

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

3 Tenglong Li<sup>1\*</sup>, Zilong Wang<sup>2</sup>, Shuyue He<sup>1</sup> and Ying Chen<sup>1</sup>

4 <sup>1</sup> Wisdom Lake Academy of Pharmacy, Xi'an Jiaotong-Liverpool University

5 <sup>2</sup> Department of Financial and Actuarial Mathematics, Xi'an Jiaotong-Liverpool  
6 University

7 \* To whom correspondence may be addressed: [Tenglong.Li@xjtlu.edu.cn](mailto:Tenglong.Li@xjtlu.edu.cn)

8

9 **Abstract**

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

26 *Keywords:* COVID-19 vaccine, herd effect, marginal effect, case fatality rate, segmented  
27 regression.

## 28 **1 Introduction**

29 The world has been living in the tunnel of COVID-19 pandemic since the outbreak in  
30 2019, yet without a clear idea about its outlet. A great hope has been placed in COVID-  
31 19 vaccines to end the pandemic, as clinical trial results suggested COVID-19  
32 vaccination can effectively prevent symptomatic infections especially severe symptoms,  
33 which protects against mortality associated with infections [1-3]. For this reason, public  
34 demand for COVID-19 vaccines was fervent and vaccination campaigns were initiated all  
35 over the world, for an early safe vaccine supply for populations at risk as well as a  
36 massive vaccine supply to match the public's demand [4-5]. For an example, the Food  
37 and Drug Administration (FDA) issued emergency use authorizations (EUA) for Pfizer-  
38 BioNTech and Moderna COVID-19 vaccines in December 2020, which marked the  
39 beginning of the vaccination campaign in the U.S. COVID-19 vaccines were then first  
40 allocated for populations at risk, the elderly population (age 65+) and the frontline  
41 (mostly healthcare and education) workers. After president Biden announced that all  
42 Americans would be eligible for COVID-19 vaccines by May 1<sup>st</sup>, 2021, the vaccination  
43 campaign was further accelerated [6]. Booster doses of COVID-19 vaccines were  
44 introduced to restore the level of protection (antibody) eroded by time [7-9]. By  
45 November 24, 2022, more than 80% of Americans have received at least one dose and  
46 more than 68% of Americans have completed a primary series of COVID-19 vaccine  
47 [10]. Literature has reported that the vaccination coverage is negatively associated with

48 case fatality rate (CFR), which refers to the mortality rate among those who are infected  
49 (i.e., confirmed COVID cases) [11-12].

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

68 The above concepts of the direct and indirect effects should be contextualized in the  
69 investigation of the impact of COVID-19 vaccination on case fatality rate (CFR). The  
70 direct effect could be interpreted as the reduction in CFR associated with one unit/percent

71 increase in the vaccination coverage, i.e., the direct effect evaluates the marginal gain  
72 during a vaccination campaign. For this reason, the direct effect is referred to as the  
73 marginal effect in this paper. The indirect effect could be interpreted as the additional  
74 reduction in CFR if the vaccination coverage passes certain unknown thresholds, i.e., the  
75 indirect effect quantifies the additional gain potentially due to herd immunity in the  
76 process of vaccinating a target population. To better characterize its nature, the indirect  
77 effect is referred to as the herd effect in this paper. It's particularly important to  
78 disentangle the herd effect from the marginal effect, for the following three reasons: First,  
79 the marginal and herd effects address different scientific questions with regard to distinct  
80 groups of people (i.e., the vaccinated individuals versus the unvaccinated individuals).  
81 Second, as discussed earlier, there are underlying thresholds for triggering the herd  
82 effects, and those thresholds essentially delineate different stages in a vaccination  
83 campaign where the marginal and herd effects may not be constant across those stages.  
84 Third, given the aforementioned two reasons, a deeper knowledge about the protection  
85 effect of vaccination is likely gained by learning the marginal and herd effects, and  
86 vaccination strategies could be optimized for a target population based on such  
87 knowledge. Unfortunately, we haven't seen research on this important topic so far.  
88 Our goal in this paper is to estimate the herd and marginal effects based on a dataset from  
89 the U.S. Centers for Disease Control and Prevention (CDC), which records various  
90 vaccination coverages for each U.S. county daily [10]. We hypothesize that both the herd  
91 and marginal effects exist and are significantly negative for modeling CFR. The  
92 segmented regression is employed first to identify the breakpoints which are considered  
93 as the thresholds for triggering the herd effect, based on data of all the U.S. counties

94 included in our study. With the identified breakpoints, we estimate the herd and marginal  
95 effects at national level using segmented regression and at county level using mixed  
96 model. Data on social vulnerability index (SVI) for individual counties is also included to  
97 control for health disparities due to sociodemographic factors at county level [18].  
98 Heterogeneity among individual counties is further evaluated by the random effects  
99 associated with the herd and marginal effects among in a mixed model.

