Association between vaccination rates and severe COVID-19 health outcomes in the United States: a population-level statistical analysis

Hongru Du MS[1,2,*], Samee Saiyed MSE[1,2], and Lauren M. Gardner PhD[1,2,3]

[1]Center for Systems Science and Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[2]Department of Civil and Systems Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[3]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA.

*Corresponding author. Johns Hopkins University, 3400 N. Charles Street, Shaffer 4, Baltimore, MD, 21218, USA. E-mail address: hdu9@jhu.edu (Hongru Du)

Abstract

Vaccine development and distribution have been at the forefront of efforts to combat the COVID-19 pandemic. As the vaccines have been widely adopted by the population, uncertainties around their effectiveness resulting from the emergence of new variants and other confounding factors make it challenging to determine their real-world impact, which is critical for understanding risk, informing public health policies, and mitigating the impact of COVID-19. We analyzed the association between time-dependent vaccination rates and COVID-19 severity for 48 states in the U.S. using Generalized Additive Models (GAMs). We controlled for additional dynamic factors such as testing rates, purpose-specific travel behaviors, underlying population immunity, and policy, and critical static factors such as comorbidities, social vulnerability, race, and state healthcare expenditures. We used SARS-CoV-2 genomic surveillance data to model the different COVID-19 variant driven waves separately, and evaluate if there is a changing role of the potential drivers of severity over time and across waves. Our study revealed a strong and statistically significant negative association between vaccine uptake and COVID-19 severity across each variant wave. Results also showed that booster shots offered additional protection against severe diseases during the Omicron wave. Additionally, higher underlying population immunity based on previous infection rates are shown to be associated with reduced COVID-19 severity. Full-service restaurant visits are associated with increased COVID-19 severity for the pre-Delta and Delta waves, while office of physician visits are associated with increased COVID-19 severity for the Omicron wave. Moreover, the states with higher government policy index scores have lower COVID-19 severity. Regarding static variables, the social vulnerability index, and the proportion of adults at high risk exhibit positive associations with COVID-19 severity, while Medicaid spending per person exhibits a negative association with COVID-19 severity. Despite the emergence of new variants, vaccines remain highly effective at reducing severe outcomes of COVID-19. Therefore, given the ongoing threat posed by COVID-19, vaccines remain a critical line of defense for protecting the public and preventing burden on healthcare systems.

Introduction

By March 1st, 2023, the COVID-19 pandemic had resulted in over 102 million reported cases and 1·1 million deaths in the United States. Vaccine development and distribution have been at the forefront of efforts to combat the impact of the disease. Three vaccines are currently available in the U.S., developed by Pfizer-BioNTech, Moderna, and Johnson & Johnson. The initial randomized clinical trials demonstrated the safety and efficacy of these vaccines, with vaccine efficacies against severe disease (hospitalization and death) ranging from 73% to 96·7%.[1–3] The clinical trials were designed to estimate the direct effect of vaccines against severe disease at the individual level.[4] However, as vaccines roll out to a broader population, uncertainties such as the emergence of new variants, variable immune responses, the quality of cold-chain storage, and other confounding factors may impact a vaccine's efficacy.[5] Hence, evaluating real-world vaccine protection against severe diseases poses a challenge.

Several published studies have attempted to quantify the real-world impact of the COVID-19 vaccines against severe disease. For example, a study in Qatar assessed the vaccines' effectiveness against severe, critical, or fatal Omicron infections using test-negative case-control analysis, and found previous infections and vaccination are effective against symptomatic Omicron infections.[6] An observational study conducted in Israel using national surveillance data showed that the two doses of the Pfizer-BioNTech mRNA vaccines are 97·2% effective in preventing COVID-19-related hospitalizations.[7] A Danish study estimated vaccine effectiveness against COVID-19 hospitalization using a cohort study design, and found that two doses of the vaccine provide high protection against hospitalization for the Alpha and Delta variant, and even higher protection against hospitalization for the Omicron variant.[8] A similar cohort study was applied in Singapore and the United Kingdom to determine whether booster shots reduce the severity of COVID-19 infections during the Omicron wave, and found consistent results - that risk of severe COVID-19 outcomes reduced after receiving booster mRNA vaccines.[9,10]

Currently, most existing literature on the population-level effects of COVID-19 vaccination is based on individual-level data and observational studies. Specifically, these studies relied upon detailed individual level data to assess the direct effectiveness of vaccination by comparing health outcomes between vaccinated and unvaccinated individuals exposed to the same environment. However, these studies may be subject to confounding by unmeasured factors and inconsistent quality of individual level data. Further, in the U.S., such high-resolution data is unavailable at the population-level, so alternative strategies must be engaged to evaluate COVID-19 vaccines effectiveness at a regional level.

One such approach is to rely on compartmental and agent-based models to simulate transmission and disease outcomes both in the presence and absence of vaccines implementation for the same population. Watson et al applied this method to estimate the impact of varying vaccine uptake rates on mortality across multiple countries and found that vaccines prevented 14·4 million COVID-19 deaths in 2021.[11] However, this approach is subject to many assumptions and limitations and therefore limited in its ability to estimate accurate effectiveness. Alternatively, statistical methods such as time series and regression analysis can be implemented to evaluate the association between vaccination coverage and healthcare outcomes across different locations. One study using this

strategy evaluated the association between vaccination coverage and the COVID-19 cases growth rate for all 50 U.S. states in the U.S. using a structural nested mean model and found a 1% increase of vaccination coverage was associated with 1·02% reduction in case growth rate.[12] However, the scope of this study is limited to cases between March and May 2021. A recent study found that vaccine coverage reduced COVID-19 mortality, but seroprevalence and prior infection rates were not associated with mortality[13] This study utilized linear regression to analyze vaccine coverage and natural immunity in relation to mortality during the Delta and Omicron waves. However, this method has limitations in capturing dynamic changes and non-linear relationships between variables. A different study by Bollyky et al[14] applied regression analysis to determine how vaccination coverage amongst other factors (e.g., presence of comorbidities, political partisanship, race, and ethnicity) impacted health outcomes (standardized infections and deaths) in the U.S. at the state level, and determined that higher vaccination rates were associated with lower death rates. The scope of this study varies from ours in its focus on the association between static variables and COVID-19 health outcomes for a fixed time window between January 1st, 2020, and July 31st, 2022, while our study expands the analysis by incorporating novel dynamic variables to capture behavioral changes over time, and explicitly evaluating the different variants independently. A recent study evaluated the time-varying relationship between vaccination, mobility, and COVID-19 health outcomes before and after the Omicron waves.[15] They found the significance of vaccine effectiveness in reducing case rates diminished during the Omicron surge, while its efficacy in lowering case-fatality rates remained substantial throughout the pandemic. Our study extends their analysis by emphasizing hospitalization rate as the outcome variable, breaking aggregated mobility into purpose-specific travel variables, modelling previous infections as a dynamic variable, and including an interaction between the completed primary series and booster rate for the Omicron wave. Despite numerous studies assessing the effectiveness of vaccines, most have not accounted for the relative impact of vaccines across different populations and variant waves, while considering dynamic potential confounding factors. Therefore, a more comprehensive understanding of vaccines' effectiveness across diverse populations and COVID-19 waves is crucial in developing informed public health policies that can effectively mitigate the spread of the virus and ensure equitable distribution of healthcare resources.

In this study, we use Generalized Additive Models (GAMs) to address these shortcomings by evaluating the time-dependent association between vaccines' effectiveness and severe COVID-19 disease for multiple populations (48 states) and across distinct variant driven COVID-19 waves, while controlling for a comprehensive set of potential confounding factors that change over time. Our framework also explicitly captures the spatial variation in the modeled relative associations through a variable transformation procedure (discussed in detail in the methods section). To distinguish between COVID-19 variant driven waves, we utilized SARS-CoV-2 genomic surveillance data and identified the dominant variant for each state and point in time, to determine time windows so the distinct variant-driven waves can be modeled independently. For the Omicron wave, we also considered the added benefit of booster doses on COVID-19. Specifically, we evaluated the interaction between completed primary series and booster rate on reducing the COVID-19 severity. Results from this analysis help improve our understanding of the real-world

relative impact of the available COVID-19 vaccines against severe COVID-19 disease at the population-level over time and can help inform future public health policies to reduce harm.

Methods

Study design

The objective of this study is to analyze the association between vaccination rates and severity of COVID-19 in the U.S., while controlling for potential confounding effects on the main associations of interest. These include novel dynamic covariates that may potentially contribute to COVID-19 severity such as naturally derived immunity from prior COVID-19 infection, local healthcare infrastructure, travel-related activities, government policies, alongside various static variables that have been identified to be significant in previous studies[14,16,17] such as comorbidities, social vulnerability index (SVI), race, and state healthcare expenditures. By controlling for these factors, we aim to provide a more comprehensive understanding of the association between vaccination rate and COVID-19 severity at the population level. The study was conducted for 48 states in the U.S. for the period between April 19th, 2021, the date at which the vaccines were approved for all adults in the U.S., to March 1st, 2022. This period covers the pre-Delta (characterized by the predominance of the Alpha variant and other variants), Delta, and Omicron waves of COVID-19, which are each evaluated independently.

Data sources and collection

We collected state-level time-series data and static variables from publicly available databases. All timeseries data were aggregated to the weekly level. A summary of the variables and their respective sources are listed in table 1, and detailed explanation and justification for the inclusion of each variable are provided in Appendix section 1·3. To mitigate the effects of potential noise and reporting issues a 3-week moving average was applied to all time series variables, with the exception of the government policy index.

Variable Name	**Variable Description**	**Source**
Output variables		
Hospitalization rate	Weekly new admissions of patient with confirmed COVID-19 normalized by reported cases for each state.	[18, 19]
Infection rate	Weekly number of confirmed cases normalized by state population.	[18]
Dynamic input variables		

Partial vaccination rate	Percentage of the total population who received at least one dose of COVID-19 vaccine approved or authorized for use in the United States.	20
Completed primary series rate	Percentage of the population who have received the second dose in a two-dose COVID-19 vaccines primary series or one dose of a single-dose COVID-19 vaccine primary series approved or authorized for use in the United States.	20
Booster vaccination rate	Percentage of the total population who received an updated (bivalent) booster dose.	21
Weekly testing rate	Total number of weekly tests conducted for each state normalized by population.	22
Full-service restaurant visits	Weekly number of visits per person to full-service restaurants.	23
Gym visits	Weekly number of visits per person to gyms.	23
University visits	Weekly number of visits per person to universities.	23
Office of physicians visits	Weekly number of visits per person to Office of physicians.	23
Government policy index	Quantitative measure of government policies implemented in response to the COVID-19 pandemic across various domains including health, social, and economic policies.	24
Previous infection	The total infections reported within a recent time window, e.g., 12 to 16 weeks ahead of hospitalization.	18
Static input variables		
Black proportion	Proportion of the population identifying as Black.	25
Social vulnerability index (SVI)	A single metric based on the combination of 15 social factors, including poverty, lack of vehicle access, and crowded housing.	26
Proportion of adults at high risk	Proportion of adults at higher risk of serious illness if infected with COVID-19 due to age, heart disease, COPD, asthma, diabetes, or BMI.	27
Medicaid spending	Total Medicaid spending for each state normalized	28

	by the population.	

Table 1. Summary of variables.

Dynamic variable transformation

To ensure the precise estimation of each dynamic variable's impact, a variable transmission mechanism must be used to account for the effects of time trends and autocorrelation in the data. For example, the completed primary series rate is always increasing with time for all locations modeled, hence it can be difficult to distinguish how much of the observed associations between vaccination rate and COVID-19 health outcomes are due to the variable interaction or the passage of time. Therefore, we applied the following transformation to all dynamic variables to remove the time trend and redefine the relative variable (RV_i^t):

$$RV_i^t = \frac{V_i^t}{\frac{1}{n}\sum_j V_j^t}$$

Where RV_i^t represents the transformed variable for state i at week t, V_i^t represents the original variable for state i at week t without the transformation, $\frac{1}{n}\sum_j V_j^t$ represents the mean of the original variable at week t, over all locations being modeled n, e.g., the national mean across the U.S. A RV_i^t larger than one indicates that state i has a higher variable value compared to the national mean at week t, while RV_i^t lower than one indicates that state i has a lower variable value compared to the national mean, at week t. After normalization, the final set of time-dependent variables included in the analysis are: Relative hospitalization rates (RHR_i^t), relative completed primary series rate ($RCPSR_i^t$), relative booster rate (RBR_i^t), relative weekly testing rate ($RWTR_i^t$), relative full-service restaurant visits ($RFSRV_i^t$), relative gym visits (RGV_i^t), relative office of physician visits ($ROPV_i^t$), relative university visits (RUV_i^t), relative previous infection (RPI_i^t), and relative government policy (RGP_i^t). These newly transformed variables enable an explicit evaluation of the relative association between each of them and the COVID-19 severity outcome of interest within a single multi-state model.

Statistical analysis

The generalized additive model (GAM) was selected as the statistical model for this analysis because of its ability to capture complex and non-linear relationships between the independent variables and the outcome variables of interest, namely, the weekly hospitalization rate in each state and weekly infection rate in each state. In efforts to identify potential discrepancies in vaccination on COVID-19 severity for different variants, we model each variant-driven wave during the study period independently. To define the variant waves we clustered each state-week pair based on the dominant circulating variant based on SARS-CoV-2 genomic surveillance data downloaded from GISAID.[29] The three waves are classified as: 1) Pre-Delta Wave, 2) Delta Wave, 3) Omicron Wave, and each state is labeled with its most dominant variant each week to define the windows. Details of this classification are described in Appendix section 1·1, and the

assignment of state-week pairs is shown in Appendix figure S1. In addition to modeling each wave separately, we also implement a single model covering the entire the study period.

Eight different GAMs were implemented in this study, which differ in their set of input covariates, response variable, and model structure. The first four models treat hospitalization rate as the outcome variable, while the second four model infection rate.

Model 1, Dyn-Hosp, treats hospitalization rates as the explanatory variable and includes only dynamic covariates, namely relative completed primary series rate, relative weekly testing rate, relative full-service restaurant visits, relative gym visits, relative office of physician visits, relative university visits, relative government policy, and relative previous infections. Model 1 is presented in equation (1):

$$RHR_i^t \sim Gamma(\mu, \phi)$$
$$\log(\mu) = f_1(RCPSR_i^{t-2}) + f_2(RWTR_i^{t-2}) + f_3(RFSRV_i^{t-2}) + f_4(RGV_i^{t-2}) + f_5(ROPV_i^{t-2}) + f_6(RUV_i^{t-2}) + f_8(RGP_i^{t-2}) + f_7(RPI_i^t) \quad (1)$$

In Model 1, the Gamma distribution is utilized to model the hospitalization rate variable with mean μ and variance ϕ. f_i denotes the smoothed, non-parametric functions of each independent variable. To account for the sequential process of infection leading to hospitalization we introduce a time lag between each of the input variables relative to the outcome variable, which is denoted by the superscript. The timeline of this model is introduced as follows: the modeled hospitalizations rate, (RHR_i^t), occur at time *t*. Infections resulting in hospitalization, are assumed to occur at time *t-2*, to account for a one week incubation period[30], and one additional week between symptoms onset and hospitalization.[31] Note, this timeline aligns with the definition of the hospitalization rate variable, which is normalized by the number of reported infections one week prior, which assumes a one week delay between when infection occurred and when it is reported. To accurately reflect the conditions presented at the time of infection, each of the variables related to vaccination ($RCPSR_i^{t-2}$), mobility ($RFSRV_i^{t-2}$, RGV_i^{t-2}, $ROPV_i^{t-2}$, RUV_i^{t-2}) , policy (RGP_i^{t-2}), and testing ($RWTR_i^{t-2}$) are also lagged by two weeks relative to the hospitalization variable. Lastly, the past infections variable, defined as stated above to capture the role of recently acquired immunity from infection in protecting from severe disease upon reinfection, is equal to the total infection rate in the population summed over the prior four to 16 weeks. This time window is explicit in the definition of RPI_i^t (see Appendix section 1·3). The lagged values for dynamic variables in Models 2-4 incorporate identical explanations as those utilized in Model 1.

