the level of protection (antibody) eroded by time [7-9]. By November 24, 2022, more than 80% of Americans have received at least one dose and more than 68% of Americans have completed a primary series of COVID-19 vaccine [10]. Literature has reported that the vaccination coverage is negatively associated with
case fatality rate (CFR), which refers to the mortality rate among those who are infected (i.e., confirmed COVID cases) [11-12].

It is necessary to decompose the protection effect of COVID-19 vaccines in order to better understand the underlying mechanism [13]. The protection effect of COVID-19 vaccines is in general a mix of two different effects, i.e., the direct effect and the indirect effect [14]. The direct effect refers to direct protection of inoculated individuals, as the vaccines can effectively reduce individual susceptibility to COVID-19 infection and severe symptoms [13-14]. The indirect effect, however, is a bit abstract and attributed to herd immunity, which is a conception states that transmission of the agent can be largely prevented if a fixed proportion of the population is immunized (either by vaccination or by recovery from infection; this proportion is called herd immunity threshold), rendering an infectious disease insignificantly dangerous for public health [15-16]. The indirect effect is defined as the protection gained by unvaccinated people, through the reduced number of infected people in the population as well as their reduced infectiousness, which can be achieved by vaccinating certain proportions of the population [13]. It should be noted here that those proportions we mentioned above are different from herd immunity threshold as they potentially correspond to different levels of herd immunity in a population [17]. In fact, those proportions are thresholds for triggering the indirect effect (with different sizes) in the course of a vaccination campaign for a target population.

The above concepts of the direct and indirect effects should be contextualized in the investigation of the impact of COVID-19 vaccination on case fatality rate (CFR). The direct effect could be interpreted as the reduction in CFR associated with one unit/percent
increase in the vaccination coverage, i.e., the direct effect evaluates the marginal gain during a vaccination campaign. For this reason, the direct effect is referred to as the marginal effect in this paper. The indirect effect could be interpreted as the additional reduction in CFR if the vaccination coverage passes certain unknown thresholds, i.e., the indirect effect quantifies the additional gain potentially due to herd immunity in the process of vaccinating a target population. To better characterize its nature, the indirect effect is referred to as the herd effect in this paper. It’s particularly important to disentangle the herd effect from the marginal effect, for the following three reasons: First, the marginal and herd effects address different scientific questions with regard to distinct groups of people (i.e., the vaccinated individuals versus the unvaccinated individuals). Second, as discussed earlier, there are underlying thresholds for triggering the herd effects, and those thresholds essentially delineate different stages in a vaccination campaign where the marginal and herd effects may not be constant across those stages. Third, given the aforementioned two reasons, a deeper knowledge about the protection effect of vaccination is likely gained by learning the marginal and herd effects, and vaccination strategies could be optimized for a target population based on such knowledge. Unfortunately, we haven’t seen research on this important topic so far. Our goal in this paper is to estimate the herd and marginal effects based on a dataset from the U.S. Centers for Disease Control and Prevention (CDC), which records various vaccination coverages for each U.S. county daily [10]. We hypothesize that both the herd and marginal effects exist and are significantly negative for modeling CFR. The segmented regression is employed first to identify the breakpoints which are considered as the thresholds for triggering the herd effect, based on data of all the U.S. counties.
included in our study. With the identified breakpoints, we estimate the herd and marginal 
effects at national level using segmented regression and at county level using mixed 
model. Data on social vulnerability index (SVI) for individual counties is also included to 
control for health disparities due to sociodemographic factors at county level [18]. 
Heterogeneity among individual counties is further evaluated by the random effects 
associated with the herd and marginal effects among in a mixed model. 

This paper is structured as follows: In the next section of materials and methods, the data 
used in this paper will be described in details, along with the models adopted for analyses 
at both national level and county level. The results from our analyses at national level and 
county level are presented and explained in the section of results, with a focus on the 
estimation and interpretation of the herd and marginal effects of vaccination regarding 
CFR. Our findings will be summarized in the discussion section, where important 
implications and limitations of our study will also be discussed. 