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

## 107 **2 Materials and Methods**

### 108 **2.1 Data**

109 Our data comes from three different sources. The US vaccine administration and equity  
110 dataset is obtained from the CDC website and has vaccination coverages of the general  
111 population and its subpopulations (defined by age) recorded daily at county level [10].  
112 The percent of people who completed a primary series of vaccination in the general  
113 population was extracted from the dataset and served as the main covariate in our model.  
114 The daily CFR at county level was calculated as the ratio between the daily count of  
115 deaths and the daily count of COVID-19 cases, based on the time series summary tables  
116 of COVID-19 deaths and confirmed cases, which were accessed from the COVID-19

117 data repository by the Center for Systems Science and Engineering (CSSE) at John  
118 Hopkins University [19]. To further control for county heterogeneity, we used a dataset  
119 from the Centers for Disease Control and Prevention Social Vulnerability Index (CDC  
120 SVI) database, created by the Geospatial research, Analysis & Services Program under  
121 the Agency for Toxic Substances and Disease Registry [18]. The CDC SVI database was  
122 established to help health officials and emergency response planners identify counties  
123 that will most likely need support before, during, and after a hazardous event. CDC SVI  
124 ranks counties on 15 social factors and further groups them into four themes, namely  
125 socioeconomic status, household composition & disability, minority status & language,  
126 and housing type & transportation [20]. We chose to use the theme-specific ranking  
127 which was constructed by summing the percentiles of the factors under each theme. The  
128 theme-specific ranking was set in the range from 0 to 1, with higher values indicating  
129 greater vulnerability.

130 The vaccination coverages and daily CFR for the period between March 11<sup>th</sup>, 2021 and  
131 Jan 26<sup>th</sup>, 2022 were selected. We chose March 11<sup>th</sup>, 2021 as it was the date when  
132 president Biden announced that COVID-19 vaccine would be available for all American  
133 adults by May 1<sup>st</sup>, 2021, an event marked the beginning of massive vaccination campaign  
134 in the U.S. We chose Jan 26<sup>th</sup>, 2022 as the ending date of our study as it was reported on  
135 this date that Omicron variant accounted for 99.9% of the new infections. This would  
136 alleviate the concern of potential confounding effect of Omicron variant regarding the  
137 relationship between the vaccination coverage and CFR. 31 counties with missing values  
138 on county FIPS code, vaccination coverages, the CDC SVI or CFR were excluded, and  
139 the final dataset has 1001098 observations clustered by 3109 U.S. counties. To prepare

140 the dataset for analysis at national level, we further extracted the average CFR and  
141 average vaccination coverage (i.e., the percent of people who completed a primary series  
142 of COVID-19 vaccine) across all the counties in our dataset for each day during our study  
143 period.

## 144 **2.2 Models**

145 Segmented regression models were employed to estimate the herd and marginal effects.  
146 Segmented regression is very similar to ordinary regression, with the only difference that  
147 regression coefficients should be estimated repeatedly for different local regions whose  
148 boundaries are defined by breakpoints, which represent the thresholds of structural  
149 changes in regression models [21-22]. Typically, the first step is to determine the number  
150 of breakpoints, which can be achieved by a model selection alike procedure, i.e., models  
151 with different number of breakpoints are compared in terms of their model fit indices  
152 (such as AIC or BIC) to determine the optimized number of breakpoints. The second step  
153 is to estimate the locations of breakpoints given the number of breakpoints. The third  
154 step, based on the estimated breakpoints, is then to fit regression models to different local  
155 regions separated by the breakpoints. Normally, one would expect all regression  
156 coefficients to be changeable across different regions, unless otherwise specified.  
157 For our analysis at national level, we intend to examine the relationship between CFR  
158 and the vaccination coverage, based on the dataset comprising only the average  
159 vaccination coverage and CFR in the U.S. The following regression model is formulated  
160 for the analysis at national level:

$$y_t = \beta_0 + \beta_1 X_t + \sum_{k=1}^m \alpha_k I_{\{X_t \in \Psi_k\}} + \sum_{k=1}^m \delta_k X_t I_{\{X_t \in \Psi_k\}} + \varepsilon_t \quad (1)$$

161 where  $y_t$  and  $X_t$  denote the average CFR and vaccinate coverage in the U.S on day  $t$ . The  
162 model (1) is built on the estimated breakpoints  $b_1 < b_2 < \dots < b_m$ , which implies there  
163 are  $m + 1$  different local regions and  $m$  different breakpoints in total (except  $b_0$  and  
164  $b_{m+1}$  which are the minimum and maximum of  $X_t$ ). The local regions separated by the  
165 breakpoints are denoted by  $\Psi_k = [b_k, b_{k+1})$  for  $k = 1, 2, \dots, m$ . The reference local  
166 region  $\Psi_0$ , although omitted from the model (1), refers to the local region  $\Psi_0 = [b_0, b_1)$ .  
167 The indicator function  $I_{\{X_t \in \Psi_k\}}$  creates the dummy variable which assigns value 1 if the  
168 value of  $X_t$  falls in the local region  $\Psi_k$  and 0 otherwise, which operationally divides the  
169 range of  $X_t$  into the local regions. The marginal effects in those local regions are  
170 characterized by  $\beta_1$  for the reference region and  $\beta_1 + \delta_k$  for the local region  $\Psi_k$ , and  
171 these parameters quantify the marginal gain/drop in CFR if the vaccination coverage  
172 increases by one percent. The herd effects for the local region  $\Psi_k$  relative to its previous  
173 region are characterized by  $\alpha_k - \alpha_{k-1}$  (for  $\Psi_1$  it is just  $\alpha_1$ ), as  $\alpha_k$  quantifies the  
174 additional gain/drop in CFR if the vaccination coverage passes the threshold  $b_k$ ,  
175 compared to the intercept term  $\beta_0$  in the reference region  $\Psi_0$ .  
176 The breakpoints  $b_k$   $k = 1, 2, \dots, m$ , are estimated based on the model (1) and the dataset  
177 for the analysis at national level (i.e., with only average daily CFR and vaccination rate in  
178 the U.S.). Naturally, they reflect the structural changes in the relationship between CFR  
179 and the vaccination coverage in general, and they can be applied to the analysis at county  
180 level where we use the longitudinal data (322 days) for all the counties (3109 counties),  
181 along with the CDC SVI indicators for explaining county heterogeneity. We build the  
182 following mixed model for the analysis at county level:



$$\begin{aligned}
 y_{it} = & \beta_0 + \beta_1 X_{it} + \sum_{k=1}^m \alpha_k I_{\{X_{it} \in \Psi_m\}} + \sum_{k=1}^m \delta_k X_{it} I_{\{X_{it} \in \Psi_m\}} + \gamma Z_i + u_1 X_{it} + u_0 \\
 & + \sum_{k=1}^m \tilde{\alpha}_k I_{\{X_{it} \in \Psi_m\}} + \sum_{k=1}^m \tilde{\delta}_k X_{it} I_{\{X_{it} \in \Psi_m\}} + \varepsilon_{it}
 \end{aligned}
 \tag{2}$$

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

## 196 **3 Results**

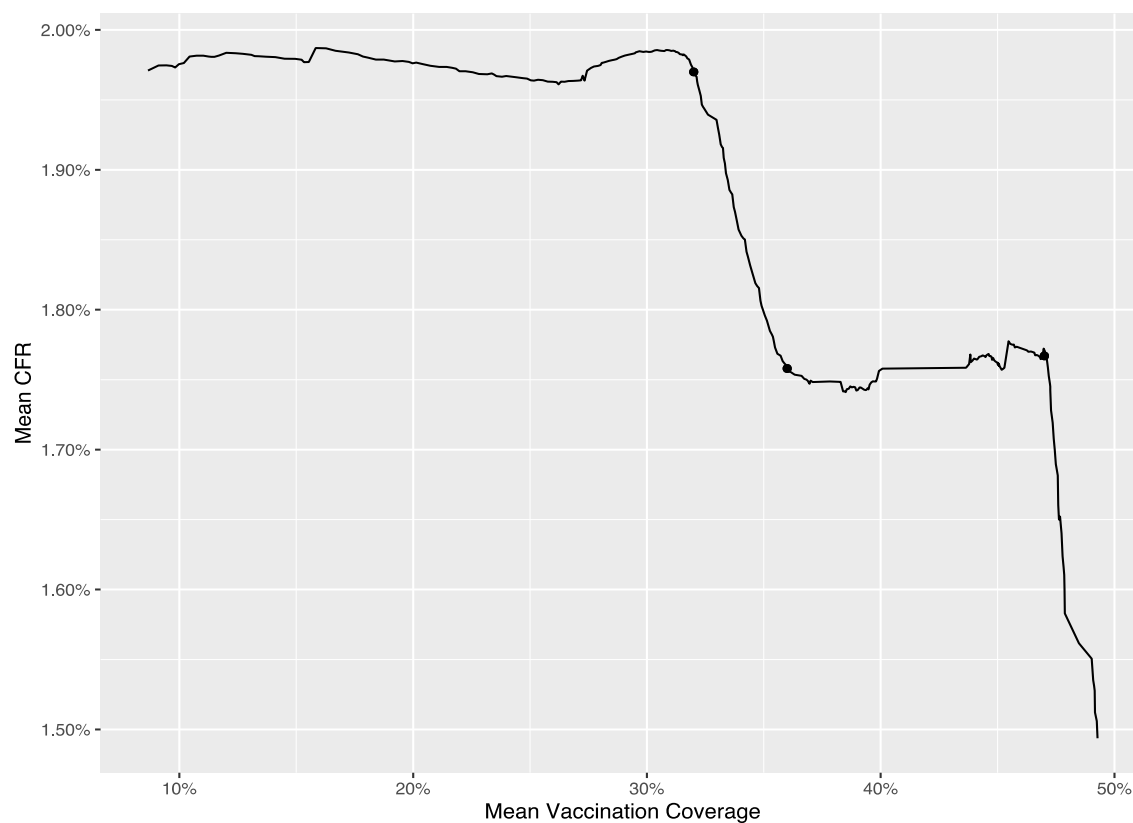
### 197 **3.1 The results of the analysis at national level**

198 As mentioned above, the dataset used for the analysis at national level has two variables,  
 199 i.e., average daily CFR and average vaccination coverage in the U.S.. The breakpoints  
 200 were estimated based on this dataset using the “segmented” package in R (version 4.2.0)  
 201 [22]. To avoid overfitting, we set the maximum number of the breakpoints as 3, based on

202 the curve between the average daily CFR and the average daily vaccination coverage  
203 depicted in Figure 1. The segmented package then does an automatic selection of the  
204 number of breakpoints based on BIC, and it estimated the locations of the breakpoints  
205 conditional on the optimized number of the breakpoints. The estimated breakpoints were  
206 superimposed on the curve in Figure 1, to further validate those estimates align with the  
207 observed structural changes.