Model 2, Dyn-Hosp-Booster, is a direct extension of Model 1, applied for only the Omicron wave, and includes an additional interaction effect between the relative completed primary series rate ($RCPSR_i^t$) and the relative booster rate (RBR_i^t). Model 2 is presented in equation (2):

$$RHR_i^t \sim Gamma(\mu, \phi)$$
$$\log(\mu) = f_1(RCPSR_i^{t-2}, RBR_i^t) + f_2(RWTR_i^{t-2}) + f_3(RFSRV_i^{t-2}) + f_4(RGV_i^{t-2}) +$$

$$f_5(ROPV_i^{t-2}) + f_6(RUV_i^{t-2}) + f_8(RGP_i^{t-2}) + f_7(RPI_i^t) \quad (2)$$

Model 3, Int-Hosp, also models hospitalization rates as the explanatory variable and includes both static and dynamic covariates, specifically removing the dynamic variables relative full-service restaurant visits, relative gym visits, relative office of physician visits, and relative university visits, and including the static variables, black proportion, the social vulnerability index (SVI), the proportion of adults at high risk, and Medicaid spending. Model 3 is presented in equation (3):

$$RHR_i^t \sim Gamma(\mu, \phi)$$

$$\begin{aligned} log(\mu) = {} & f_1(RCPSR_i^{t-2}) + f_2(RWTR_i^{t-2}) + f_3(RGP_i^{t-2}) + f_4(RPI_i^t) \\ & + f_5(Black\ proportion) + f_6(social\ vulnerability\ index) \\ & + f_7(adults\ at\ high\ risk) + f_8(Medicaid\ spending) \end{aligned} \quad (3)$$

The selection of static variables included in the model was based on those found significant in prior research, and the removal of the subset of dynamic variables was based on a correlation analysis amongst the static and dynamic variables. Detail of static variables selections are documented in Appendix section 2·6.

Model 4, Int-Hosp-Booster, is an extension of Model 3 that includes an additional interaction effect between the relative completed primary series rate ($RCPSR_i^t$) and the relative booster rate (RBR_i^t). Model 4 is presented in equation (4):

$$RHR_i^t \sim Gamma(\mu, \phi)$$

$$\begin{aligned} log(\mu) = {} & f_1(RCPSR_i^{t-2}, RBR_i^{t-2}) + f_2(RWTR_i^{t-2}) + f_3(RGP_i^{t-2}) + f_4(RPI_i^t) \\ & + f_5(Black\ proportion) + f_6(social\ vulnerability\ index) \\ & + f_7(adults\ at\ high\ risk) + f_8(Medicaid\ spending) \end{aligned} \quad (4)$$

Each of these four models are also implemented with the infection rate as the outcome variable of interest, making eight total models. The hospitalization and infection rates were assumed to follow a Gamma distribution, and the parameters were estimated using the Residual Maximum Likelihood (REML) method. The choice of Gamma distribution is based on the observed right-skewed nature of the relative hospitalization rate and relative infection rate, which has been validated through distribution fitting (Appendix section 1·2). Further details of each model are documented in Appendix section 2·7. All models are fitted separately to the full study period and each variant wave.

The results for Models 1 and 2, Dyn-Hosp and Dyn-Hosp-Booster, are presented in the section below. The results for Models 3 and 4, Int-Hosp, Int-Hosp-Booster, are presented in Appendix sections 3·2 and 3·3. The equivalent results for Models 4-8, with infection rate as the outcome

variable, are presented in Appendix section 3·4 to 3·7. Data processing, visualization, and analysis were carried out using R 4·0 and Python 3·8.

Results

Figure 1 illustrates the results for Dyn-Hosp for each variant, namely, the Pre-Delta (Blue), Delta (Orange), and Omicron waves (Red), respectively.

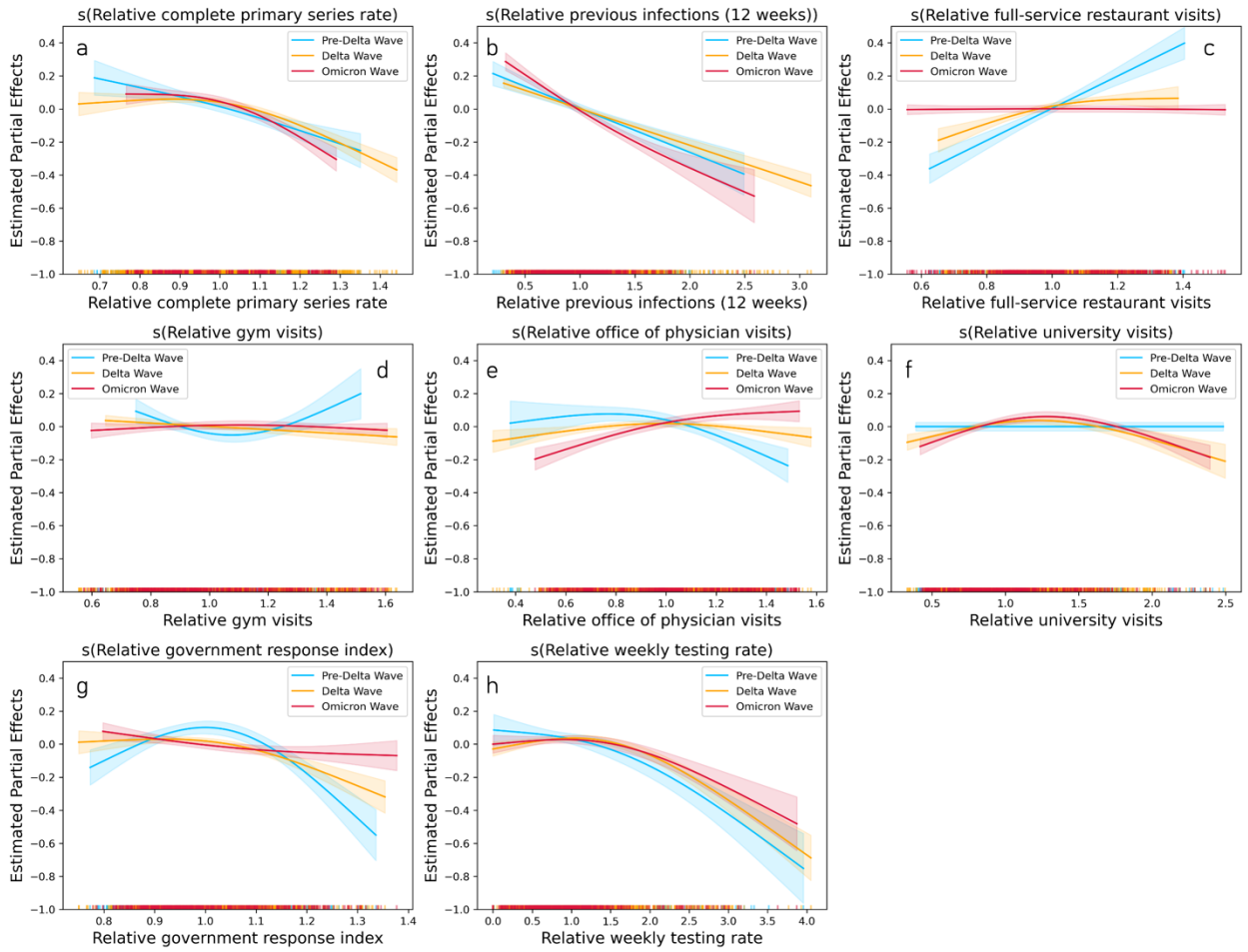


Figure 1: Estimated partial effects of dependent variables by different variant waves from Dyn-Hosp. The dots on the x-axis represent the distribution of the variables.

Dyn-Hosp results for the three variants revealed similar patterns as the full model (Appendix figure S22). The relative completed primary series rate, relative previous infections, and relative weekly testing rate consistently displayed strong negative associations with relative hospitalization rate across different waves. Moreover, the relative full-service restaurant visits maintained a positive effect except for the Omicron wave. The relative office visits to physicians slightly changed from negative to positive effects as the analysis progressed from the pre-Delta to the Omicron wave. On the other hand, the results from the relative government response index revealed a different pattern which resulted in its impact gradually flattening out from the pre-Delta to the Omicron wave.

Figure 2 illustrates the results of Dyn-Hosp-Booster, for just the Omicron wave with the additional inclusion of an interaction effect between the completed primary series rate and the relative booster

rate. This interaction effect is presented as a 3-dimensional surface and a 2-dimension contour map in figure 2·a, and 2·b.

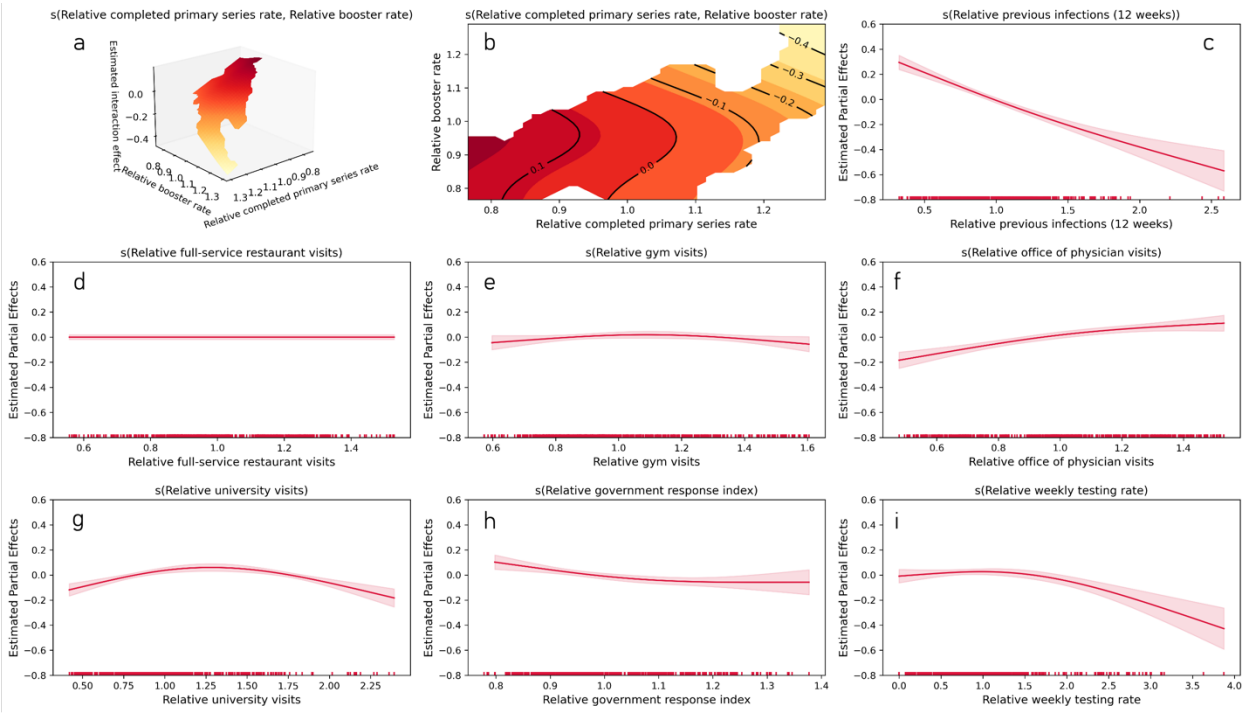


Figure 2: Estimated partial effects of dependent variables for the Omicron wave from Dyn-Hosp-Booster. The figure 2·a shows a three-dimensional partial dependent plot for the interaction and figure 2·b shows the two-dimensional contour plot. The dots on the x-axis represent the distribution of the variables.

With the exception of the completed primary series rate, the effects of all other variables shown modeled remained consistent with the results for the Omicron wave shown in figure 1. Figures 2·a, and 2·b show the interaction between two vaccine-related variables in three-dimensional space and the projection into two-dimensional variable space, respectively. The deeper red indicates a more positive effect on the hospitalization rate, and the light yellow indicates a more negative effect to the hospitalization rate. The solid black lines in figure 2·b represent the contour lines. The contour lines correspond to points that have an equivalent impact on the hospitalization rate, with the values marked on each line indicating the actual interaction effect of these points on the hospitalization rate. Those two plots reveal that the relative hospitalization rate decreases along the direction of increasing the relative booster rate and the relative completed primary series rate.

Discussion

The primary objective of administering a COVID-19 vaccine is to safeguard against severe illness from the virus. However, despite the wide availability of vaccines in the U.S. since April 19th, 2021, their real-world impact on minimizing COVID-19 severity at a population level is yet to be fully understood. In addition, it is difficult to ascertain how the effectiveness of the vaccines may vary for different variants. In this study, we aimed to address this gap in knowledge by evaluating the association between COVID-19 severity and vaccine uptake across different variant waves in the United States, while explicitly taking into account additional potentially dynamic confounding

factors such as purpose-specific travel behaviors, government policy, previous infection rates, and testing rates. The revealed associations can help improve our understanding of vaccines effectiveness in real world settings and compliment the findings from clinical trials and survey-based studies.

Vaccines protect against severe COVID-19 for pre-Delta, Delta and Omicron waves

Our study reveals a strong and statistically significant negative association between vaccine uptake rates and COVID-19 severity. This relationship was consistent across each of the variant waves modeled, as well as the entire 11-month study period. As illustrated in figure 1·a, the results indicated no decline in protection from the completed primary series through the Omicron waves, consistent with earlier findings that vaccine protection against severe illnesses does not significantly wane with the new variants circulated to date. In contrast, when we modeled infection rates as the response variable, we observed a decreasing effect of vaccines from the pre-Delta to the Omicron wave (Appendix section 3·4). This outcome aligns with existing literature highlighting the rapid waning of the vaccines' effectiveness against infection.[32,33] Nonetheless, while vaccines may offer reduced protection against infection, they continue to provide substantial protection against severe outcomes of COVID-19. A separate GAM analysis (Figure 2) including booster data for the Omicron variant indicates that boosters provide additional protection against severe disease caused by the Omicron variant, particularly in states with higher completed primary series rates. Although the value of booster shots in protecting against severe cases of COVID-19 is still being studied and controversial,[34] results from our analysis provide evidence supporting the effectiveness of booster doses against severe cases caused by the Omicron variant. Nonetheless, the findings obtained from model 6 (Dyn-Inf-Booster) reveal that the interaction between the booster and completed primary series rates has a relatively limited impact on Omicron infection (Appendix figure S28). However, it is crucial to emphasize that despite the diminished effectiveness of mRNA boosters against Omicron infections, vaccines still serve the essential purpose in reducing the severity of COVID-19, even in the face of emerging variants.

Immunity from recent infection protects against severe COVID-19 upon reinfection

Higher past COVID-19 infection levels in a population are associated with a decrease in COVID-19 severity, indicating immunity gained from infection can provide some protection against severe disease in the event of re-infection in the future, but only for a limited period of time. Our study utilized the total number of cases reported in a 12-week window, ranging from 4 to 16 weeks prior to the time period modeled, as a proxy for recently acquired immunity, and found a strong negative association between the previous infection rate and future COVID-19 severity. These results were consistent across the different variant waves. This finding aligns with other case-control studies that found previous infections showed strong effectiveness against severe, critical, or fatal COVID-19.[6,35] Additionally, as shown in the sensitivity analysis in Appendix section 2·4, evidence for protection from prior infection exists across time windows up to 24 weeks, indicating protection from previous infection can last for up to six months. Results from the sensitivity analysis also highlight that the negative association between previous infection rates and severity of COVID-19 weakens over time, (see Appendix section 2·5), suggesting that infection-acquired immunity protection against severe COVID-19 outcomes decreases with time since the last

infection. While the waning of natural immunity has been established in molecular and clinical research[36], our analysis provides additional insight at the population level. Our models with infection rate as the outcome variables reveal an insignificant association between previous infection and infection rate during the pre-Delta and Delta waves but a significant negative association during the Omicron wave (Appendix figure S27). This finding partially aligns with existing literature that found suggests, at the individual level, previous infection protected against infection pre-Omicron, but this effectiveness decreases substantially during the Omicron wave.[37] Nevertheless, at the population level, the number of infected individuals is considerably higher during the Omicron wave than earlier, while a smaller proportion remains susceptible. Consequently, the cumulative impact of previous infections becomes more pronounced. Our results highlight the variability in the impact of previous infections on infection at the individual and population levels.