2 Materials and Methods 

2.1 Data 

Our data comes from three different sources. The US vaccine administration and equity 
dataset is obtained from the CDC website and has vaccination coverages of the general 
population and its subpopulations (defined by age) recorded daily at county level [10]. 
The percent of people who completed a primary series of vaccination in the general 
population was extracted from the dataset and served as the main covariate in our model. 
The daily CFR at county level was calculated as the ratio between the daily count of 
deaths and the daily count of COVID-19 cases, based on the time series summary tables 
of COVID-19 deaths and confirmed cases, which were accessed from the COVID-19
data repository by the Center for Systems Science and Engineering (CSSE) at John Hopkins University [19]. To further control for county heterogeneity, we used a dataset from the Centers for Disease Control and Prevention Social Vulnerability Index (CDC SVI) database, created by the Geospatial research, Analysis & Services Program under the Agency for Toxic Substances and Disease Registry [18]. The CDC SVI database was established to help health officials and emergency response planners identify counties that will most likely need support before, during, and after a hazardous event. CDC SVI ranks counties on 15 social factors and further groups them into four themes, namely socioeconomic status, household composition & disability, minority status & language, and housing type & transportation [20]. We chose to use the theme-specific ranking which was constructed by summing the percentiles of the factors under each theme. The theme-specific ranking was set in the range from 0 to 1, with higher values indicating greater vulnerability.

The vaccination coverages and daily CFR for the period between March 11th, 2021 and Jan 26th, 2022 were selected. We chose March 11th, 2021 as it was the date when president Biden announced that COVID-19 vaccine would be available for all American adults by May 1st, 2021, an event marked the beginning of massive vaccination campaign in the U.S. We chose Jan 26th, 2022 as the ending date of our study as it was reported on this date that Omicron variant accounted for 99.9% of the new infections. This would alleviate the concern of potential confounding effect of Omicron variant regarding the relationship between the vaccination coverage and CFR. 31 counties with missing values on county FIPS code, vaccination coverages, the CDC SVI or CFR were excluded, and the final dataset has 1001098 observations clustered by 3109 U.S. counties. To prepare
the dataset for analysis at national level, we further extracted the average CFR and average vaccination coverage (i.e., the percent of people who completed a primary series of COVID-19 vaccine) across all the counties in our dataset for each day during our study period.

2.2 Models

Segmented regression models were employed to estimate the herd and marginal effects. Segmentated regression is very similar to ordinary regression, with the only difference that regression coefficients should be estimated repeatedly for different local regions whose boundaries are defined by breakpoints, which represent the thresholds of structural changes in regression models [21-22]. Typically, the first step is to determine the number of breakpoints, which can be achieved by a model selection alike procedure, i.e., models with different number of breakpoints are compared in terms of their model fit indices (such as AIC or BIC) to determine the optimized number of breakpoints. The second step is to estimate the locations of breakpoints given the number of breakpoints. The third step, based on the estimated breakpoints, is then to fit regression models to different local regions separated by the breakpoints. Normally, one would expect all regression coefficients to be changeable across different regions, unless otherwise specified.

For our analysis at national level, we intend to examine the relationship between CFR and the vaccination coverage, based on the dataset comprising only the average vaccination coverage and CFR in the U.S. The following regression model is formulated for the analysis at national level:

\[
y_t = \beta_0 + \beta_1 x_t + \sum_{k=1}^{m} \alpha_k I(x_t \in \psi_k) + \sum_{k=1}^{m} \delta_k x_t I(x_t \in \psi_k) + \varepsilon_t
\] (1)
where $y_t$ and $X_t$ denote the average CFR and vaccinate coverage in the U.S on day $t$. The model (1) is built on the estimated breakpoints $b_1 < b_2 < \cdots < b_m$, which implies there are $m + 1$ different local regions and $m$ different breakpoints in total (except $b_0$ and $b_{m+1}$ which are the minimum and maximum of $X_t$). The local regions separated by the breakpoints are denoted by $\Psi_k = [b_k, b_{k+1})$ for $k = 1, 2, \cdots, m$. The reference local region $\Psi_0$, although omitted from the model (1), refers to the local region $\Psi_0 = [b_0, b_1)$. The indicator function $I_{\{X_t \in \Psi_k\}}$ creates the dummy variable which assigns value 1 if the value of $X_t$ falls in the local region $\Psi_k$ and 0 otherwise, which operationally divides the range of $X_t$ into the local regions. The marginal effects in those local regions are characterized by $\beta_1$ for the reference region and $\beta_1 + \delta_k$ for the local region $\Psi_k$, and these parameters quantify the marginal gain/drop in CFR if the vaccination coverage increases by one percent. The herd effects for the local region $\Psi_k$ relative to its previous region are characterized by $\alpha_k - \alpha_{k-1}$ (for $\Psi_1$ it is just $\alpha_1$), as $\alpha_k$ quantifies the additional gain/drop in CFR if the vaccination coverage passes the threshold $b_k$, compared to the intercept term $\beta_0$ in the reference region $\Psi_0$.