208 The breakpoints were estimated as 32%, 36% and 47%, which suggested that the herd  
209 effect of vaccination may be associated with the thresholds of 32%, 36% and 47% in the  
210 vaccination coverage. Based on those breakpoints, we have four different local regions,  
211 namely  $\Psi_1 = [8.66\%, 32\%]$ ;  $\Psi_2 = [32\%, 36\%]$ ;  $\Psi_3 = [36\%, 47\%]$ ;  $\Psi_4 =$   
212  $[47\%, 49.28\%]$ , with the minimum and maximum of the average daily vaccination  
213 coverages as 8.66% and 49.28% respectively. Table 1 lists the estimates of the regression  
214 coefficients based on the model (1). We further calculated the herd and marginal effect  
215 estimates which are tabulated in Table 2. The marginal effect in the first local region  $\Psi_1$   
216 was insignificant, which suggested that the drop in the CFR per percent increase in the  
217 vaccination coverage was not significantly different from 0, if the vaccination coverage  
218 did not surpass 32%. The herd effect at the threshold 32% was also insignificant, which  
219 was largely due to the insignificant marginal effect in the region  $\Psi_1$ . We found a  
220 significant marginal effect in the second local region  $\Psi_2$  (-0.057), which indicated that  
221 there was a drop of 0.057 percent in the CFR for every percent increase in the vaccination  
222 coverage in this region, evidencing that the protection effect of COVID vaccination  
223 against mortality. In addition, the herd effect at the threshold 36% was significant too (-  
224 0.233), suggesting that there was a further drop of 0.233 percent in the CFR besides the

225 marginal CFR reduction per percent increase in the vaccination coverage. In the third  
226 local region  $\Psi_3$ , however, we observed a slight positive marginal effect in the CFR  
227 (0.003), which means the marginal gain of vaccination (in terms of the reduction in CFR)  
228 disappeared and vaccination was somehow harmful for protecting against mortality.  
229 Correspondingly, the herd effect at the threshold 47% was also positively significant  
230 (0.009), suggesting again that vaccination was not helpful at this stage. The marginal  
231 effect in the fourth local region  $\Psi_4$  was strongly negative, specifically there was a drop of  
232 0.115 percent in the CFR associated with every percent increase in the vaccination  
233 coverage at this stage.



234

235 Figure 1. The relationship between the average vaccination coverage and the average  
236 CFR in the U.S. The solid dots represent the breakpoints estimated by the “segmented”  
237 package in R.  
238

239 Table 1. The regression model parameter estimates for the analysis at national level.

Parameter	Estimate	T Ratio	p-value
$\beta_0$	1.976	624.70	<0.001
$\beta_1$	0.000	0.28	0.78
$\alpha_1$	-0.003	-0.90	0.37
$\delta_1$	-0.057	-45.32	<0.001
$\alpha_2$	-0.236	-80.67	<0.001
$\delta_2$	0.003	7.50	<0.001
$\alpha_3$	-0.227	-55.09	<0.001
$\delta_3$	-0.115	-47.03	<0.001

240

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

Effect	Location	Estimate	p-value
1 <sup>st</sup> Marginal Effect	$\Psi_1 = [8.66\%, 32\%]$	0.000	0.78
1 <sup>st</sup> Herd Effect	$b_1 = 32\%$	-0.003	0.37
2 <sup>nd</sup> Marginal Effect	$\Psi_2 = [32\%, 36\%]$	-0.057	<0.001
2 <sup>nd</sup> Herd Effect	$b_2 = 36\%$	-0.233	<0.001
3 <sup>rd</sup> Marginal Effect	$\Psi_3 = [36\%, 47\%]$	0.003	<0.001
3 <sup>rd</sup> Herd Effect	$b_3 = 47\%$	0.009	0.02
4 <sup>th</sup> Marginal Effect	$\Psi_4 = [47\%, 49.28\%]$	-0.115	<0.001

242

### 243 **3.2 The results of the analysis at county level**

244 We further investigated the marginal and herd effects of COVID vaccination based on an  
245 analysis at county level, where the daily CFR and vaccination coverages from March  
246 11<sup>th</sup>, 2021 to Jan 26<sup>th</sup>, 2022 as well as the CDC SVI rankings for 3109 U.S. counties were  
247 used. The estimated breakpoints of 32%, 36% and 47%, obtained based on the analysis at  
248 national level, were adopted for our analysis at county level. The mixed model (2) was  
249 employed to account for the clustered data at county level, and its fixed and random  
250 effect estimates are tabulated in the Table 3. Furthermore, the estimates of herd and  
251 marginal effects, as well as their corresponding random effect estimates, are listed in the  
252 Table 4. To determine the significance of the random effects, we compared the full model  
253 (i.e., the model (2)) with two different reduced models (one without the random effects  
254 associated with all the marginal effects, i.e.,  $u_1, \tilde{\delta}_1, \tilde{\delta}_2, \tilde{\delta}_3$ ; another one without the  
255 random effect associated with the first marginal effect only, i.e.,  $u_1$ ), and the resultant  
256 tests gave p-values smaller than 0.001, suggesting that it was necessary to include  
257 random effects for all the marginal and herd effect parameters.

258 Across all the U.S. counties in our data, the marginal effect was significantly negative in  
259 the first local region  $\Psi_1$  (i.e., when the vaccination coverage was between 8.66% and  
260 32%), specifically one percent increase in the vaccination coverage was associated with  
261 0.004 percent drop in the CFR. The first herd effect at the threshold of 32% was -0.025  
262 and significant, meaning there was an additional drop of 0.025 percent in the CFR as the  
263 vaccination coverage reached 32%, beyond the marginal effect observed in  $\Psi_1$ . The  
264 marginal effect in the second local region  $\Psi_2$  was also significantly negative (-0.01),  
265 which showed that there was 0.01 percent drop in the CFR per one percent increase in the

266 vaccination coverage, when the vaccination coverage was between 32% and 36%. The  
267 second herd effect, however, was overall insignificant, which suggested additional  
268 protection effect at the threshold 36% may not exist. Similarly, we found significant  
269 marginal effect (-0.023) for the third local region  $\Psi_3$  but insignificant herd effect at the  
270 threshold 47%. The marginal effect within the fourth local region  $\Psi_4$  was the strongest,  
271 as every percent increase in the vaccination coverage was associated with 0.043 percent  
272 reduction in the CFR, if the vaccination coverage surpassed 47%.