Local factors contribute to variation in COVID-19 health outcomes

Existing clinical and statistical studies[14,16,17] have identified critical indicators for COVID-19 severity including demographics, comorbidities, SVI, and healthcare expenditures. Results from the integrated model that includes both dynamic and static variables indicate the proportion of adults at high risk and the social vulnerability index (SVI) are positively associated with COVID-19 severity across all variant waves modeled (Appendix figure S24·f and S24·g). This finding is consistent with existing literature[14,17], which suggests that individuals with comorbidities and those from socially vulnerable regions are more likely to experience severe COVID-19 outcomes. Our results reveal a negligible association between black proportion and COVID-19 severity, which differs from a previously identified positive associations.[14] This discrepancy is likely due to the more comprehensive nature of our model, which includes novel covariates that have strong relationships with ethnicity rates, e.g., full-service restaurant and office of physician visits are positively correlated with black proportion, thus reducing the impact of the standalone ethnicity variable. In the United States, the eligibility for Medicaid varies by state, but generally, individuals and families with incomes up to 138% of the federal poverty level may qualify for coverage.[38] Our results reveal a consistent negative association between state-level Medicaid spending per person and COVID-19 severity (Appendix figure S24·h), which indicates the potential protective effect of healthcare expenditures in mitigating the impact of the pandemic on vulnerable groups. In contrast to the hospitalization rate models, when infection rate is treated as the outcome variable (see Appendix section 3·7), there is no evidence for consistent or significant associations with demographics, comorbidities, SVI, or healthcare expenditures across variant waves. These results suggests that dynamic COVID-19 infection risk is complex and changed over time, and the factors which contributed to transmission varied across waves. Further research is needed for a more comprehensive understanding of the complex and evolving nature of COVID-19 transmission.

Mobility-derived activity patterns are associated with COVID-19 health outcomes

At the beginning of the pandemic, several studies evaluated the association between mobility and COVID-19 transmission with inconsistent findings.[39,40] One possible reason for this inconsistency is that aggregated mobility data may not accurately reflect dynamically changing human behaviors. Furthermore, the connection between mobility and severe health outcomes remains unclear. This

study uses disaggregated mobility patterns to capture diverse behaviors between populations, specifically relative engagement levels in particular activities, to explore the association between these variables and COVID-19 severity. Results shown in figure 1·c indicate that increased visits to full-service restaurants are positively associated with severe COVID-19 outcomes in the pre-Delta and Delta waves, but not significant during the Omicron wave. Furthermore, regarding the lack of evidence for this relationship during Omicron, the sensitivity analysis conducted that relaxed the basis dimension k in the model (allowing for a more complex, nonlinear relationship between the input and output variables in the model), revealed that full-service restaurant visits also exhibit a strong nonlinear positive relationship with COVID-19 severity for the Omicron wave (refer to Appendix figure S9). These results indicate a more complex, but still significant relationship between dining out and COVID-19 severity during Omicron relative to earlier waves. In contrast, when we modeled the infection rate as the outcome variable, our analysis revealed a slight negative to neutral effect of full-service restaurant visits (refer to Appendix figure S27·c). Taken together, our findings suggest that while more visits to full-service restaurants may not contribute to increased transmission, they may be linked to higher-risk infections because they are occurring among more vulnerable demographic groups. This hypothesis is supported by additional analysis (shown in Appendix figures S19 to S21), which highlights the positive correlation between regional full-service restaurant visits and both the social vulnerability index and poverty rate. In addition to restaurant activity, we also observed a significant association between disease severity and visits to hospitals, medical centers, and Outpatient Care Centers, however unlike with restaurant activity, this association transitioned from negative to positive between the pre-Delta to Omicron wave. One possible explanation for this finding is that as the pandemic evolved, the public became more familiar with the disease and more tolerant to at-home symptom management, thus those COVID-19 patients that sought medical care were more likely to be those with more severe symptoms. Finally, visits to the university and gym were found not to be significantly associated with COVID-19 severity. This is likely due to the young and relatively healthy demographic that frequents universities and gyms, while still vulnerable to contracting the SARS-CoV-2 virus, they are less likely to experience severe outcomes from COVID-19 infection. This hypothesis is further supported by the findings from the infection rate model, which identified a positive association with university visits during the Delta and Omicron waves (Appendix figure S27·f).

More stringent government public health policy is associated with reduced COVID-19 severity

Our results indicate that more stringent government policies were associated with reduced COVID-19 severity, consistent with previous studies.[41] More precisely, it was found that the state-week pairs with a significantly high government response index (indicating stricter policy) have a stronger negative effect on the severity of COVID-19. (Appendix figure S22). However, this negative effect reduced over time, and was least evident during the Omicron wave. The reduced effect of the policy during Omicron is likely due to a complex combination of factors, including the increasing population level immunity from both widespread adoptions of vaccines and prior exposure providing more protection from severe disease during this period, combined with a reduction in the government's response to the pandemic over time.

Additionally, weekly testing rates were shown to be negatively associated with COVID-19 severity levels. While this result does not imply a causative relationship between testing rate and COVID-19 severity, there are various reasons why testing rates may be linked to severity rates. Firstly, higher testing rates during periods of high transmission can provide more public awareness of risk, in which case the more vulnerable population make be more likely to take precautions to protect themselves, thus lowering transmission amongst high-risk groups and lowering severity rates. Second, testing rates directly affect the hospitalization rate because they impact the denominator, i.e., confirmed cases. Third, the testing rate can capture various confounding effects. For example, as shown in Appendix figure S19 to S21, the weekly testing rate positively correlates with medical aid spending at the state level, and may indirectly capture the effect of state-level healthcare expenditures on population level COVID-19 health outcomes.

Limitations

As with all modeling studies, this work is subject to several limitations. Firstly, this study was primarily designed to determine the association between various potential risk factors and COVID-19 outcomes, rather than to establish causality between these variables. Thus, it is possible our findings may be reflecting the role of unobserved confounding factors that are excluded from our study. Another potential limitation is due to the application at the state-level. The aggregation of the data to the state-level is unable to capture the heterogeneities of the communities within each state, and it is possible different associations exists at the local level, than are identified at the state-level. Additionally, while we believe the use of the hospitalization rate in a given state at a given time is a plausible choice as a proxy for COVID-19 severity, it is subject to variable case reporting and data quality across states, which may arise due to uneven testing capacity, reporting delays or at-home testing, among others.

Conclusion

This research utilizes publicly available real-world data to provide robust evidence of the efficacy of vaccines against COVID-19 severity across various variant waves in the United States. More importantly, this paper concludes that booster shots offer additional protection against severe COVID-19 during the Omicron waves. Despite the emergence of new variants, vaccines remain the most effective intervention for mitigating the severe effects of COVID-19. Therefore, given the ongoing threat posed by COVID-19 and its potential variants, vaccines continue to be the best line of defense for protecting public health and preventing the further spread of the virus.

Contributors

LG, and HD contributed to the conceptualization and design of the study. HD and SS collected the data and conducted the analysis. HD led the writing of the original draft. HD, SS, and LG edited the manuscript, discussed results, and provided feedback regarding the manuscript. LG supervised the study and acquired funding. HD and SS have verified the underlying data. All authors had full access to the data and approved the manuscript for publication.

Data sharing statement

All the data and code used for the analysis is available from https://github.com/hongru94/vaccination_rate_GAMs.

Acknowledgments

This work was funded the NSF RAPID Award ID 2108526, NSF Award ID 2229996, and CDC Contract #75D30120C09570.

Declaration of interests

The authors declare no potential conflicts of interest.

Reference

1 Baden LR, El Sahly HM, Essink B, *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* 2021; **384**: 403–16.

2 Sadoff J, Gray G, Vandebosch A, *et al.* Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *N Engl J Med* 2021; **384**: 2187–201.

3 Thomas SJ, Moreira ED, Kitchin N, *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months. *N Engl J Med* 2021; **385**: 1761–73.

4 Kim JH, Marks F, Clemens JD. Looking beyond COVID-19 vaccine phase 3 trials. *Nat Med* 2021; **27**: 205–11.

5 Patel MK, Bergeri I, Bresee JS, *et al.* Evaluation of post-introduction COVID-19 vaccine effectiveness: Summary of interim guidance of the World Health Organization. *Vaccine* 2021; **39**: 4013–24.

6 Altarawneh HN, Chemaitelly H, Ayoub HH, *et al.* Effects of Previous Infection and Vaccination on Symptomatic Omicron Infections. *N Engl J Med* 2022; **387**: 21–34.

7 Haas EJ, Angulo FJ, McLaughlin JM, *et al.* Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *The Lancet* 2021; **397**: 1819–29.

8 Gram MA, Emborg H-D, Schelde AB, *et al.* Vaccine effectiveness against SARS-CoV-2 infection or COVID-19 hospitalization with the Alpha, Delta, or Omicron SARS-CoV-2 variant: A nationwide Danish cohort study. *PLOS Med* 2022; **19**: e1003992.

9 Ng OT, Marimuthu K, Lim N, *et al.* Analysis of COVID-19 Incidence and Severity Among Adults Vaccinated With 2-Dose mRNA COVID-19 or Inactivated SARS-CoV-2 Vaccines With and Without Boosters in Singapore. *JAMA Netw Open* 2022; **5**: e2228900.

10 Agrawal U, Bedston S, McCowan C, *et al.* Severe COVID-19 outcomes after full vaccination of primary schedule and initial boosters: pooled analysis of national prospective cohort studies of 30 million individuals in England, Northern Ireland, Scotland, and Wales. *The Lancet* 2022; **400**: 1305–20.

11 Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *Lancet Infect Dis* 2022; **22**: 1293–302.

12 Wang R, Wang J, Hu T, Zhou X-H. Population-Level Effectiveness of COVID-19 Vaccination Program in the United States: Causal Analysis Based on Structural Nested Mean Model. *Vaccines* 2022; **10**: 726.

13 Shioda K. Population-Level Relative Effectiveness of the COVID-19 Vaccines and the Contribution of Naturally Acquired Immunity. .

14 Bollyky TJ, Castro E, Aravkin AY, *et al.* Assessing COVID-19 pandemic policies and behaviours and their economic and educational trade-offs across US states from Jan 1, 2020, to July 31, 2022: an observational analysis. *The Lancet* 2023; : S0140673623004610.

15 Hu S, Xiong C, Zhao Y, Yuan X, Wang X. Vaccination, human mobility, and COVID-19 health outcomes: Empirical comparison before and during the outbreak of SARS-Cov-2 B.1.1.529 (Omicron) variant. *Vaccine* 2023; : S0264410X23006175.

16 Ogedegbe G, Ravenell J, Adhikari S, *et al.* Assessment of racial/ethnic disparities in hospitalization and mortality in patients with COVID-19 in New York City. *JAMA Netw Open* 2020; **3**: e2026881–e2026881.

17 Gallo Marin B, Aghagoli G, Lavine K, *et al.* Predictors of COVID-19 severity: a literature review. *Rev Med Virol* 2021; **31**: 1–10.

18 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; **20**: 533–4.

19 HHS Protect Public Data Hub. https://public-data-hub-dhhs.hub.arcgis.com/ (accessed Jan 13, 2023).

20 COVID-19/data_tables/vaccine_data/us_data at master · govex/COVID-19. GitHub. https://github.com/govex/COVID-19 (accessed Jan 13, 2023).

21 CDC. COVID Data Tracker. Cent. Dis. Control Prev. 2020; published online March 28. https://covid.cdc.gov/covid-data-tracker (accessed Jan 13, 2023).

22 COVID-19/data_tables/testing_data at master · govex/COVID-19. GitHub. https://github.com/govex/COVID-19 (accessed Jan 13, 2023).

23 Places Data Curated for Accurate Geospatial Analytics | SafeGraph. https://www.safegraph.com (accessed Jan 13, 2023).

24 Hale T, Webster S, Petherick A, Phillips T, Kira B. Oxford COVID-19 government response tracker (OxCGRT). *Last Updat* 2020; **8**: 30.

25 Bureau UC. State Population by Characteristics: 2010-2019. Census.gov. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-detail.html (accessed April 30, 2023).

26 CDC/ATSDR SVI Data and Documentation Download | Place and Health | ATSDR. 2022; published online Dec 22. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html (accessed Feb 13, 2023).

27 Adults at Higher Risk of Serious Illness if Infected with Coronavirus. KFF. 2020; published online May 13. https://www.kff.org/other/state-indicator/adults-at-higher-risk-of-serious-illness-if-infected-with-coronavirus/ (accessed Feb 13, 2023).

28 Health Insurance Coverage of the Total Population. KFF. 2022; published online Oct 28. https://www.kff.org/other/state-indicator/total-population/ (accessed April 30, 2023).

29 Khare S, Gurry C, Freitas L, *et al.* GISAID's role in pandemic response. *China CDC Wkly* 2021; **3**: 1049.

30 Quesada JA, López-Pineda A, Gil-Guillén VF, Arriero-Marín JM, Gutiérrez F, Carratala-Munuera C. Incubation period of COVID-19: A systematic review and meta-analysis. *Rev Clínica Esp Engl Ed* 2021; **221**: 109–17.

31 Faes C, Abrams S, Van Beckhoven D, *et al.* Time between symptom onset, hospitalisation and recovery or death: statistical analysis of Belgian COVID-19 patients. *Int J Environ Res Public Health* 2020; **17**: 7560.

32 Goldberg Y, Mandel M, Bar-On YM, *et al.* Waning Immunity after the BNT162b2 Vaccine in Israel. *N Engl J Med* 2021; **385**: e85.

33 Tenforde MW, Self WH, Naioti EA, *et al.* Sustained effectiveness of Pfizer-BioNTech and Moderna vaccines against COVID-19 associated hospitalizations among adults—United States, March–July 2021. *Morb Mortal Wkly Rep* 2021; **70**: 1156.

34 Chenchula S, Karunakaran P, Sharma S, Chavan M. Current evidence on efficacy of COVID-19 booster dose vaccination against the Omicron variant: A systematic review. *J Med Virol* 2022; **94**: 2969–76.

35 Carazo S, Skowronski DM, Brisson M, *et al.* Protection against omicron (B. 1.1. 529) BA. 2 reinfection conferred by primary omicron BA. 1 or pre-omicron SARS-CoV-2 infection among health-care workers with and without mRNA vaccination: a test-negative case-control study. *Lancet Infect Dis* 2023; **23**: 45–55.

36 Bobrovitz N, Ware H, Ma X, *et al.* Protective effectiveness of previous SARS-CoV-2 infection and hybrid immunity against the omicron variant and severe disease: a systematic review and meta-regression. *Lancet Infect Dis* 2023; **23**: 556–67.

37 Altarawneh HN, Chemaitelly H, Hasan MR, *et al.* Protection against the Omicron Variant from Previous SARS-CoV-2 Infection. *N Engl J Med* 2022; **386**: 1288–90.

38 Blavin F, Karpman M, Kenney GM, Sommers BD. Medicaid versus marketplace coverage for near-poor adults: effects on out-of-pocket spending and coverage. *Health Aff (Millwood)* 2018; **37**: 299–307.

39 Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect Dis* 2020; **20**: 1247–54.

40 Kraemer MU, Yang C-H, Gutierrez B, *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020; **368**: 493–7.

41 Hale T, Angrist N, Hale AJ, *et al.* Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. *PLoS One* 2021; **16**: e0253116.

Supplementary Materials for

Association between vaccination rates and severe COVID-19 health outcomes in the United States: a population-level statistical analysis

Hongru Du MS[1,2,*], Samee Saiyed MSE[1,2], and Lauren M. Gardner PhD[1,2,3]

[1]Center for Systems Science and Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[2]Department of Civil and Systems Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[3]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA.