The breakpoints $b_k$ $k = 1, 2, \cdots, m$, are estimated based on the model (1) and the dataset for the analysis at national level (i.e., with only average daily CFR and vaccination rate in the U.S.). Naturally, they reflect the structural changes in the relationship between CFR and the vaccination coverage in general, and they can be applied to the analysis at county level where we use the longitudinal data (322 days) for all the counties (3109 counties), along with the CDC SVI indicators for explaining county heterogeneity. We build the following mixed model for the analysis at county level:
\[ y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \sum_{k=1}^{m} \alpha_k I_{\{x_{i,t} \in \psi_m\}} + \sum_{k=1}^{m} \delta_k X_{i,t} I_{\{x_{i,t} \in \psi_m\}} + \gamma Z_i + u_1 X_{i,t} + u_0 \]

\[ + \sum_{k=1}^{m} \tilde{\alpha}_k I_{\{x_{i,t} \in \psi_m\}} + \sum_{k=1}^{m} \tilde{\delta}_k X_{i,t} I_{\{x_{i,t} \in \psi_m\}} + \epsilon_{i,t} \]

where \( y_{i,t} \) and \( X_{i,t} \) denote the CFR and vaccination coverage for county \( i \) at day \( t \). \( Z_i \) is the covariate vector that contains CDC SVI theme-specific rankings on the four main themes for county \( i \). \( \beta_0, \beta_1, \alpha_k, \delta_k \) are the parameters characterize the herd and marginal effects, as similarly defined in the model (1), except in the model (2) they are fixed effects. Correspondingly, we have their random effects characterized by \( u_0, u_1, \tilde{\alpha}_k, \tilde{\delta}_k \) that are due to the heterogeneity among the counties that cannot be explained away by the fixed effects of county rankings on CDC SVI, which are represented by \( \gamma \). The Model (2) is built on the same set of breakpoints \( b_k, k = 1, 2, \ldots, m \), that are obtained based on the model (1) and the dataset for the analysis at national level. This means the model (2) shares the same local regions \( \Psi_k = [b_k, b_{k+1}) \) for \( k = 1, 2, \ldots, m \), across all the counties in our study. The significances of the fixed effects \( \beta_{0,i}, \beta_{1,i}, \alpha_k, \delta_k \) as well as their corresponding random effects \( u_0, u_1, \tilde{\alpha}_k, \tilde{\delta}_k \) will be checked via model outputs and comparison tests.

3 Results

3.1 The results of the analysis at national level

As mentioned above, the dataset used for the analysis at national level has two variables, i.e., average daily CFR and average vaccination coverage in the U.S.. The breakpoints were estimated based on this dataset using the “segmented” package in R (version 4.2.0) [22]. To avoid overfitting, we set the maximum number of the breakpoints as 3, based on
the curve between the average daily CFR and the average daily vaccination coverage depicted in Figure 1. The segmented package then does an automatic selection of the number of breakpoints based on BIC, and it estimated the locations of the breakpoints conditional on the optimized number of the breakpoints. The estimated breakpoints were superimposed on the curve in Figure 1, to further validate those estimates align with the observed structural changes.