273 Furthermore, heterogeneity among the U.S. counties regarding the herd and marginal  
274 effect estimates was evident. For the first herd effect (at 32%), the fixed effect estimate  
275 was -0.025 with a random effect of 0.024, and this means roughly 56% of the counties  
276 had negative herd effects as expected, but the other 44% of the counties could have no  
277 herd effect or even positive herd effects at the threshold 32%. For the second and third  
278 herd effect (at 36% and 47% respectively), roughly 49% of the counties have negative  
279 herd effects, which further demonstrated that those two herd effects were not significant  
280 among the counties. Regarding the marginal effects: although their fixed effect estimates  
281 were all very significant (p-value < 0.001), their random effect estimates suggested the  
282 fourth marginal effect was the strongest one (was negative in 73% of the counties). The  
283 first, second and third marginal effects were negative in approximately 57%, 54% and  
284 62% of the U.S. counties. All taken, the protection effect of COVID vaccination was  
285 confirmed in general and for the majority of the U.S. counties, while substantial  
286 heterogeneity that defined the size and the validity of the protection effect for individual  
287 county still existed. We also found that only one CDC SVI theme ranking, i.e., rankings  
288 on household composition & disability, could help explain the county heterogeneity.

289 Unsurprisingly, this CDC SVI theme ranking was positively related to CFR, and  
 290 specifically one percentile rise in the theme ranking could result in 0.8 percent increase in  
 291 CFR.

292

293 Table 3. The mixed model parameter estimates pertaining to fixed effects (F.E.) and  
 294 random effects (R.E.) for the analysis at county level. The standard errors (S.E.) of the  
 295 random effects (R.E.) are also provided.

Parameter	F.E. Est	F.E. T Ratio	F.E. p-value	R.E. Est	S.E. of R.E.
$\beta_0$	1.64	28.88	<0.001	1.33	0.03
$\beta_1$	-0.004	-8.85	<0.001	0.0005	0.00001
$\alpha_1$	-0.025	-7.35	<0.001	0.024	0.0008
$\delta_1$	-0.007	-3.74	<0.001	0.007	0.0002
$\alpha_2$	-0.018	-2.84	0.005	0.091	0.003
$\delta_2$	-0.02	-13.07	<0.001	0.006	0.0002
$\alpha_3$	-0.004	-0.35	0.728	0.2	0.007
$\delta_3$	-0.039	-23.49	<0.001	0.004	0.0002
$\gamma_1$	0.0004	0.33	0.744	n/a	n/a
$\gamma_2$	0.807	11.00	<0.001	n/a	n/a
$\gamma_3$	-0.057	-0.73	0.467	n/a	n/a
$\gamma_4$	0.004	0.05	0.961	n/a	n/a

296

297

298

299

300

301 Table 4. The herd and marginal effect estimates for the analysis at county level. The fixed  
 302 effect estimates (F.E. Est), the random effect estimates (R.E. Est) as well as the p-value  
 303 for the F.E Est are provided.

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

304

#### 305 **4 Discussion**

306 Vaccination has been acknowledged as an effective tool to reduce hospitalization and  
 307 mortality related to COVID-19 infections, and vaccination campaign has been rolled out  
 308 in virtually every country that has access to COVID-19 vaccines. Understanding the  
 309 effect of COVID-19 vaccination in terms of case fatality rate (CFR) reduction has  
 310 unquestionably profound meaning, for a successful implementation of the COVID-19  
 311 vaccination campaign. Drawing on the direct and indirect effects of vaccination from  
 312 literature, we rename the direct effect as the marginal effect of vaccination and the  
 313 indirect effect as the herd effect of vaccination, to better describe the nature of those  
 314 effects in terms of reducing the CFR. Defining the herd and marginal effects also helps  
 315 build regression models for obtaining their estimates, as those two kinds of effects require  
 316 different parameterization in the model. Analysis at the national level and county level



317 for the United States, were then implemented based on datasets containing the daily  
318 vaccination coverages and case reports in the U.S. Theme rankings for individual  
319 counties from the CDC SVI were also included to explain heterogeneity at county level.  
320 Our analysis at national level suggested three different locations (i.e., when the  
321 vaccination rate reached 32%, 36% and 47%) for possible herd effects and strong  
322 significance for the marginal effects, which was further confirmed by our analysis at  
323 county level after controlling for county heterogeneity.

324 Our analyses have demonstrated how COVID-19 vaccination protects against COVID  
325 related mortality over the course of COVID-19 vaccination campaign in the U.S. In  
326 general, COVID-19 vaccination indeed can significantly reduce the CFR, but its effect is  
327 not constant during the vaccination campaign. The estimated breakpoints have divided  
328 the vaccination campaign into four different regions based on the vaccination coverage,  
329 i.e.,  $\Psi_1 = [8.66\%, 32\%)$ ,  $\Psi_2 = [32\%, 36\%)$ ,  $\Psi_3 = [36\%, 47\%)$  and  $\Psi_4 =$   
330  $[47\%, 49.3\%)$ . The marginal effects in those four regions are correspondingly -0.004, -  
331 0.01, -0.023 and -0.043, which are all significant. This shows the vaccination can directly  
332 result in meaningful reduction in the CFR and thus it should be recommended especially  
333 for the unvaccinated population, as the marginal effects largely quantify the reduced risks  
334 of mortality that one would benefit from the vaccination if he/she chooses to get  
335 vaccinated. We also observe that the sizes of the marginal effects enlarge as the  
336 vaccination coverages increases, which suggests that the direct benefit of COVID-19  
337 vaccination is becoming more and more significant as the vaccination coverage in the  
338 population increases. Our results also indicate the existence of herd effect, specifically at  
339 the threshold 32%. The herd effect at the threshold 32% is statistically significant (-