*Corresponding author. Johns Hopkins University, 3400 N. Charles Street, Shaffer 4, Baltimore, MD, 21218, USA. E-mail address: hdu9@jhu.edu (Hongru Du)

Table of Contents

1. Supplementary Data ***20***

1.1 Preprocessing of genomic data **20**

1.2 Distribution of outcome variables **20**

1.3 Variables description **23**

1.4 Preprocessing of mobility data **25**

2. Supplementary Methods ***25***

2.1 Dynamic mobility variables selection **25**

2.2 Robustness check of vaccination data **27**

2.3 Sensitivity analysis of smooth basis dimension (k) **29**

2.4 Sensitivity analysis of prior window length for previous infection **33**

2.5 Sensitivity analysis of lags for previous infection **35**

2.6 Static variables selection **37**

2.7 GAMs with infection rate as the outcome variable **39**

3. Supplementary Results ***41***

3.1 Results for Model 1 (Dyn-Hosp) **41**

3.2 Results for Model 3 (Int-Hosp) **42**

3.3 Results for Model 4 (Int-Hosp-Booster) **44**

3.4 Results for Model 5 (Dyn-Inf) **45**

3.5 Results for Model 6 (Dyn-Inf-Booster) **47**

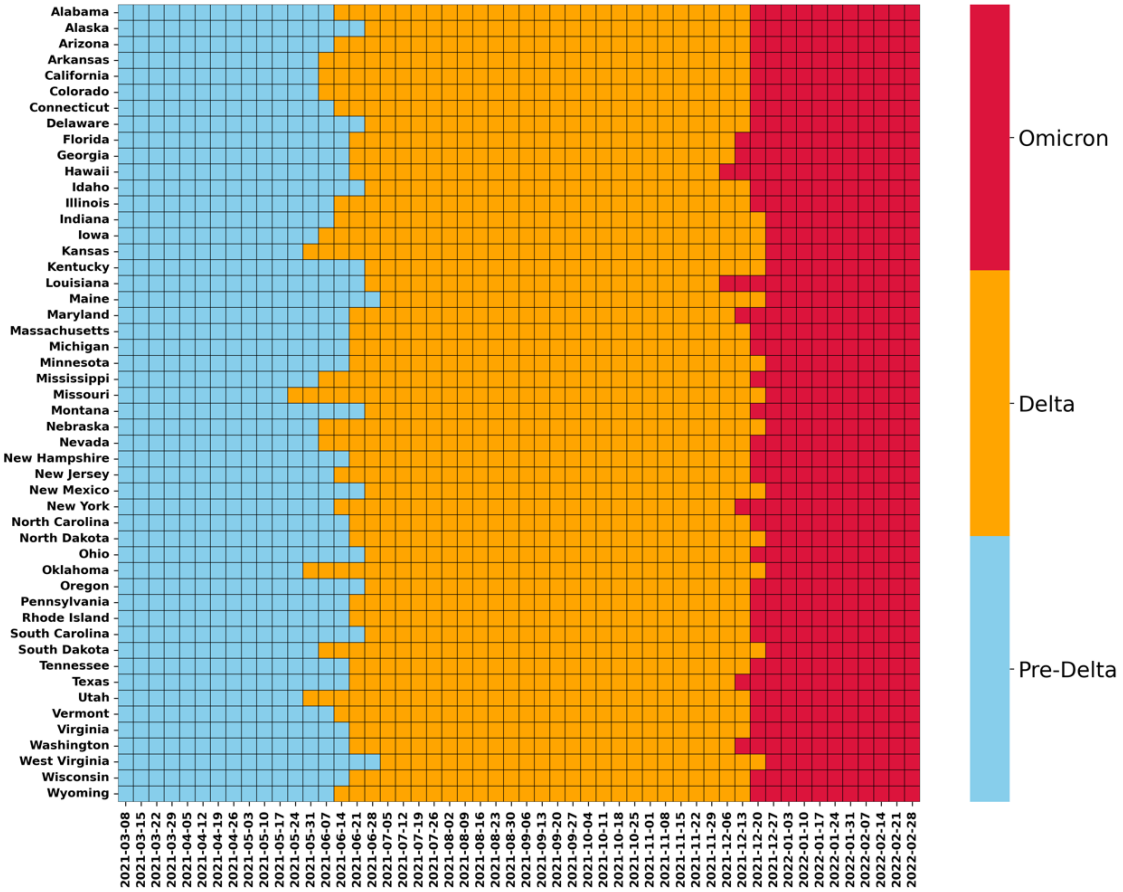
3.6 Results for Model 7 (Int-Inf) **48**

3.7 Results for Model 8 (Int-Inf-Booster) 50
Reference ***51***

1. Supplementary Data

1.1 Preprocessing of genomic data

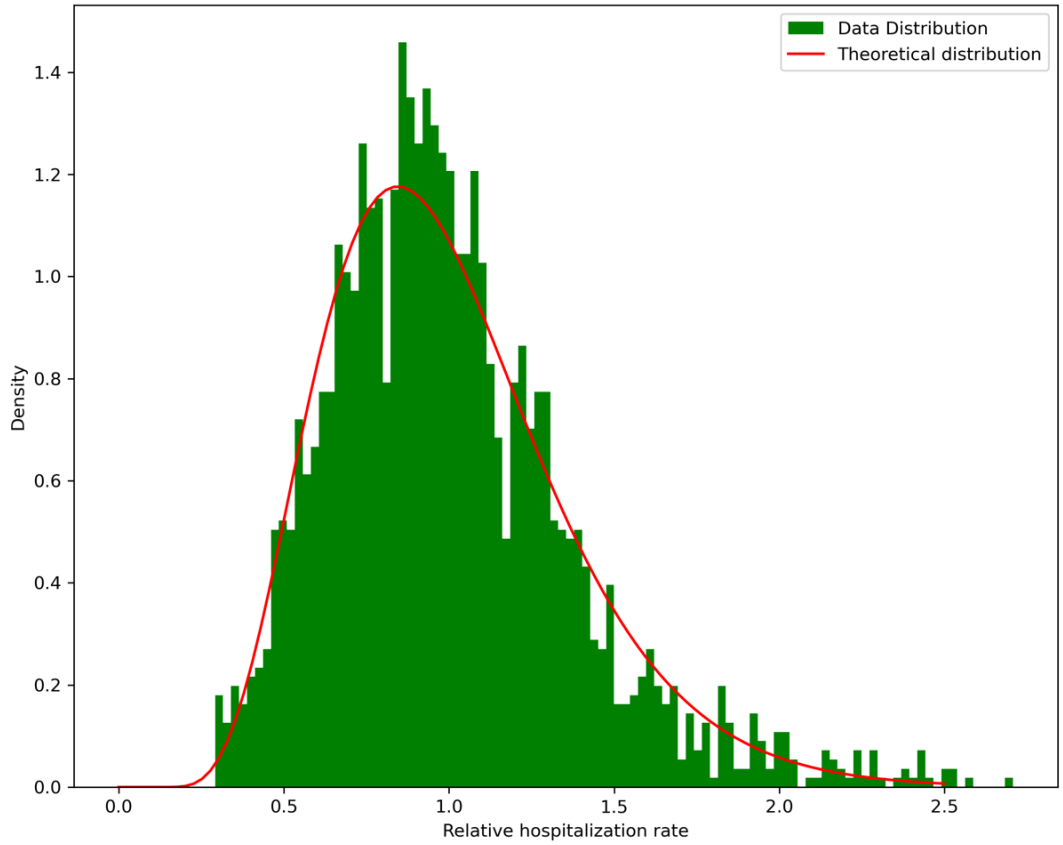
All genomic data were collected from GISAID[1] on October 27th, 2022. GISAID is a publicly accessible repository of dataset that sharing of genomic data on various pathogens, including influenza and COVID-19. We analyzed the available set of sequences, to determine the proportion of each variant theoretically in circulation. Specifically, we calculated the proportion of each variant for each week in each state from March 1st, 2021, to March 1st, 2022. To identify the most dominant variant for each state-week pair during the analyzed period, we labeled the state-week pairs based on the variant with the highest proportion. This enables us to track the dominant variant in each state and cluster the state-week pairs based on the most dominant variant. The assignment of state-week pairs is shown in Appendix figure S1 below:



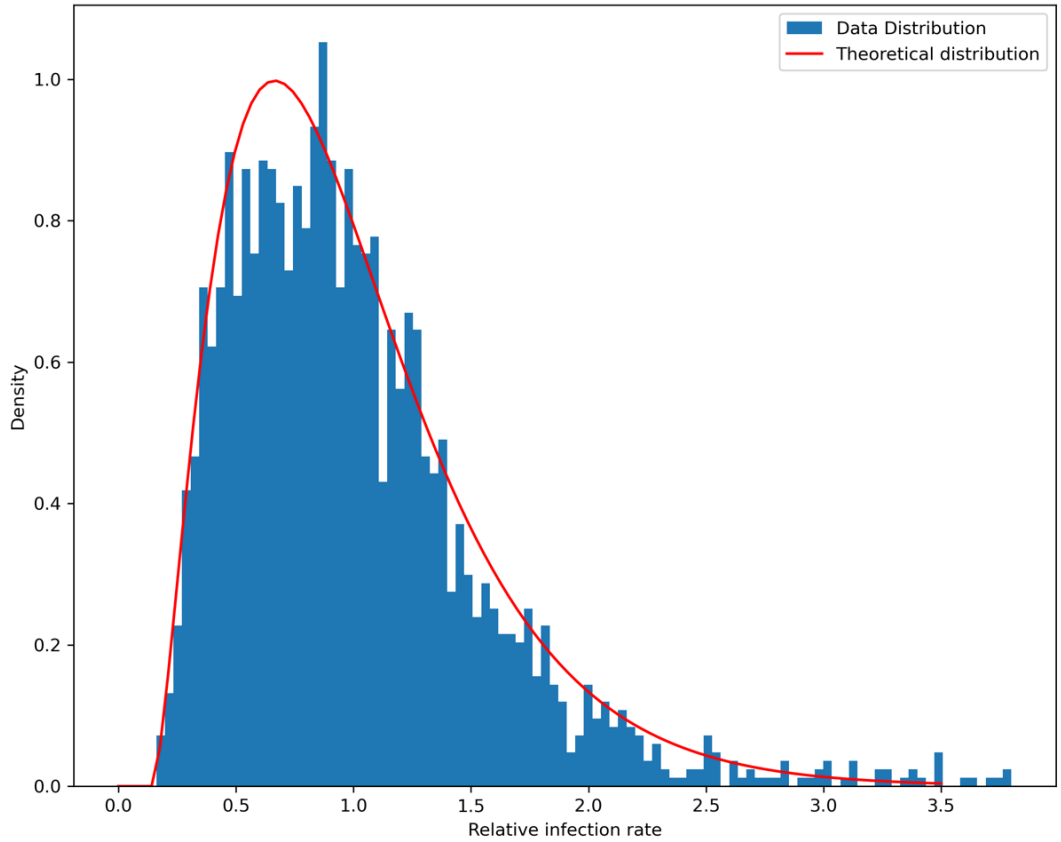
Appendix figure S1: State-week group assignment based on the dominant variant. The x-axis represents each week, the y-axis represents each state, and the color represents the assignment of each state-week pair.

1.2 Distribution of outcome variables

We selected Gamma distribution for the outcome variable to account for the right skewed distribution of the variable. We further verified the choose of Gamma distribution by fitting data with theoretical Gamma distribution. The results are shown in figures below:



Appendix figure S2: Comparing relative hospitalization rate distribution with a fitted Gamma distribution. The red line represents fitted Gamma distribution and green histogram is the distribution of relative hospitalization rate.



Appendix figure S3: Comparing relative infection rate distribution with a fitted Gamma distribution. The red line represents fitted Gamma distribution and blue histogram is the distribution of relative infection rate.

1.3 Variables description

Outcome variable

Hospitalization rates: In this study we used hospitalization rates (HR_i^t) as a proxy for COVID-19 severity, and the outcome variable of interest for our model. COVID-19 hospitalization rates represent both the severity of COVID-19 disease in a population and the burden it places on the healthcare system. Hospitalization rate (HR_i^t) for each state i and week t is defined as follow:

$$HR_i^t = \frac{H_i^t}{C_i^{t-1}},$$

where H_i^t is the number of hospitalizations for state i and week t and C_i^{t-1} is the number of confirmed cases for state i and week $t-1$. We applied a one-week lag between reported confirmed cases and hospitalizations to account for the time between symptom onset and hospital admission.[2] Weekly state-level HR_i^t was treated as the outcome variable in this analysis and served as an indicator for the severity of COVID-19 for a given time and location. The infection rate is also modeled as an outcome variable, the related models are presented in Appendix section 2.7 and the results are presented in Appendix section 3.4 to 3.7.

Time-dependent Covariates

Vaccination rate: In this analysis, the weekly cumulative COVID-19 vaccination rate is the primary variable under examination, as we hypothesize it to be a critical determinant in protecting populations against severe COVID-19 disease. The completed primary series rate was chosen over the partial vaccination rate because it represents the recommended dosage by the U.S. CDC. To address the fact that vaccine eligibility was not available to all U.S. adults until April 19th, 2021, we also ran the model using the partial vaccination rate for all variant waves, and the results are consistent with the completed primary series rate (results are presented in Appendix section 2.2). As booster shots became widely available during the Omicron wave, we also include the booster vaccination rate as a covariate to investigate its potential impact on COVID-19 health outcomes. Due to errors and anomalies in the data, the vaccination data excludes West Virginia and New Hampshire, while the booster vaccination data excludes West Virginia, New Hampshire, and North Carolina. As a result, West Virginia and New Hampshire are excluded from all analyses, and North Carolina is excluded from the booster analysis for the Omicron wave.

Trip-purpose-specific mobility: We adapted multiple mobility-derived metrics from a previous study[3] to represent destination-specific travel behaviors and engagement levels for specific types of activities, namely full-service restaurant, gym, university, and office of physician visits. Specifically, the mobility variables represent the weekly number of visits to a given destination of interest per person per week. This variable allows us to compare the relative frequency of visits to each point of interest across states and to investigate their potential impact on COVID-19 health outcomes. The metrics were generated based on anonymized mobility data from Safegraph[4], which tracks the number of visits to different types of destinations for a sample of the population at the census tract level in the U.S. Examples of destinations include full-service restaurants, gyms, and grocery stores. The original Safegraph dataset includes over 20 destination categories; thus, to reduce the complexity of the model we identified a smaller representative set of destinations to include as input in the final model. This was accomplished by first organizing the destination categories into six distinct destination groups based on the first two digits of the NAICS code,[5] namely Retail Trade (44-45), Education Services (61), Healthcare and Social Assistance (62), Arts, Entertainment, and Recreation (71), Accommodation and Food Services (72), and Other Services (81). From each group, we selected one destination category as the representative variable for the group based on the correlations between other variables within the group (Details are documented in Appendix section 1.4). Subsequently, we conducted a model selection process to identify the most appropriate subset of mobility variables from these six to be included in the final model based on concurvity and explained model variance (details are documented in Appendix section 2.1). The destinations included in the final model are the office of physicians, gym, university, and full-service restaurants.

Previous infection rates: Several studies have demonstrated the effectiveness of previous infections against reinfections and severe COVID-19 outcomes. Studies have illustrated that individuals retain a substantial level of natural immunity for six months after infection.[6–8] To attempt to account for the role of recently acquired immunity

from infection in protecting from severe disease upon reinfection in our study we generate a variable to represent the total population infected and recovered within a recent window, i.e., the total infections reported between weeks (t-16) and (t-4), which allows for time to recover and build up immunity[9] by the time period t at which the hospitalizations are modeled, but short enough that immunity has not waned. These specific prior infections variable (PI_i^t), requires multiple parameters, namely the length of the interval that infections are summed over and the start and end period of the window. To identify the best window and evaluate the sensitivity of our analysis to the chosen window length, start and end time, we conducted a sensitivity analysis. The time window with the largest deviance explained in the GAMs was selected for the final model, which was a three-month window ranging from 4 to 16 weeks prior to time t. Additional details of this sensitivity analysis and the results are included in Appendix section 2.4. The mathematical formulation of this metric is defined as follows:

$$PI_i^t = \frac{\sum_{j:t-16}^{t-4} C_i^j}{pop_i}$$

where PI_i^t represents cumulative infection rate for state i from 16 to 4 weeks prior of week t, C_i^j is the weekly confirmed cases for state i at week j, and pop_i is the population for state i. The sum in the numerator defines the summation of C_i^j for the $t-16$ to $t-4$ weeks prior to t.