The breakpoints were estimated as 32%, 36% and 47%, which suggested that the herd effect of vaccination may be associated with the thresholds of 32%, 36% and 47% in the vaccination coverage. Based on those breakpoints, we have four different local regions, namely \( \Psi_1 = [8.66\%, 32\%); \ \Psi_2 = [32\%, 36\%); \ \Psi_3 = [36\%, 47\%); \ \Psi_4 = [47\%, 49.28\%) \), with the minimum and maximum of the average daily vaccination coverages as 8.66% and 49.28% respectively. Table 1 lists the estimates of the regression coefficients based on the model (1). We further calculated the herd and marginal effect estimates which are tabulated in Table 2. The marginal effect in the first local region \( \Psi_1 \) was insignificant, which suggested that the drop in the CFR per percent increase in the vaccination coverage was not significantly different from 0, if the vaccination coverage did not surpass 32%. The herd effect at the threshold 32% was also insignificant, which was largely due to the insignificant marginal effect in the region \( \Psi_1 \). We found a significant marginal effect in the second local region \( \Psi_2 (-0.057) \), which indicated that there was a drop of 0.057 percent in the CFR for every percent increase in the vaccination coverage in this region, evidencing that the protection effect of COVID vaccination against mortality. In addition, the herd effect at the threshold 36% was significant too (-0.233), suggesting that there was a further drop of 0.233 percent in the CFR besides the
marginal CFR reduction per percent increase in the vaccination coverage. In the third
local region $\Psi_3$, however, we observed a slight positive marginal effect in the CFR
(0.003), which means the marginal gain of vaccination (in terms of the reduction in CFR)
disappeared and vaccination was somehow harmful for protecting against mortality.
Correspondingly, the herd effect at the threshold 47% was also positively significant
(0.009), suggesting again that vaccination was not helpful at this stage. The marginal
effect in the fourth local region $\Psi_4$ was strongly negative, specifically there was a drop of
0.115 percent in the CFR associated with every percent increase in the vaccination
coverage at this stage.

Figure 1. The relationship between the average vaccination coverage and the average
CFR in the U.S. The solid dots represent the breakpoints estimated by the “segmented”
package in R.
Table 1. The regression model parameter estimates for the analysis at national level.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>T Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>1.976</td>
<td>624.70</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.000</td>
<td>0.28</td>
<td>0.78</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.003</td>
<td>-0.90</td>
<td>0.37</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>-0.057</td>
<td>-45.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.236</td>
<td>-80.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.003</td>
<td>7.50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-0.227</td>
<td>-55.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>-0.115</td>
<td>-47.03</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 2. The herd and marginal effect estimates for the analysis at national level.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Location</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Marginal Effect</td>
<td>$\Psi_1 = [8.66%, 32%]$</td>
<td>0.000</td>
<td>0.78</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Herd Effect</td>
<td>$b_1 = 32%$</td>
<td>-0.003</td>
<td>0.37</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Marginal Effect</td>
<td>$\Psi_2 = [32%, 36%]$</td>
<td>-0.057</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Herd Effect</td>
<td>$b_2 = 36%$</td>
<td>-0.233</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Marginal Effect</td>
<td>$\Psi_3 = [36%, 47%]$</td>
<td>0.003</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Herd Effect</td>
<td>$b_3 = 47%$</td>
<td>0.009</td>
<td>0.02</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Marginal Effect</td>
<td>$\Psi_4 = [47%, 49.28%]$</td>
<td>-0.115</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
3.2 The results of the analysis at county level

We further investigated the marginal and herd effects of COVID vaccination based on an analysis at county level, where the daily CFR and vaccination coverages from March 11th, 2021 to Jan 26th, 2022 as well as the CDC SVI rankings for 3109 U.S. counties were used. The estimated breakpoints of 32%, 36% and 47%, obtained based on the analysis at national level, were adopted for our analysis at county level. The mixed model (2) was employed to account for the clustered data at county level, and its fixed and random effect estimates are tabulated in the Table 3. Furthermore, the estimates of herd and marginal effects, as well as their corresponding random effect estimates, are listed in the Table 4. To determine the significance of the random effects, we compared the full model (i.e., the model (2)) with two different reduced models (one without the random effects associated with all the marginal effects, i.e., $u_1, \delta_1, \delta_2, \delta_3$; another one without the random effect associated with the first marginal effect only, i.e., $u_1$), and the resultant tests gave p-values smaller than 0.001, suggesting that it was necessary to include random effects for all the marginal and herd effect parameters.