340 0.025), which demonstrates the indirect (additional) benefit brought by the vaccination  
341 once the vaccination coverage reaches 32% in the population. This implies that one  
342 would indirectly benefit from the COVID-19 vaccination even if he/she is not vaccinated  
343 as long as the vaccination coverage passes 32%, by a 0.025% reduction in the CFR.  
344 Our results have important implications for the COVID-19 vaccination strategies. First,  
345 our findings suggest that vaccination campaign should be rapidly carried out at the initial  
346 stage, to trigger the threshold for herd effects, in order to procure additional protection of  
347 COVID-19 vaccination against the CFR for the entire population regardless of individual  
348 vaccination statuses. This echoes our earlier finding of the significant herd effect at the  
349 first breakpoint 32% and is consistent with recommendations offered by the literature  
350 [3,4,23,24,25]. It is noteworthy that, a rapid effective implementation at the initial stage  
351 can pose considerable logistical challenges for a vaccination campaign [23,26,27].  
352 Therefore, careful resource planning is required for the access, transportation, storage and  
353 distribution of vaccines, which has been exemplified by the vaccination campaign in the  
354 U.S. [25,28]. Second, eligible unvaccinated individuals should be encouraged (even  
355 urged) to get vaccinated at all stages of a vaccination campaign, as the marginal effects  
356 were evident across all the local regions defined for the U.S. vaccination campaign in our  
357 analysis. More profoundly, we found the whole population would benefit more if more  
358 people got vaccinated, as the size of marginal effect was positively correlated with the  
359 vaccination coverage in the population. The gain from the marginal effects, on average,  
360 also outweighed the gain from the herd effects, as manifested by the Table 4. These key  
361 observations suggest that the marginal effect is more important than the herd effect for  
362 the protection against COVID mortality [15]. Thus, vaccination strategy should focus on

363 how to capitalize on the marginal effect, i.e., promote individual vaccination willingness  
364 and accessibility, in order to continuously push for a higher vaccination rate in the  
365 population [15,29]. Based on our results, the goal of a vaccination campaign should be  
366 pursuing a higher vaccination coverage in the population, rather than meeting a  
367 predefined threshold for triggering the herd effect [4,29,30].

368 Heterogeneity among the U.S. counties in terms of the marginal and herd effects is  
369 considerable. The sizes and even the signs of the marginal and herd effects could vary  
370 across all the counties, which signals that the protection effect of COVID-19 vaccination  
371 is not constant and partially determined by county idiosyncrasy. For example, we took  
372 the social vulnerability index (SVI) into account in our analysis and did find the theme of  
373 household composition & disability significantly was significantly associated with the  
374 CFR after controlling for the vaccination coverage. This indicates that demographical  
375 features of individual county, such as the age distribution and disability proportion, play  
376 vital roles in explaining the heterogeneity existed for the relationship between the  
377 vaccination coverage and the CFR [31]. Although the other three SVI themes, namely  
378 socioeconomic status, minority status & language and housing type & transportation,  
379 were not statistically significant, factors such as environmental conditions [32], political  
380 atmosphere [33] and non-pharmaceutical interventions [34] could contribute to county  
381 heterogeneity, and potentially confound the relationship between the vaccination  
382 coverage and the CFR. Most notably, research has shown that vaccine hesitancy  
383 (willingness) is a key determinant of vaccination coverage, and it potentially mediates the  
384 relationship between the factors influencing the CFR (like SVI) and the CFR itself, and

385 therefore variation of vaccine hesitancy among the U.S. counties potentially accounts for  
386 a significant portion of the county heterogeneity observed in our paper [35-36].  
387 There are limitations in our analysis: We did not investigate the impact of COVID-19  
388 variants on the CFR and the vaccine effectiveness, considering there were different  
389 COVID-19 variants (and their lineages and sublineages), such as alpha, delta and  
390 omicron, spreading during our study period, as we can hardly identify the boundaries of  
391 the spreading period of each variant from the data. For the similar reason, the potential  
392 impact of different brands of vaccines (such as BioNTech and Moderna) was also not  
393 considered in our model, as the data did not contain information about the number of  
394 administered doses of every specific brand. Most importantly, our model treats the  
395 breakpoints as the fixed values across all the counties, which may not be true as the  
396 breakpoints could vary across different counties as a result of unique evolvement of  
397 vaccination campaign in individual counties. Unfortunately, allowing each county to have  
398 its own breakpoints would require a huge number of parameters and a complex Bayesian  
399 model, which goes beyond the scope of this paper [37]. Therefore, further robustness and  
400 sensitivity analyses may be warranted [38-41].  
401 To summarize, we have shown the existence of the herd effects via a segmented  
402 regression model. Specifically, we identified three different breakpoints that represented  
403 the locations of the herd effects. Accounting for county heterogeneity, we found one of  
404 the three herd effects to be statistically significant, and it suggested that additional  
405 indirect benefit of COVID-19 vaccination may exist at the earlier stage of a vaccination  
406 campaign. We also found the marginal effect size varied at different stages of the  
407 vaccination campaign, and specifically the marginal (direct) benefit of COVID-19

408 vaccination likely became larger as the vaccination coverage increased. Our findings  
409 demonstrate that the herd and marginal effects should be carefully differentiated and  
410 assessed in analyzing vaccination data, to better inform vaccination campaign strategies  
411 as well as evaluate vaccination effectiveness.