Government policy: The stringency and timing of implementing government policies to mitigate the impacts of COVID-19, such as school closure, cancellation of public events, and international travel controls, are associated with different measures of epidemic severity.[10] We selected the government response index from Oxford Coronavirus Government Response Tracker (OxCGRT)[11] as our indicator for government policy. The index tracks the diversity of government responses across various policies, ranging from containment measures and closures to healthcare systems, vaccination strategies, and economic policies. This index reflects the government's response level with a number ranging from 0 to 100, the larger the number, the more substantial the response. It is available for all 50 states in the U.S. at a weekly timescale for the entire period of analysis.

Weekly testing rate: The weekly testing rates were included in the model as a potential confounding factor for multiple reasons. Firstly, it represents a proxy input feature to capture the level of healthcare infrastructure available to a population. Second, it directly impacts the hospitalization rate through the denominator (i.e., total confirmed cases), as the number of reported cases in a region is a direct function of local testing availability, thus increased testing will lead to higher case rates, and lower hospitalization rates. For example, in two locations with the same true hospitalization rate (e.g., the likelihood of a COVID-19 infection needing admittance is equal), a location with twice as much testing will detect more cases, and therefore appear to have a lower hospitalization rate. Third, higher testing rates correlate to higher Medicaid and healthcare spending (see Appendix figure S19 to S21), which can play a critical role in how a population is impacted by a disease via access to treatment and vaccinations. Fourth, increased testing can lead to more cases being identified, and thus impact people's awareness and behavior during an outbreak. For these reasons testing rate is included as a potential confounding factor in our model. We normalized the raw weekly total testing count by population to get the weekly testing rate.

Static *Covariates*

Black proportion: The proportion of the population identified as black.

Social vulnerability index: The Social Vulnerability Index utilizes data from the U.S. Census to assess the relative level of social vulnerability in each census tract. By analyzing 14 social factors, the SVI categorizes tracts into four closely interrelated themes and then aggregates them as a single indicator of social vulnerability.

Proportion of adults at high risk: The proportion of the population over 18 years old is at high risk of serious illness if infected with Coronavirus.

Medicaid spending: Total Medicaid spending for each state normalized by the population.

1.4 Preprocessing of mobility data

The 21 mobility destination categories from Safegraph were organized into six distinct industry groups based on the NAICS code. The relevant groups per the NAICS code are Retail Trade (44-45), Education Services (61), Healthcare and Social Assistance (62), Arts, Entertainment, and Recreation (71), Accommodation and Food Services (72), and Other Services (81).[2] The details of generating visits to each POI are documented in previous publication.

Appendix table S1: Description of each industry group and the corresponding destination categories.

Retail Trade (44-45)	Education Services (61)	Healthcare and Social Assistance (62)	Arts, Entertainment, and Recreation (71)	Accommodation and Food Services (72)	Other Services (81)
Automotive Store (441310)	Elementary School (611110)	Office of Physician (621111)	Parks (712190)	Hotels (721110)	Religious Organizations (813100)
Hardware Store (444130)	University (611310)	Child Day Care (624410)	Gym (713940)	Full-Service Restaurant (722511)	
Grocery Store (445110)				Cafes, Snacks, Bars (722515)	
Convenience Store (445120)				Limited-Service Restaurant (722513)	
Pharmacies (446110)					
Gas Station (447110)					
Sporting Goods Store (451110)					
Department Store (452210)					
Other General Store (452319)					
Used Merchandise Store (453310)					

To reduce the complexity of the model, we selected one destination category as the representative variable for each industry group. For the industry groups with more than 3 destination categories, we conducted a Pearson's correlation analysis and selected the variable that had the highest correlation to the other destination categories in each group. This method selected Gas Stations and Full-Service Restaurant from the Retail Trade (44-45) and Accommodation and Food Services (72) as the representative variable for each industry group. For the Educational Services (61) and Healthcare and Social Assistance (62) groups, we selected University and Office of Physician as the representative variables based on studies that indicated SARS-CoV-2 infection severity is lower in adolescents than adults.[3] For the Arts, Entertainment, and Recreation (71) industry group, we selected Gym as the representative variable instead of Parks because studies have identified park use to have a minor effect on COVID-19 transmission compared to other mobility destinations.[4] Religious Organizations was selected from the Other Services (81) industry group because it is the only destination category present.

2. Supplementary Methods

2.1 Dynamic mobility variables selection

The GAMs fit outcome variables with smoothed independent variables, allowing the nonlinear relationships between input and output. However, the nonlinear variables smoothing sometimes can result in concurvity issues. Concurvity occurs when some smooth term in a model could be approximated by one or more of the other smooth terms, leading to inaccurate estimates of the effect for given variables. In this section, we conduct model selection to ensure the validity of our model and to detect and mitigate any concurvity issues that may arise, using mobility data selected

from Appendix section 1.4 and other independent variables. The significance level and the concurvity for each variable for every model are reported in table S2 below:

Appendix table S2: Significance level and concurvity for all variables.

	All variants		Pre-Delta wave		Delta wave		Omicron wave	
Variable	Signif	Concurvity	Signif	Concurvity	Signif	Concurvity	Signif	Concurvity
Relative completed primary series rate	***	0.71	***	0.70	***	0.76	***	0.79
Relative previous infection rate (12 weeks)	***	0.12	***	0.56	***	0.28	***	0.37
Relative full-service restaurant rate	***	0.75	***	0.75	***	0.75		0.81
Relative university visits	***	0.48		0.67	***	0.49	***	0.50
Relative gas station visits	*	0.81		0.87	***	0.80		0.87
Relative gym visits	**	0.60	**	0.63	*	0.67	*	0.62
Relative religious organization visits	*	0.80	*	0.80		0.82	**	0.84
Relative office of physician visits	***	0.45	***	0.62	*	0.46	***	0.52
Relative weekly testing rate	***	0.44	***	0.69	***	0.45	***	0.41
Relative government response index	***	0.25	***	0.76	***	0.27	***	0.23

Significance codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, '': > 0.1.

The value of concurvity range from 0 to 1, the higher the concurvity the more a smooth variable can be approximated by the smooth of other variables. Based on the results form table S2, we removed relative gas station visits and relative religious organization visits from the model. This decision was based on their lack of significance and/or their high concurvity values. The equivalent results for selected variables are presented in table S3.

Appendix table S3: Significance level and concurvity for selected variables.

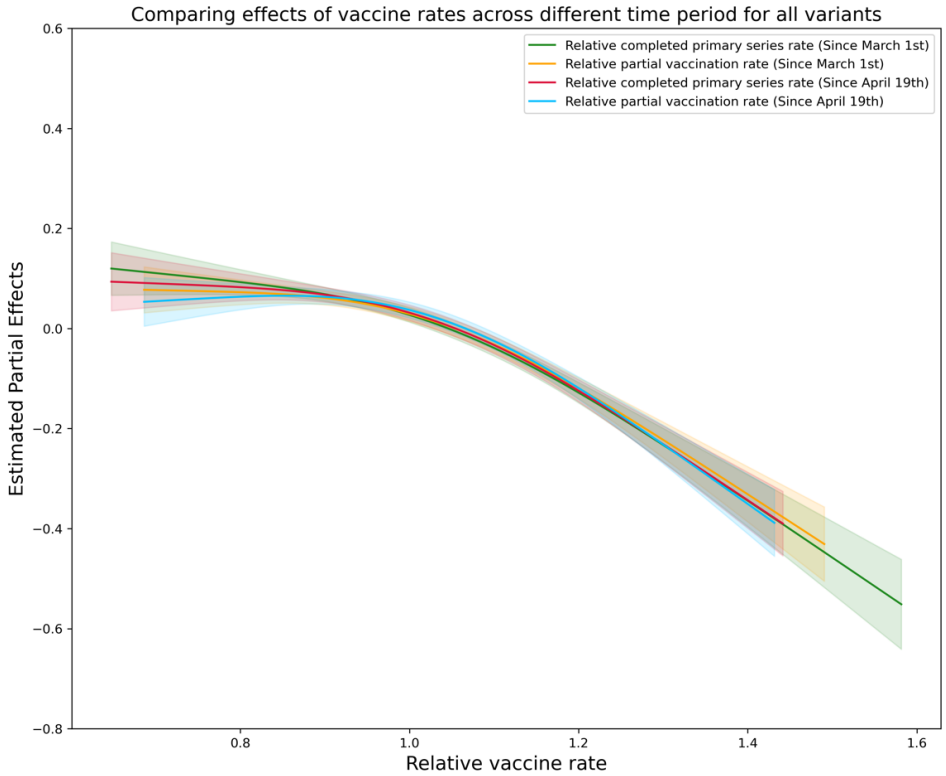
	All variants		Pre-Delta wave		Delta wave		Omicron wave	
Variable	Signif	Concurvity	Signif	Concurvity	Signif	Concurvity	Signif	Concurvity
Relative completed primary series rate	***	0.53	***	0.60	***	0.56	***	0.59
Relative previous infection rate (12 weeks)	***	0.12	***	0.49	***	0.24	***	0.34
Relative full-service restaurant rate	***	0.57	***	0.58	***	0.55		0.68
Relative university visits	***	0.34		0.53	***	0.39	***	0.35
Relative gym visits	***	0.46	***	0.54	**	0.52	.	0.51
Relative office of physician visits	***	0.43	***	0.59	**	0.43	***	0.52
Relative weekly testing rate	***	0.41	***	0.69	***	0.41	***	0.37
Relative government response index	***	0.24	***	0.69	***	0.25	***	0.20

Significance codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, '': > 0.1.

The results from table S3 reveal that each variable exhibits a concurvity value below 0.7 and is significant in at least three out of four models.

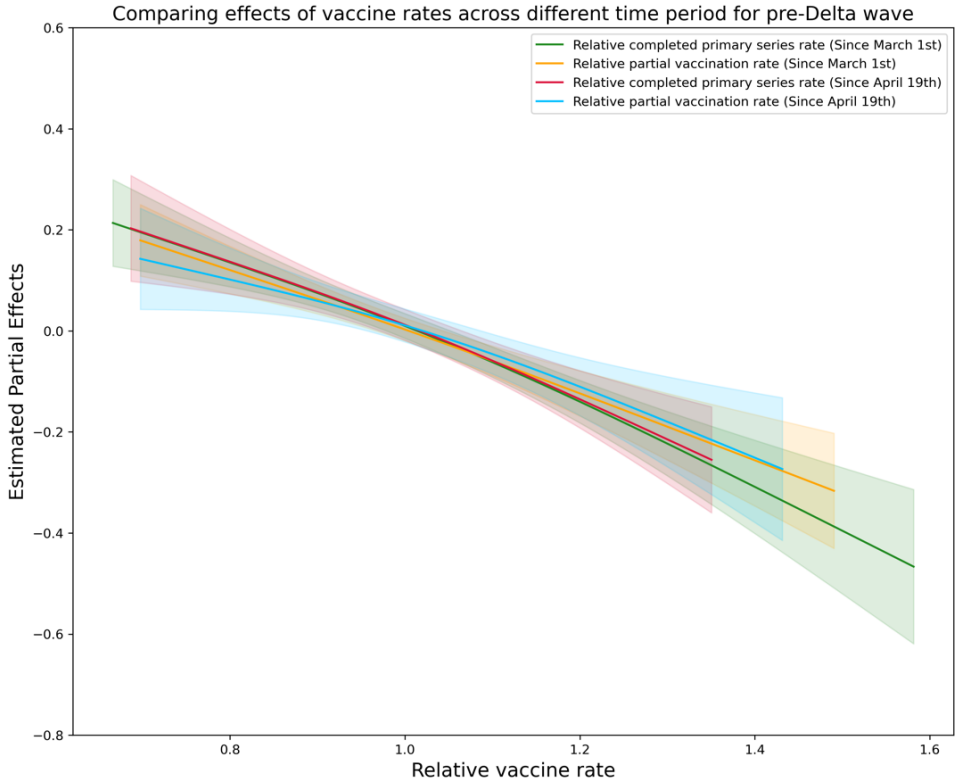
2.2 Robustness check of vaccination data

We selected the completed primary series rate as the main vaccination variable in the main analysis. To assess the robustness and validity of our findings, we conducted additional analyses using different vaccination data (completed primary series rate, and partial vaccination rate) and varying starting dates (March 8th, 2021, and April 19th, 2021) for the analysis. We applied our sensitivity analysis to the models involving all variants and the pre-Delta waves, as these were the only ones that could be affected by the analysis. The results of four different combinations of vaccination data and analysis starting date for all variants are shown in Appendix figure S4 below:



Appendix figure S4: Robustness check of vaccination rate with different vaccination data and varying starting date for the model of all variants.

The results of four different combination of vaccination data and starting date for the pre-Delta wave are shown in Appendix figure S5 below:



Appendix figure S5: Robustness check of vaccination rate with different vaccination data and varying starting date for the model of the pre-Delta wave.

The finding from this robustness check demonstrated a strong and consistent impact of vaccination, independent of the chosen vaccination data or the starting date of the analysis. This consistency suggests the robustness of our results and highlights the robustness of the completed primary series rate as the main vaccination variable.

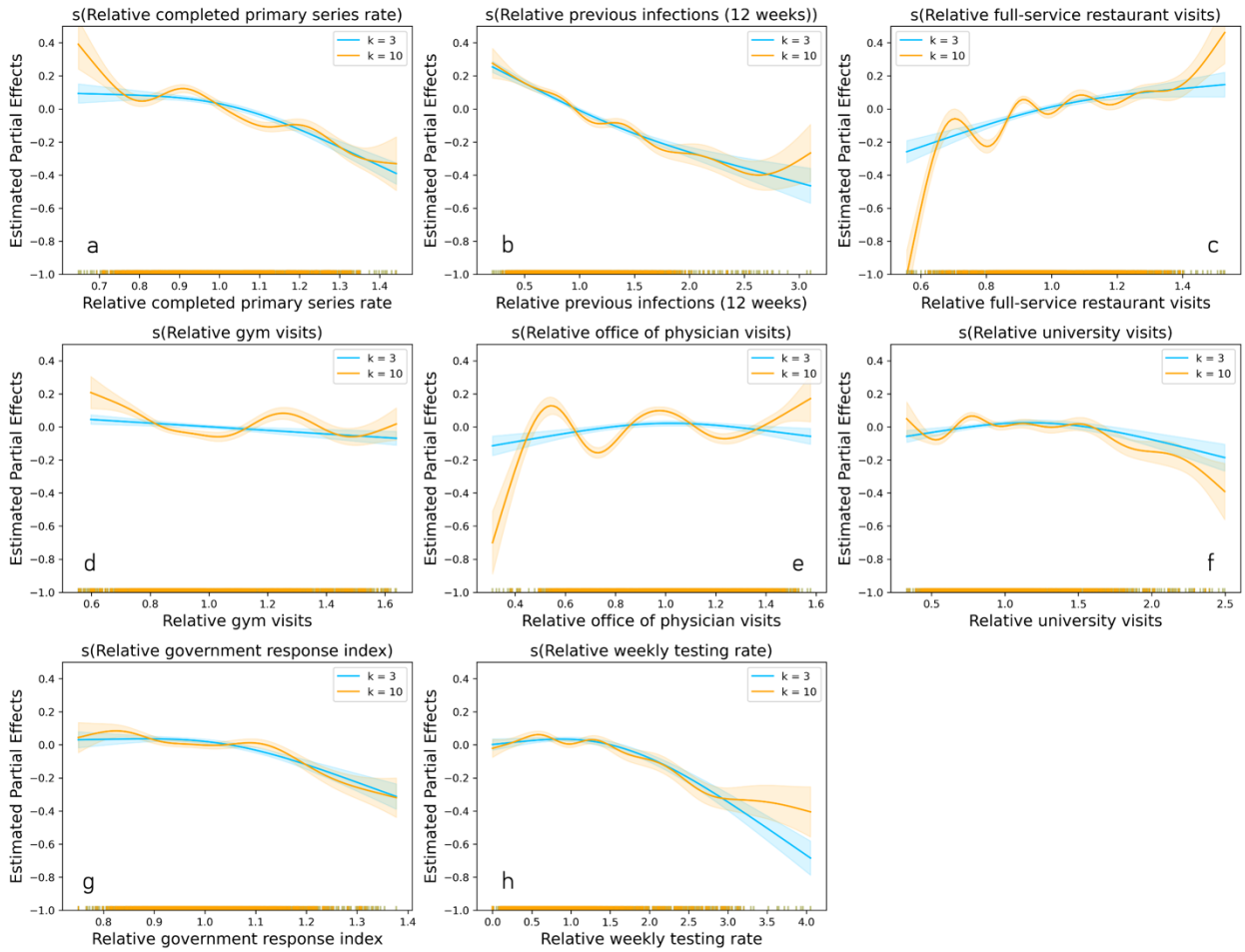
2.3 Sensitivity analysis of smooth basis dimension (k)

The parameter k controls the number of basis functions used to construct the smoothed curve. The larger the k, the more wiggles of the smoothed curve. We conduct the sensitivity analysis for all GAMs with k = 3 and k = 10. The results of the significance of smooth terms and estimated effects for all the GAMs are shown below:

Appendix table S4. Approximate significance of smooth terms for the model of all variants:

	k = 3			k = 10		
Variable	edf	F	p-value	edf	F	p-value
completed primary series rate	1.93	86.79	***	7.68	32.10	***
Previous infection rates (12 weeks)	1.82	220.46	***	7.43	64.36	***
Full-service restaurant visits	1.80	43.23	***	8.92	32.61	***
University visits	1.87	12.93	***	8.75	12.28	***
Gym visits	0.90	5.63	***	6.35	9.81	***
Office_of_physician visits	1.55	7.88	***	8.70	26.07	***
Weekly testing rate	1.96	89.51	***	7.40	15.27	***
Government response index	1.91	42.10	***	5.66	10.12	***

Significance codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, '': > 0.1.