Across all the U.S. counties in our data, the marginal effect was significantly negative in the first local region $\Psi_1$ (i.e., when the vaccination coverage was between 8.66% and 32%), specifically one percent increase in the vaccination coverage was associated with 0.004 percent drop in the CFR. The first herd effect at the threshold of 32% was -0.025 and significant, meaning there was an additional drop of 0.025 percent in the CFR as the vaccination coverage reached 32%, beyond the marginal effect observed in $\Psi_1$. The marginal effect in the second local region $\Psi_2$ was also significantly negative (-0.01), which showed that there was 0.01 percent drop in the CFR per one percent increase in the
vaccination coverage, when the vaccination coverage was between 32% and 36%. The second herd effect, however, was overall insignificant, which suggested additional protection effect at the threshold 36% may not exist. Similarly, we found significant marginal effect (-0.023) for the third local region $\Psi_3$ but insignificant herd effect at the threshold 47%. The marginal effect within the fourth local region $\Psi_4$ was the strongest, as every percent increase in the vaccination coverage was associated with 0.043 percent reduction in the CFR, if the vaccination coverage surpassed 47%.

Furthermore, heterogeneity among the U.S. counties regarding the herd and marginal effect estimates was evident. For the first herd effect (at 32%), the fixed effect estimate was -0.025 with a random effect of 0.024, and this means roughly 56% of the counties had negative herd effects as expected, but the other 44% of the counties could have no herd effect or even positive herd effects at the threshold 32%. For the second and third herd effect (at 36% and 47% respectively), roughly 49% of the counties have negative herd effects, which further demonstrated that those two herd effects were not significant among the counties. Regarding the marginal effects: although their fixed effect estimates were all very significant (p-value < 0.001), their random effect estimates suggested the fourth marginal effect was the strongest one (was negative in 73% of the counties). The first, second and third marginal effects were negative in approximately 57%, 54% and 62% of the U.S. counties. All taken, the protection effect of COVID vaccination was confirmed in general and for the majority of the U.S. counties, while substantial heterogeneity that defined the size and the validity of the protection effect for individual county still existed. We also found that only one CDC SVI theme ranking, i.e., rankings on household composition & disability, could help explain the county heterogeneity.
Unsurprisingly, this CDC SVI theme ranking was positively related to CFR, and specifically one percentile rise in the theme ranking could result in 0.8 percent increase in CFR.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>F.E. Est</th>
<th>F.E. T Ratio</th>
<th>F.E. p-value</th>
<th>R.E. Est</th>
<th>S.E. of R.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>1.64</td>
<td>28.88</td>
<td>&lt;0.001</td>
<td>1.33</td>
<td>0.03</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.004</td>
<td>-8.85</td>
<td>&lt;0.001</td>
<td>0.0005</td>
<td>0.00001</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.025</td>
<td>-7.35</td>
<td>&lt;0.001</td>
<td>0.024</td>
<td>0.0008</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>-0.007</td>
<td>-3.74</td>
<td>&lt;0.001</td>
<td>0.007</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.018</td>
<td>-2.84</td>
<td>0.005</td>
<td>0.091</td>
<td>0.003</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>-0.02</td>
<td>-13.07</td>
<td>&lt;0.001</td>
<td>0.006</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-0.004</td>
<td>-0.35</td>
<td>0.728</td>
<td>0.2</td>
<td>0.007</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>-0.039</td>
<td>-23.49</td>
<td>&lt;0.001</td>
<td>0.004</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.0004</td>
<td>0.33</td>
<td>0.744</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.807</td>
<td>11.00</td>
<td>&lt;0.001</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.057</td>
<td>-0.73</td>
<td>0.467</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>0.004</td>
<td>0.05</td>
<td>0.961</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Table 4. The herd and marginal effect estimates for the analysis at county level. The fixed effect estimates (F.E. Est), the random effect estimates (R.E. Est) as well as the p-value for the F.E Est are provided.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Location</th>
<th>F.E. Est</th>
<th>F.E. p-value</th>
<th>R.E. Est</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Marginal Effect</td>
<td>$\Psi_1 = [8.66%, 32%]$</td>
<td>-0.004</td>
<td>&lt;0.001</td>
<td>0.0005</td>
</tr>
<tr>
<td>1st Herd Effect</td>
<td>$b_1 = 32%$</td>
<td>-0.025</td>
<td>&lt;0.001</td>
<td>0.024</td>
</tr>
<tr>
<td>2nd Marginal Effect</td>
<td>$\Psi_2 = [32%, 36%]$</td>
<td>-0.01</td>
<td>&lt;0.001</td>
<td>0.008</td>
</tr>
<tr>
<td>2nd Herd Effect</td>
<td>$b_2 = 36%$</td>
<td>0.008</td>
<td>0.70</td>
<td>0.115</td>
</tr>
<tr>
<td>3rd Marginal Effect</td>
<td>$\Psi_3 = [36%, 47%]$</td>
<td>-0.023</td>
<td>&lt;0.001</td>
<td>0.006</td>
</tr>
<tr>
<td>3rd Herd Effect</td>
<td>$b_3 = 47%$</td>
<td>0.014</td>
<td>0.545</td>
<td>0.29</td>
</tr>
<tr>
<td>4th Marginal Effect</td>
<td>$\Psi_4 = [47%, 49.3%]$</td>
<td>-0.043</td>
<td>&lt;0.001</td>
<td>0.005</td>
</tr>
</tbody>
</table>