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

## References

- 432 [1] Dal-Ré R, Bekker LG, Gluud C, et al. Ongoing and future COVID-19 vaccine clinical  
433 trials: challenges and opportunities. *Lancet Infect Dis*. 2021;21(11):e342-e347.  
434 doi:10.1016/S1473-3099(21)00263-2  
435
- 436 [2] Yan ZP, Yang M, Lai CL. COVID-19 vaccines: A review of the safety and efficacy  
437 of current clinical trials. *Pharmaceuticals (Basel)*. 2021;14(5). doi:10.3390/ph14050406  
438
- 439 [3] Ahmed S, Khan S, Imran I, et al. Vaccine development against COVID-19: Study  
440 from pre-clinical phases to clinical trials and global use. *Vaccines (Basel)*. 2021;9(8):836.  
441 doi:10.3390/vaccines9080836  
442
- 443 [4] World Health Organization. *Global Covid-19 Vaccination Strategy in a Changing*  
444 *World.*; 2022. [https://www.who.int/publications/m/item/global-covid-19-vaccination-](https://www.who.int/publications/m/item/global-covid-19-vaccination-strategy-in-a-changing-world--july-2022-update)  
445 [strategy-in-a-changing-world--july-2022-update](https://www.who.int/publications/m/item/global-covid-19-vaccination-strategy-in-a-changing-world--july-2022-update)  
446
- 447 [5] Mathieu E, Ritchie H, Ortiz-Ospina E, et al. A global database of COVID-19  
448 vaccinations. *Nat Hum Behav*. 2021;5(7):947-953. doi:10.1038/s41562-021-01122-8  
449
- 450 [6] Hodge JG. Nationalizing public health emergency legal responses. *J Law Med Ethics*.  
451 2021;49(2):315-320. doi:10.1017/jme.2021.45  
452
- 453 [7] Bar-On YM, Goldberg Y, Mandel M, et al. Protection of BNT162b2 vaccine booster  
454 against Covid-19 in Israel. *N Engl J Med*. 2021;385(15):1393-1400.  
455 doi:10.1056/NEJMoa2114255  
456
- 457 [8] Mahase E. Covid-19 booster vaccines: What we know and who's doing what. *BMJ*.  
458 2021;374:n2082. doi:10.1136/bmj.n2082  
459
- 460 [9] Andrews N, Stowe J, Kirsebom F, et al. Effectiveness of COVID-19 booster vaccines  
461 against COVID-19-related symptoms, hospitalization and death in England. *Nat Med*.  
462 2022;28(4):831-837. doi:10.1038/s41591-022-01699-1  
463
- 464 [10] Centers for Disease Control and Prevention. COVID Data Tracker. Accessed  
465 November 24, 2022. <https://covid.cdc.gov/covid-data-tracker>.  
466
- 467 [11] Zhao S, Lou J, Cao L, et al. Differences in the case fatality risks associated with  
468 SARS-CoV-2 Delta and non-Delta variants in relation to vaccine coverage: An early  
469 ecological study in the United Kingdom. *Infect Genet Evol*. 2022;97(105162):105162.  
470 doi:10.1016/j.meegid.2021.105162  
471
- 472 [12] Lee YC, Chang KY, Mirsaeidi M. Association of COVID-19 case-fatality rate with  
473 state health disparity in the United States. *Front Med (Lausanne)*. 2022;9:853059.  
474 doi:10.3389/fmed.2022.853059  
475

- 476 [13] Lipsitch M, Dean NE. Understanding COVID-19 vaccine efficacy. *Science*.  
477 2020;370(6518):763-765. doi:10.1126/science.abe5938  
478
- 479 [14] Halloran ME, Haber M, Longini IM Jr, Struchiner CJ. Direct and indirect effects in  
480 vaccine efficacy and effectiveness. *Am J Epidemiol*. 1991;133(4):323-331.  
481 doi:10.1093/oxfordjournals.aje.a115884  
482
- 483 [15] Morens DM, Folkers GK, Fauci AS. The concept of classical herd immunity may  
484 not apply to COVID-19. *J Infect Dis*. 2022;226(2):195-198. doi:10.1093/infdis/jiac109  
485
- 486 [16] Randolph HE, Barreiro LB. Herd immunity: Understanding COVID-19. *Immunity*.  
487 2020;52(5):737-741. doi:10.1016/j.immuni.2020.04.012  
488
- 489 [17] Gallagher ME, Sieben AJ, Nelson KN, et al. Indirect benefits are a crucial  
490 consideration when evaluating SARS-CoV-2 vaccine candidates. *Nat Med*. 2021;27(1):4-  
491 5. doi:10.1038/s41591-020-01172-x  
492
- 493 [18] Flanagan BE, Hallisey EJ, Adams E, Lavery A. Measuring community vulnerability  
494 to natural and anthropogenic hazards: The centers for disease control and prevention's  
495 social vulnerability index. *J Environ Health*. 2018;80(10):34-36.  
496
- 497 [19] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19  
498 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1  
499
- 500 [20] Place and Health, Agency for Toxic Substances and Disease Registry. *CDC SVI*  
501 *2018 Documentation*.; 2020. Accessed June 7, 2022.  
502 [https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI\\_documentation\\_2018.](https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2018.html)  
503 [html](https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2018.html).  
504
- 505 [21] Muggeo VMR. Estimating regression models with unknown breakpoints. *Stat Med*.  
506 2003;22(19):3055-3071. doi:10.1002/sim.1545  
507
- 508 [22] Muggeo VMR. Segmented: an R package to fit regression models with broken-line  
509 relationships. *R news*. 2008;8(1):20-25.  
510
- 511 [23] McKee M, Rajan S. What can we learn from Israel's rapid roll out of COVID 19  
512 vaccination? *Isr J Health Policy Res*. 2021;10(5):1-4.  
513
- 514 [24] Sah P, Vilches TN, Moghadas SM, et al. Accelerated vaccine rollout is imperative to  
515 mitigate highly transmissible COVID-19 variants. *EClinicalMedicine*.  
516 2021;35(100865):100865. doi:10.1016/j.eclinm.2021.100865  
517
- 518 [25] Moghadas SM, Sah P, Fitzpatrick MC, et al. COVID-19 deaths and hospitalizations  
519 averted by rapid vaccination rollout in the United States. *bioRxiv*. Published online 2021.  
520 doi:10.1101/2021.07.07.21260156  
521