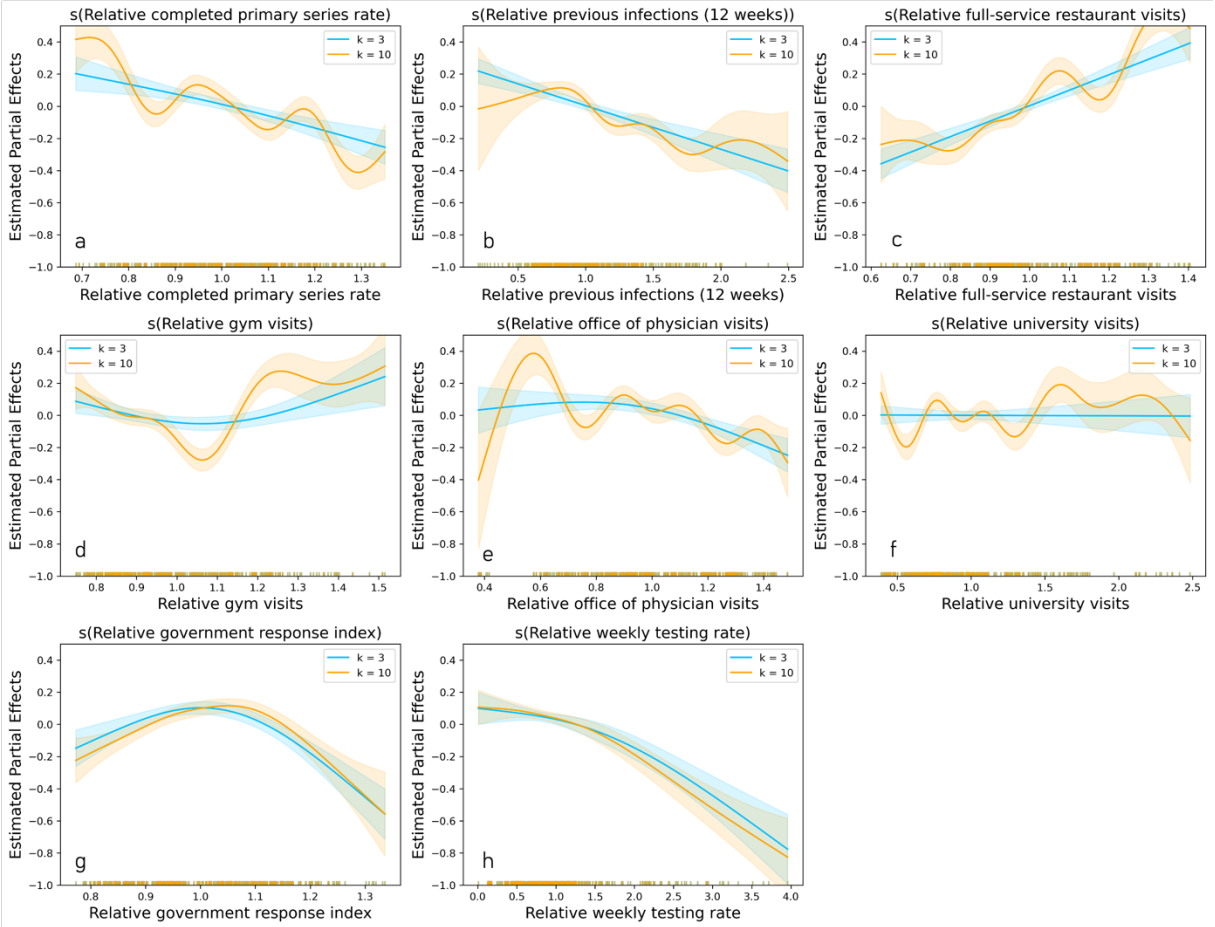


Appendix figure S6: Comparison of the estimated partial effects for the model of all variants with different basis dimensions (k = 3 and k = 10).

Appendix table S5. Approximate significance of smooth terms for the model of the pre-Delta wave:

	k = 3			k = 10		
Variable	edf	F	p-value	edf	F	p-value
completed primary series rate	0.95	10.37	***	7.19	14.94	***
Previous infection rates (12 weeks)	0.94	18.75	***	6.59	7.51	***
Full-service restaurant visits	1.00	40.58	***	7.50	20.81	***
University visits	0.26	0.16		8.80	6.36	***
Gym visits	0.97	7.35	***	6.65	9.90	***
Office_of_physician visits	1.81	13.83	***	7.63	7.18	***
Weekly testing rate	1.85	21.20	***	2.47	6.58	***
Government response index	1.92	28.17	***	3.50	6.36	***

Significance codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, '': > 0.1.

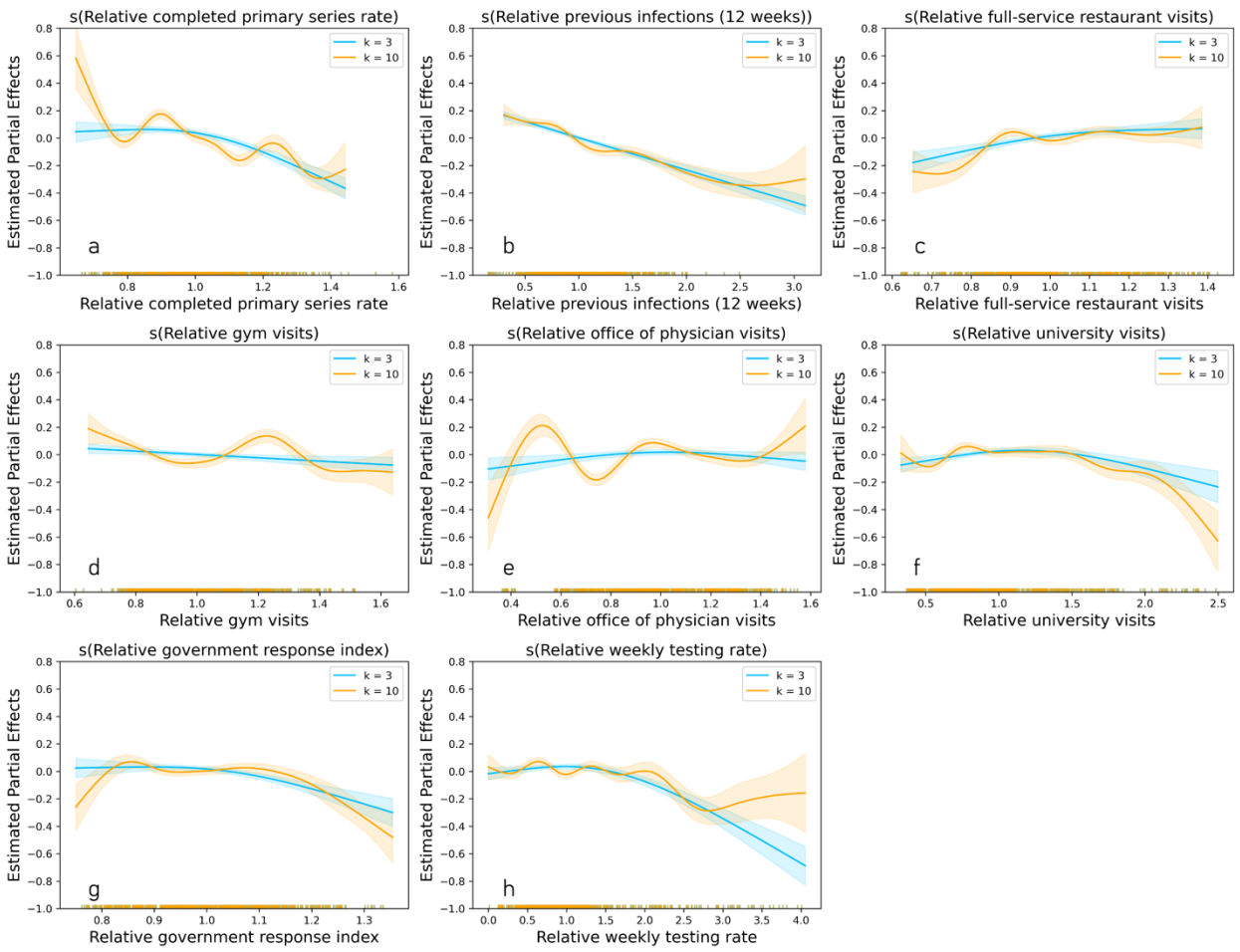


Appendix figure S7: Comparison of the estimated partial effects for the model the pre-Delta wave with different basis dimensions (k = 3 and k = 10).

Appendix table S6. Approximate significance of smooth terms for the model of the Delta wave:

	k = 3			k = 10		
Variable	edf	F	p-value	edf	F	p-value
completed primary series rate	1.92	50.40	***	8.50	24.93	***
Previous infection rates (12 weeks)	1.00	91.75	***	5.79	29.62	***
Full-service restaurant visits	1.80	17.41	***	7.88	19.89	***
University visits	1.67	12.40	***	7.64	9.04	***
Gym visits	0.82	3.08	**	6.04	8.99	***
Office_of_physician visits	1.12	3.72	*	8.70	15.56	***
Weekly testing rate	1.94	49.47	***	7.66	7.71	***
Government response index	1.88	24.69	***	4.82	4.74	***

Significance codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, '': > 0.1.

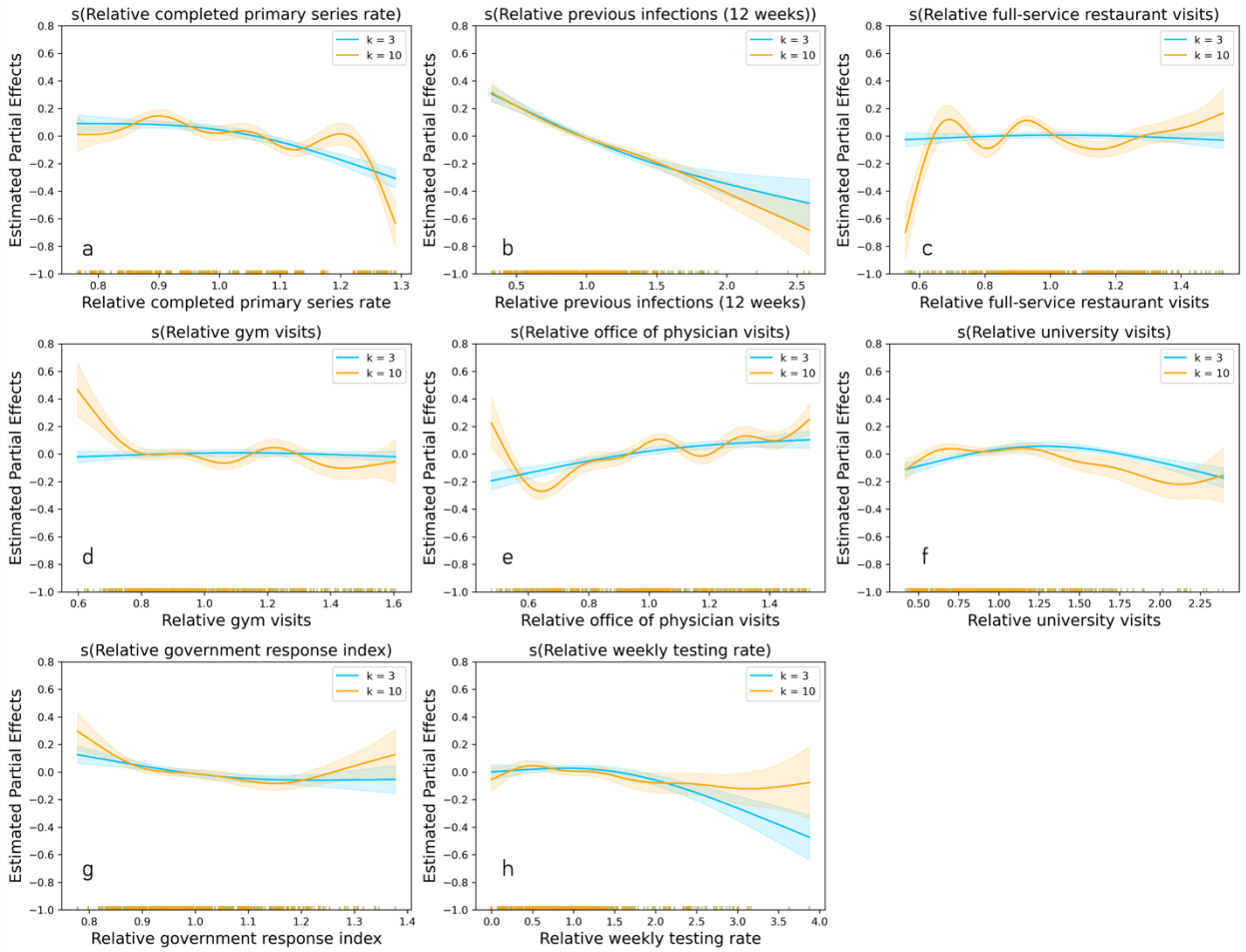


Appendix figure S8: Comparison of the estimated partial effects for the model the Delta wave with different basis dimensions (k = 3 and k = 10).