4 Discussion

Vaccination has been acknowledged as an effective tool to reduce hospitalization and mortality related to COVID-19 infections, and vaccination campaign has been rolled out in virtually every country that has access to COVID-19 vaccines. Understanding the effect of COVID-19 vaccination in terms of case fatality rate (CFR) reduction has unquestionably profound meaning, for a successful implementation of the COVID-19 vaccination campaign. Drawing on the direct and indirect effects of vaccination from literature, we rename the direct effect as the marginal effect of vaccination and the indirect effect as the herd effect of vaccination, to better describe the nature of those effects in terms of reducing the CFR. Defining the herd and marginal effects also helps build regression models for obtaining their estimates, as those two kinds of effects require different parameterization in the model. Analysis at the national level and county level
for the United States, were then implemented based on datasets containing the daily vaccination coverages and case reports in the U.S. Theme rankings for individual counties from the CDC SVI were also included to explain heterogeneity at county level. Our analysis at national level suggested three different locations (i.e., when the vaccination rate reached 32%, 36% and 47%) for possible herd effects and strong significance for the marginal effects, which was further confirmed by our analysis at county level after controlling for county heterogeneity. Our analyses have demonstrated how COVID-19 vaccination protects against COVID related mortality over the course of COVID-19 vaccination campaign in the U.S. In general, COVID-19 vaccination indeed can significantly reduce the CFR, but its effect is not constant during the vaccination campaign. The estimated breakpoints have divided the vaccination campaign into four different regions based on the vaccination coverage, i.e., $\Psi_1 = [8.66\%, 32\%)$, $\Psi_2 = [32\%, 36\%)$, $\Psi_3 = [36\%, 47\%)$ and $\Psi_4 = [47\%, 49.3\%)$. The marginal effects in those four regions are correspondingly -0.004, -0.01, -0.023 and -0.043, which are all significant. This shows the vaccination can directly result in meaningful reduction in the CFR and thus it should be recommended especially for the unvaccinated population, as the marginal effects largely quantify the reduced risks of mortality that one would benefit from the vaccination if he/she chooses to get vaccinated. We also observe that the sizes of the marginal effects enlarge as the vaccination coverages increases, which suggests that the direct benefit of COVID-19 vaccination is becoming more and more significant as the vaccination coverage in the population increases. Our results also indicate the existence of herd effect, specifically at the threshold 32%. The herd effect at the threshold 32% is statistically significant (-
0.025), which demonstrates the indirect (additional) benefit brought by the vaccination once the vaccination coverage reaches 32% in the population. This implies that one would indirectly benefit from the COVID-19 vaccination even if he/she is not vaccinated as long as the vaccination coverage passes 32%, by a 0.025% reduction in the CFR.