- 522 [26] Aguilera X, Mundt AP, Araos R, Weitzel T. The story behind Chile's rapid rollout  
523 of COVID-19 vaccination. *Travel Med Infect Dis*. 2021;42(102092):102092.  
524 doi:10.1016/j.tmaid.2021.102092  
525
- 526 [27] Glied S. Strategy drives implementation: COVID vaccination in Israel. *Isr J Health*  
527 *Policy Res*. 2021;10(1):9. doi:10.1186/s13584-021-00445-1  
528
- 529 [28] Vilches TN, Moghadas SM, Sah P, et al. Estimating COVID-19 infections,  
530 hospitalizations, and deaths following the US vaccination campaigns during the  
531 pandemic. *JAMA Netw Open*. 2022;5(1):e2142725.  
532 doi:10.1001/jamanetworkopen.2021.42725  
533
- 534 [29] Monge S, Olmedo C, Alejos B, et al. Direct and indirect effectiveness of mRNA  
535 vaccination against severe acute respiratory syndrome Coronavirus 2 in long-term care  
536 facilities, Spain. *Emerg Infect Dis*. 2021;27(10):2595-2603. doi:10.3201/eid2710.211184  
537
- 538 [30] Clemente-Suárez VJ, Hormeño-Holgado A, Jiménez M, et al. Dynamics of  
539 population immunity due to the herd effect in the COVID-19 pandemic. *Vaccines*  
540 *(Basel)*. 2020;8(2):236. doi:10.3390/vaccines8020236  
541
- 542 [31] Brown CC, Young SG, Pro GC. COVID-19 vaccination rates vary by community  
543 vulnerability: A county-level analysis. *Vaccine*. 2021;39(31):4245-4249.  
544 doi:10.1016/j.vaccine.2021.06.038  
545
- 546 [32] Chen Y, Ma ZF, Yu D, Jiang Z, Wang B, Yuan L. Geographical distribution of trace  
547 elements (selenium, zinc, iron, copper) and case fatality rate of COVID-19: a national  
548 analysis across conterminous USA. *Environ Geochem Health*. 2022;44(12):4423-4436.  
549 doi:10.1007/s10653-022-01204-0  
550
- 551 [33] Weisel O. Vaccination as a social contract: The case of COVID-19 and US political  
552 partisanship. *Proc Natl Acad Sci U S A*. 2021;118(13):e2026745118.  
553 doi:10.1073/pnas.2026745118  
554
- 555 [34] Li T, White LF. Bayesian back-calculation and nowcasting for line list data during  
556 the COVID-19 pandemic. *PLoS Comput Biol*. 2021;17(7):e1009210.  
557 doi:10.1371/journal.pcbi.1009210  
558
- 559 [35] Bergen N, Kirkby K, Fuertes CV, et al. Global state of education-related inequality  
560 in COVID-19 vaccine coverage, structural barriers, vaccine hesitancy, and vaccine  
561 refusal: findings from the Global COVID-19 Trends and Impact Survey. *Lancet Glob*  
562 *Health*. 2023;11(2):e207-e217. doi:10.1016/S2214-109X(22)00520-4  
563
- 564 [36] Chen Y, Zhang L, Li T, Li L. Amplified effect of social vulnerability on health  
565 inequality regarding COVID-19 mortality in the USA: the mediating role of vaccination  
566 allocation. *BMC Public Health*. 2022;22(1):2131. doi:10.1186/s12889-022-14592-w  
567



568 [37] Buscot MJ, Wotherspoon SS, Magnussen CG, et al. Bayesian hierarchical piecewise  
569 regression models: a tool to detect trajectory divergence between groups in long-term  
570 observational studies. *BMC Med Res Methodol*. 2017;17(1). doi:10.1186/s12874-017-  
571 0358-9

572  
573 [38] Li T. *The Bayesian Paradigm of Robustness Indices of Causal Inferences*. Michigan  
574 State University; 2018.

575  
576 [39] Li T, Frank K. The probability of a robust inference for internal validity. *Sociol*  
577 *Methods Res*. 2022;51(4):1947-1968. doi:10.1177/0049124120914922

578  
579 [40] Li T, Frank KA. The probability of a robust inference for internal validity and its  
580 applications in regression models. *arXiv [statME]*. Published online 2020.  
581 <http://arxiv.org/abs/2005.12784>

582  
583 [41] Li T. On the probability of invalidating a causal inference due to limited external  
584 validity. *arXiv [statME]*. Published online 2022. <http://arxiv.org/abs/2206.08649>

585

586

587

588

589

590

591

592

593

594