Appendix table S7. Approximate significance of smooth terms for the model of the Omicron wave:

	k = 3			k = 10		
Variable	edf	F	p-value	edf	F	p-value
completed primary series rate	1.89	45.12	***	8.63	18.36	***
Previous infection rates (12 weeks)	1.54	91.47	***	1.00	25.94	***
Full-service restaurant visits	0.07	0.04		8.74	18.61	***
University visits	0.95	13.28	***	6.89	6.28	***
Gym visits	0.47	0.64	.	7.12	6.49	***
Office_of_physician visits	1.78	23.58	***	6.93	11.54	***
Weekly testing rate	1.89	17.92	***	1.30	0.54	*
Government response index	1.33	6.42	***	2.56	4.16	***

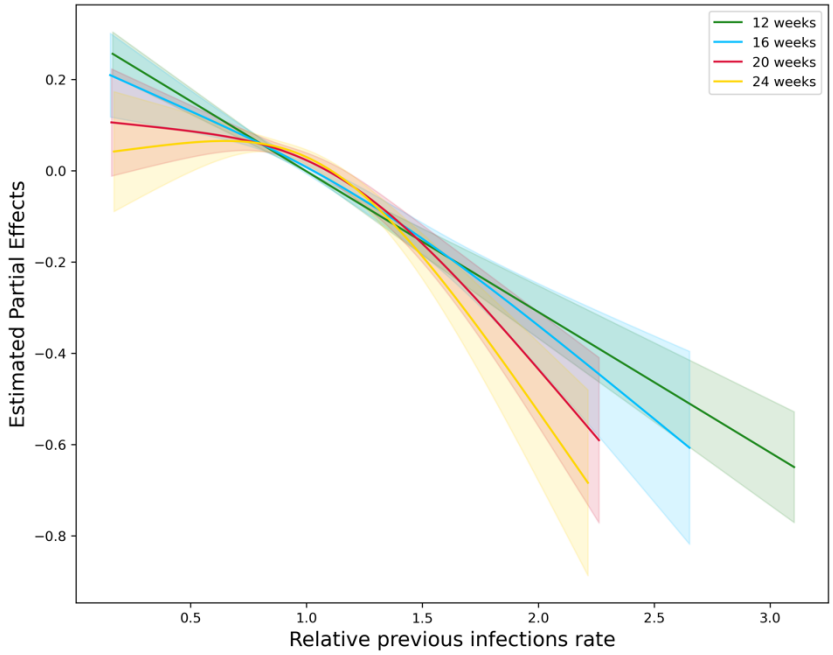
Significance codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, '': > 0.1.



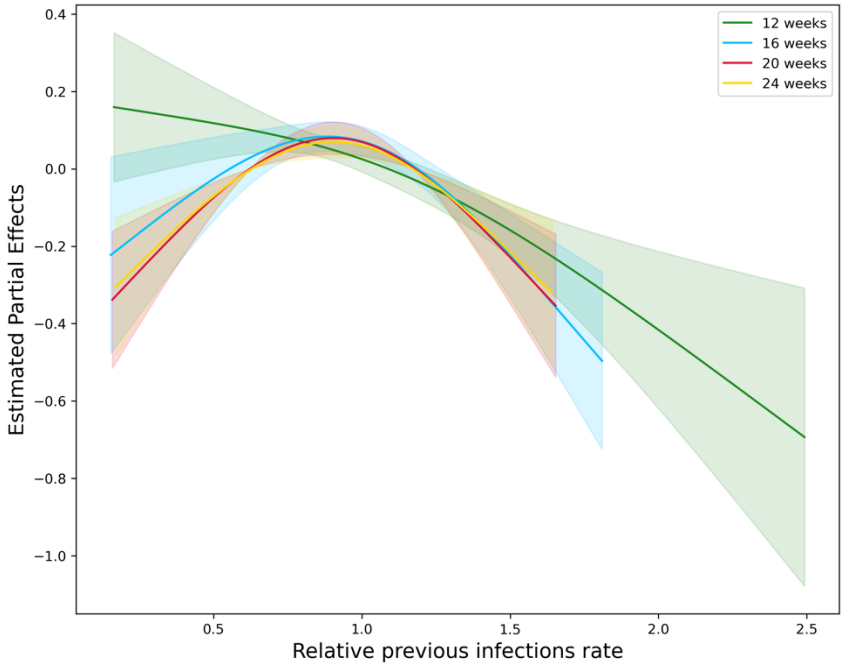
Appendix figure S9: Comparison of the estimated partial effects for the model of the Omicron wave with different basis dimensions (k = 3 and k = 10).

2.4 Sensitivity analysis of prior window length for previous infection

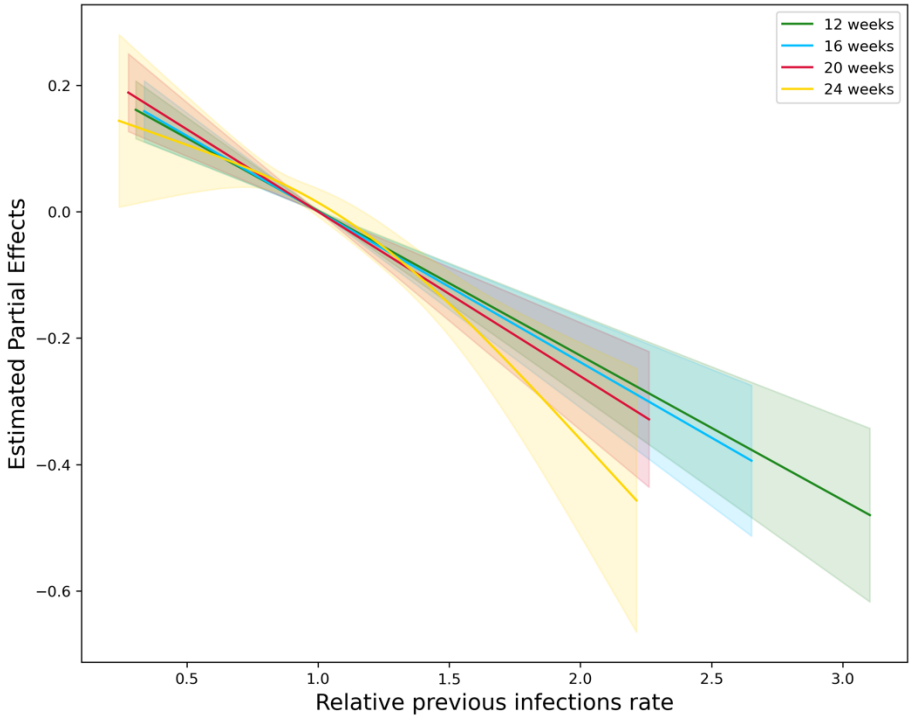
In this section, we presented a sensitivity analysis to assess the impact of the prior window length for the previous infection on our analysis. To ensure our results are robust, we fixed all other covariates and a lag of four weeks for previous infections while varying the prior window length for previous infections from 12 to 24 weeks. The results of this sensitivity analysis for each model are shown below:



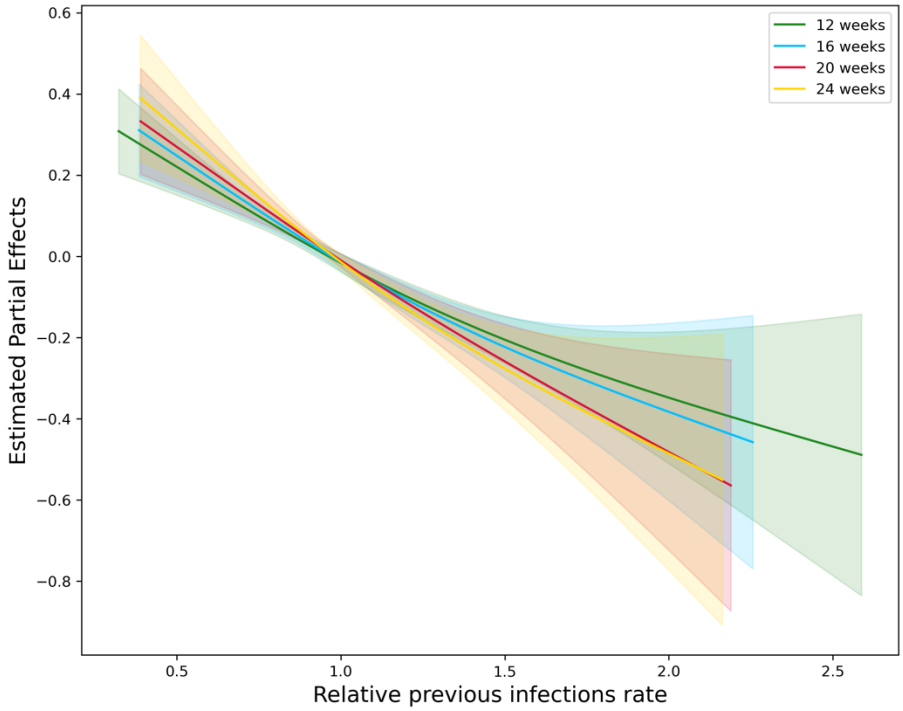
Appendix figure S10: Comparison of the estimated partial effect of the previous infection rate for the model of all variants with different prior window lengths (12, 16, 20 and 24 weeks).



Appendix figure S11: Comparison of the estimated partial effect of the previous infection rate for the model of the pre-Delta wave with different prior window lengths (12, 16, 20 and 24 weeks).



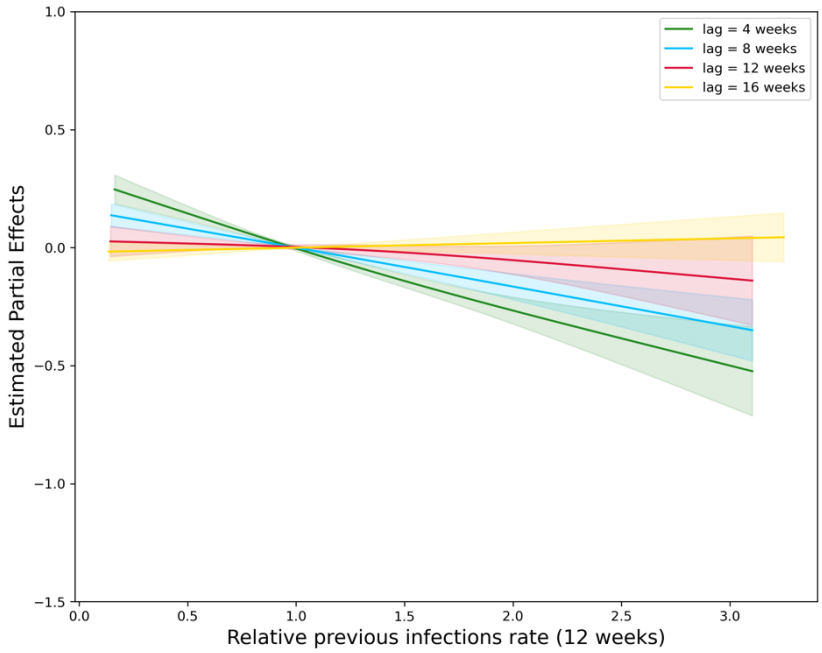
Appendix figure S12: Comparison of the estimated partial effect of the previous infection rate for the model of the Delta wave with different prior window lengths (12, 16, 20 and 24 weeks).



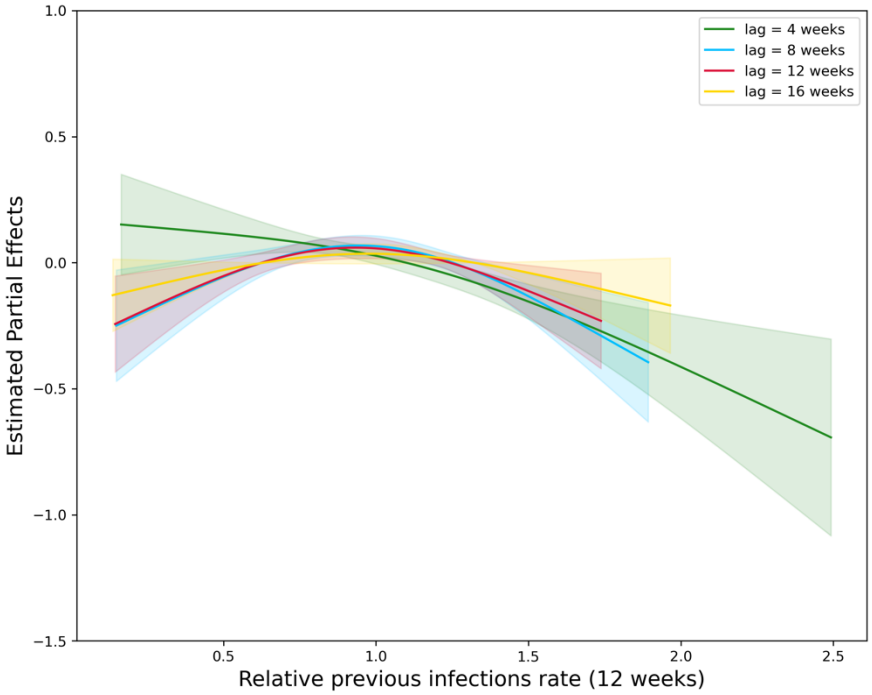
Appendix figure S13: Comparison of the estimated partial effect of the previous infection rate for the model of the Omicron wave with different prior window lengths (12, 16, 20 and 24 weeks).

2.5 Sensitivity analysis of lags for previous infection

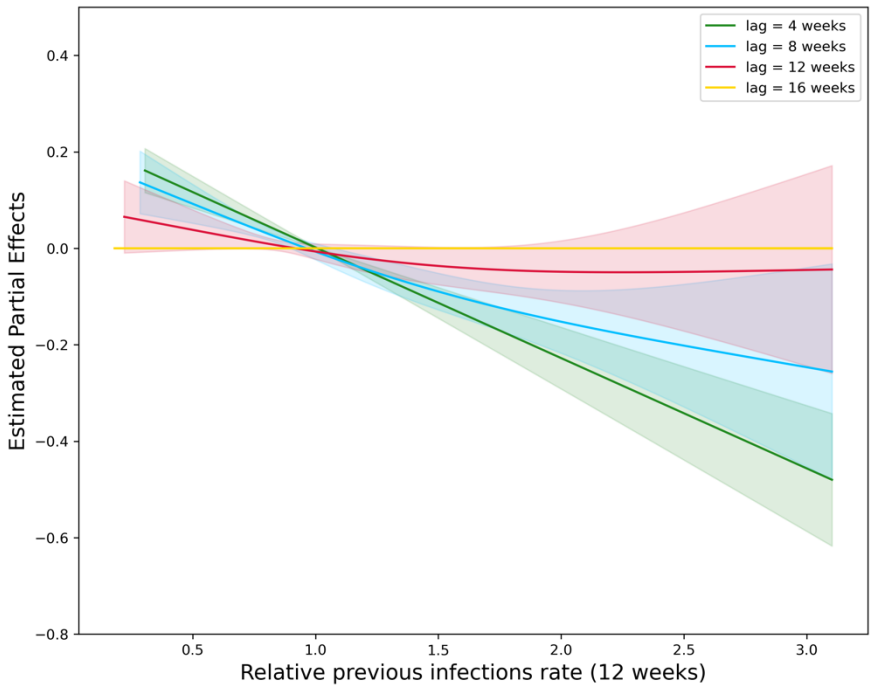
This section presents a sensitivity analysis to assess the impact of the prior window length for the previous infection on our analysis. To ensure our results are robust, we fixed all other covariates and a prior window length of 12 weeks for previous infections while varying the lag for previous infections from 4 to 16 weeks. The results of this sensitivity analysis for each model are shown below:



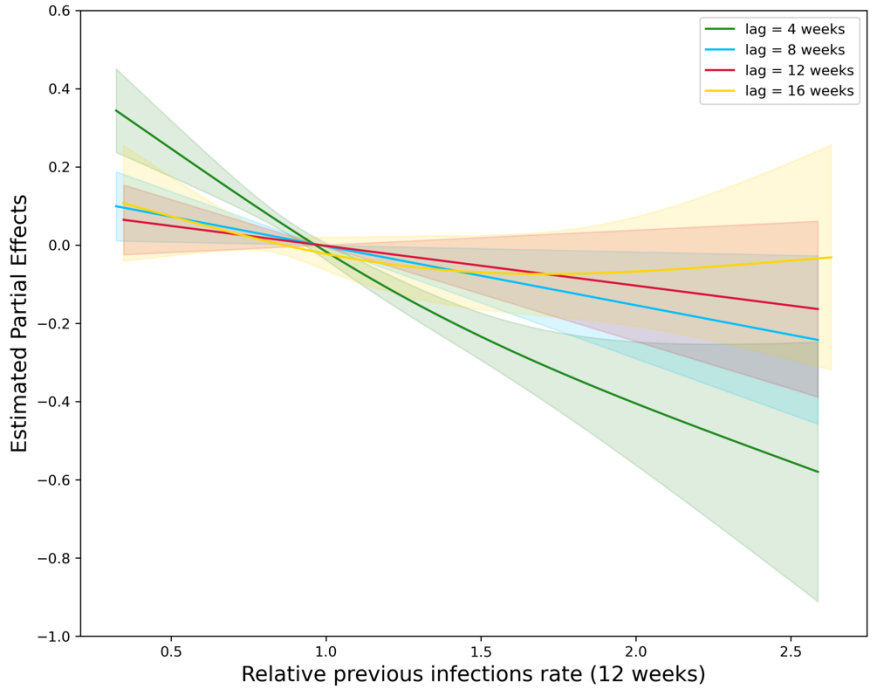
Appendix figure S14: Comparison of the estimated partial effect of the previous infection rate for the model of all variants with different lags (4, 8, 12, and 16 weeks).