Our results have important implications for the COVID-19 vaccination strategies. First, our findings suggest that vaccination campaign should be rapidly carried out at the initial stage, to trigger the threshold for herd effects, in order to procure additional protection of COVID-19 vaccination against the CFR for the entire population regardless of individual vaccination statuses. This echoes our earlier finding of the significant herd effect at the first breakpoint 32% and is consistent with recommendations offered by the literature [3,4,23,24,25]. It is noteworthy that, a rapid effective implementation at the initial stage can pose considerable logistical challenges for a vaccination campaign [23,26,27]. Therefore, careful resource planning is required for the access, transportation, storage and distribution of vaccines, which has been exemplified by the vaccination campaign in the U.S. [25,28]. Second, eligible unvaccinated individuals should be encouraged (even urged) to get vaccinated at all stages of a vaccination campaign, as the marginal effects were evident across all the local regions defined for the U.S. vaccination campaign in our analysis. More profoundly, we found the whole population would benefit more if more people got vaccinated, as the size of marginal effect was positively correlated with the vaccination coverage in the population. The gain from the marginal effects, on average, also outweighed the gain from the herd effects, as manifested by the Table 4. These key observations suggest that the marginal effect is more important than the herd effect for the protection against COVID mortality [15]. Thus, vaccination strategy should focus on
how to capitalize on the marginal effect, i.e., promote individual vaccination willingness
and accessibility, in order to continuously push for a higher vaccination rate in the
population [15,29]. Based on our results, the goal of a vaccination campaign should be
pursuing a higher vaccination coverage in the population, rather than meeting a
predefined threshold for triggering the herd effect [4,29,30].
Heterogeneity among the U.S. counties in terms of the marginal and herd effects is
considerable. The sizes and even the signs of the marginal and herd effects could vary
across all the counties, which signals that the protection effect of COVID-19 vaccination
is not constant and partially determined by county idiosyncrasy. For example, we took
the social vulnerability index (SVI) into account in our analysis and did find the theme of
household composition & disability significantly was significantly associated with the
CFR after controlling for the vaccination coverage. This indicates that demographical
features of individual county, such as the age distribution and disability proportion, play
vital roles in explaining the heterogeneity existed for the relationship between the
vaccination coverage and the CFR [31]. Although the other three SVI themes, namely
socioeconomic status, minority status & language and housing type & transportation,
were not statistically significant, factors such as environmental conditions [32], political
atmosphere [33] and non-pharmaceutical interventions [34] could contribute to county
heterogeneity, and potentially confound the relationship between the vaccination
coverage and the CFR. Most notably, research has shown that vaccine hesitancy
(willingness) is a key determinant of vaccination coverage, and it potentially mediates the
relationship between the factors influencing the CFR (like SVI) and the CFR itself, and
therefore variation of vaccine hesitancy among the U.S. counties potentially accounts for a significant portion of the county heterogeneity observed in our paper [35-36].

There are limitations in our analysis: We did not investigate the impact of COVID-19 variants on the CFR and the vaccine effectiveness, considering there were different COVID-19 variants (and their lineages and sublineages), such as alpha, delta and omicron, spreading during our study period, as we can hardly identify the boundaries of the spreading period of each variant from the data. For the similar reason, the potential impact of different brands of vaccines (such as BioNTech and Moderna) was also not considered in our model, as the data did not contain information about the number of administered doses of every specific brand. Most importantly, our model treats the breakpoints as the fixed values across all the counties, which may not be true as the breakpoints could vary across different counties as a result of unique evolvement of vaccination campaign in individual counties. Unfortunately, allowing each county to have its own breakpoints would require a huge number of parameters and a complex Bayesian model, which goes beyond the scope of this paper [37]. Therefore, further robustness and sensitivity analyses may be warranted [38-41].

To summarize, we have shown the existence of the herd effects via a segmented regression model. Specifically, we identified three different breakpoints that represented the locations of the herd effects. Accounting for county heterogeneity, we found one of the three herd effects to be statistically significant, and it suggested that additional indirect benefit of COVID-19 vaccination may exist at the earlier stage of a vaccination campaign. We also found the marginal effect size varied at different stages of the vaccination campaign, and specifically the marginal (direct) benefit of COVID-19
vaccination likely became larger as the vaccination coverage increased. Our findings
demonstrate that the herd and marginal effects should be carefully differentiated and
assessed in analyzing vaccination data, to better inform vaccination campaign strategies
as well as evaluate vaccination effectiveness.
References


[38] Li T. *The Bayesian Paradigm of Robustness Indices of Causal Inferences*. Michigan State University; 2018.