Appendix figure S15: Comparison of the estimated partial effect of the previous infection rate for the model of the pre-Delta wave with different lags (4, 8, 12, and 16 weeks).



Appendix figure S16: Comparison of the estimated partial effect of the previous infection rate for the model of the Delta wave with different lags (4, 8, 12, and 16 weeks).



Appendix figure S17: Comparison of the estimated partial effect of the previous infection rate for the model of the Omicron wave with different lags (4, 8, 12, and 16 weeks).

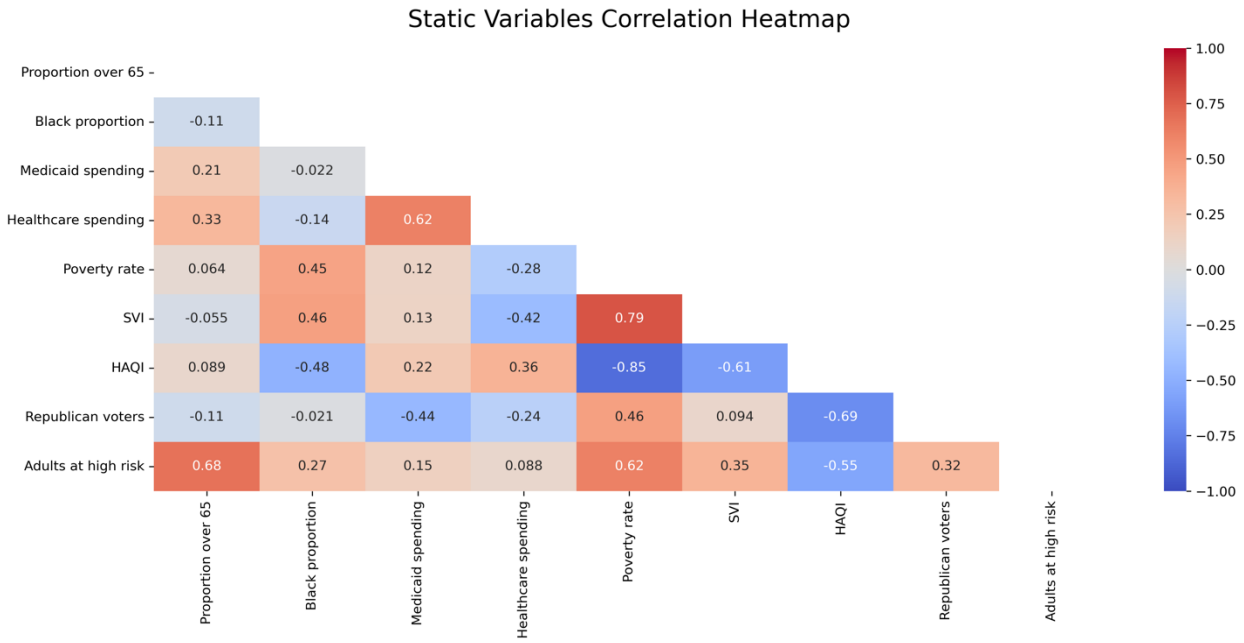
2.6 Static variables selection

We selected state-level static variables that were found to have association with COVID-19 health outcomes in a recent study.[12] These variables cover a ranging of different factors, such as socioeconomic indicators, racial demographics, age, proxy for comorbidities, political factors, and state-level healthcare expenditures. Then, a correlation analysis is performed within static variable and between static and dynamic variables to determine the suitable variables to be included in the model. Full list of static variables included in the correlation analysis are listed in Appendix table S8.

Appendix table S8: Description of static variables included in the correlation analysis.

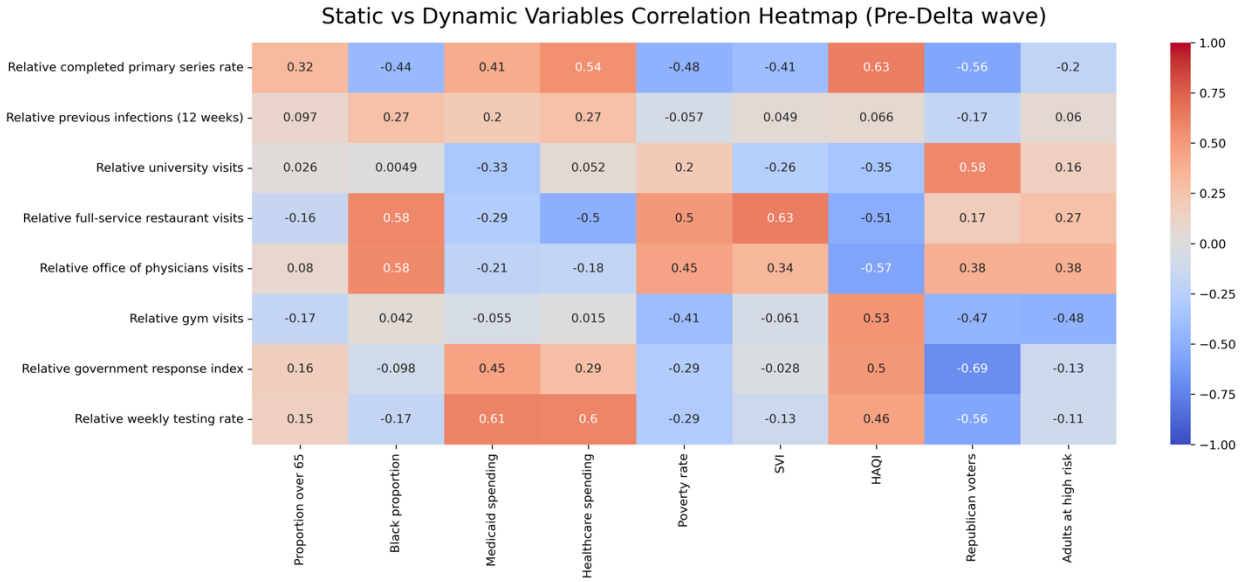
Variable name	Variable description	source
Static variables		
Black proportion	Proportion of the population identifying as Black.	13
Medicaid spending	Total Medicaid spending for each state normalized by the population.	14
Healthcare spending	Total Healthcare spending for each state normalized by the population.	15
Poverty rate	Percentage of population living below poverty line.	16
Social Vulnerability Index	A single metric based on the combination of 15 social factors, including poverty, lack of vehicle access, and crowded housing.	17
HAQI	IHME's healthcare access and quality index.	18
Republican voters	Percentage of a state's voters who voted for the 2020 Republican presidential candidate.	19
Percentage of adults at high risk	Percent of adults at higher risk of serious illness if infected with COVID-19 due to age, heart disease, COPD, asthma, diabetes, or BMI.	20
Proportion over 65	Proportion of population age 65 and older.	20

Appendix figure S18 illustrates the Pearson's correlation coefficient between each pair of static variables.



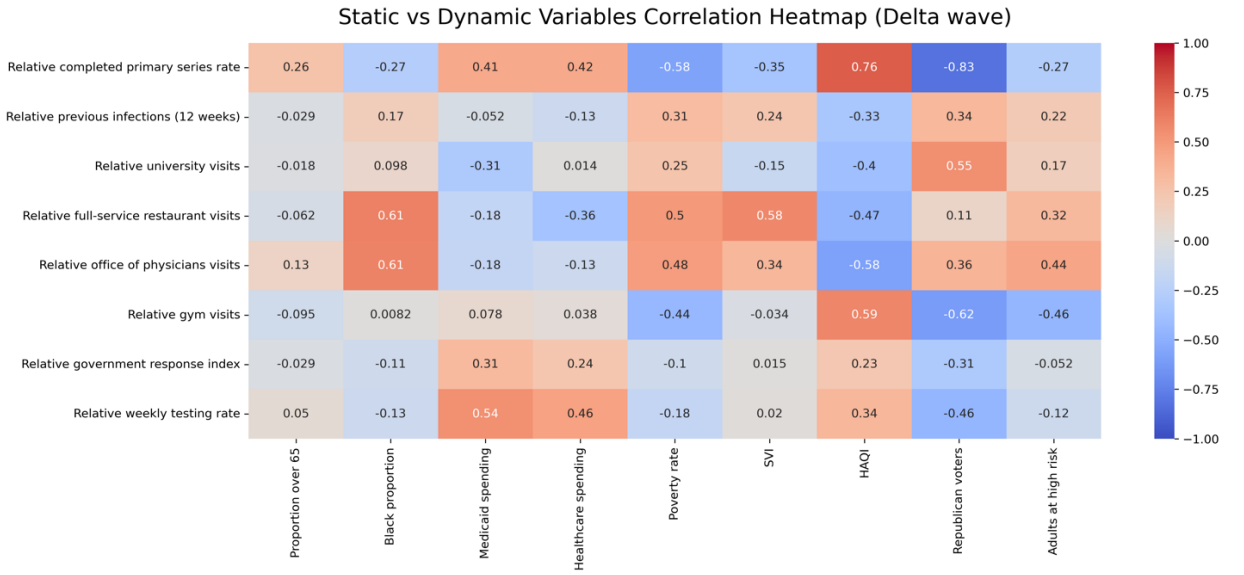
Appendix figure S18. Pearson's correlation heatmap between each pair of static variables.

Appendix figure S19 illustrates the Pearson's correlation coefficient between static and dynamic variables during the pre-Delta wave.



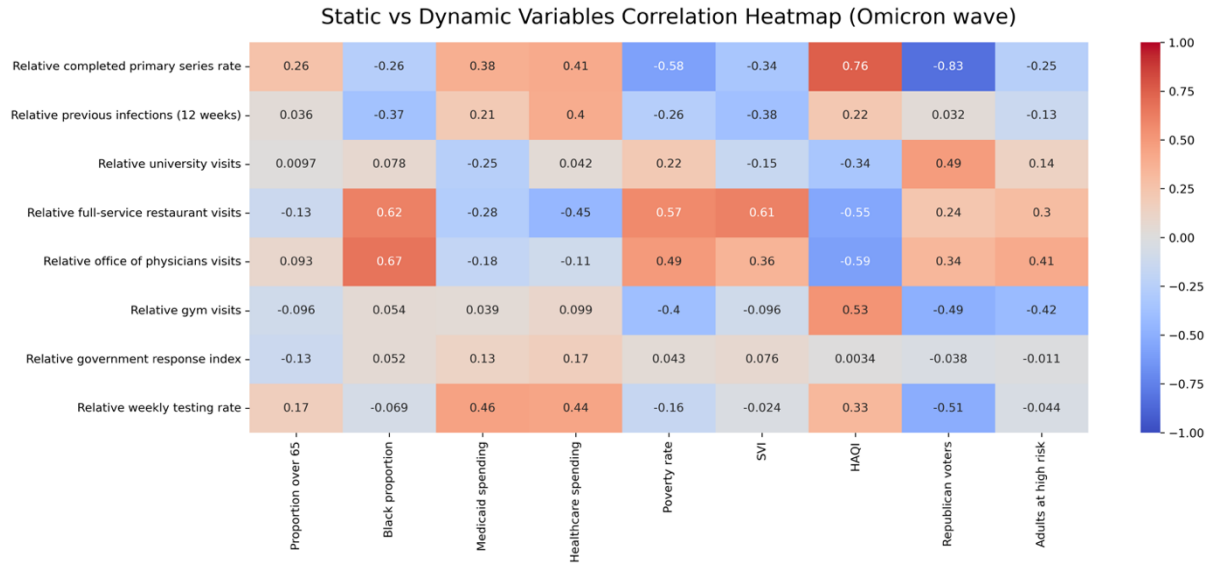
Appendix figure S19. Pearson’s correlation heatmap between static and dynamic variables during the pre-Delta wave.

Appendix figure S20 illustrates the Pearson's correlation coefficient between static and dynamic variables during the Delta wave.



Appendix figure S20. Pearson’s correlation heatmap between static and dynamic variables during the Delta wave.

Appendix figure S21 illustrates the Pearson's correlation coefficient between static and dynamic variables during the Omicron wave.



Appendix figure S21. Pearson's correlation heatmap between static and dynamic variables during the Omicron wave.

Based on the correlation analysis presented earlier, the set of static variables to be incorporated into the models was determined. The most significant dynamic mobility variables (full-service restaurant visits and office of physician visits) exhibited a high correlation with SVI and black proportion during different waves. Therefore, to ensure the accuracy and reliability of the model, we excluded all dynamic mobility variables from the GAMs with static and dynamic input variables. Next, we evaluated the complete list of static variables. We selected the black proportion as a representative variable for the race group and the SVI as a proxy for the vulnerable population. The poverty rate was dropped due to its high correlation with SVI. Additionally, we selected adults at high risk as a control variable for population-level comorbidities and dropped the proportion over 65. We also decided to include Medicaid spending over healthcare spending as the state-level proxy for healthcare expenditures. Lastly, we dropped the HAQI and Republican voters variables due to their high correlation with the completed primary series rate.

2.7 GAMs with infection rate as the outcome variable

This section elucidates models 5 to 8, which employ the relative infection rate (RIR_i^t) as the dependent variable. These models share the same framework as models 1 to 4, while the outcome variable is the infection rate. Model 5, Dyn-Inf, includes only dynamic covariates, namely relative completed primary series rate, relative weekly testing rate, relative full-service restaurant visits, relative gym visits, relative office of physician visits, relative university visits, relative government policy, and relative previous infections. Model 5 is presented in equation (S1):

$$RIR_i^t \sim Gamma(\mu, \phi)$$
$$\log(\mu) = f_1(RCPSR_i^{t-1}) + f_2(RWTR_i^{t-1}) + f_3(RFSRV_i^{t-1}) + f_4(RGV_i^{t-1}) + f_5(ROPV_i^{t-1}) + f_6(RUV_i^{t-1}) + f_8(RGP_i^{t-1}) + f_7(RPI_i^t) \quad (S1)$$

Model 6, Dyn-Inf-Booster is applied for only the Omicron wave and includes an interaction effect between the relative completed primary series rate ($RCPSR_i^t$) and the relative booster rate (RBR_i^t). Model 6 is presented in equation (S2):

$$RIR_i^t \sim Gamma(\mu, \phi)$$
$$log(\mu) = f_1(RCPSR_i^{t-1}, RBR_i^{t-1}) + f_2(RWTR_i^{t-1}) + f_3(RFSRV_i^{t-1}) + f_4(RGV_i^{t-1}) + f_5(ROPV_i^{t-1}) + f_6(RUV_i^{t-1}) + f_8(RGP_i^{t-1}) + f_7(RPI_i^t) \quad (S2)$$

Model 7, Int-Inf, includes both static and dynamic covariates, specifically removing the dynamic variables relative full-service restaurant visits, relative gym visits, relative office of physician visits, and relative university visits, and

including the static variables, black proportion, the social vulnerability index (SVI), the proportion of adults at high risk, and Medicaid spending. Model 7 is represented by equation (S3) as follows:

$$RIR_i^t \sim Gamma(\mu, \phi)$$
$$log(\mu) = f_1(RCPSR_i^{t-1}) + f_2(RWTR_i^{t-1}) + f_8(RGP_i^{t-1}) + f_7(RPI_i^t) + f_5(\text{Black proportion}) + f_6(\text{social vulnerability index}) + f_7(\text{adults at high risk}) + f_8(\text{Medicaid spending}) \quad (S3)$$

Model 8, Int-Hosp-Booster, is analogous to Dyn-Hosp-Booster, with an additional interaction effect between the relative completed primary series rate ($RCPSR_i^t$) and the relative booster rate (RBR_i^t). Model 8 is represented by equation (S4) as follows:

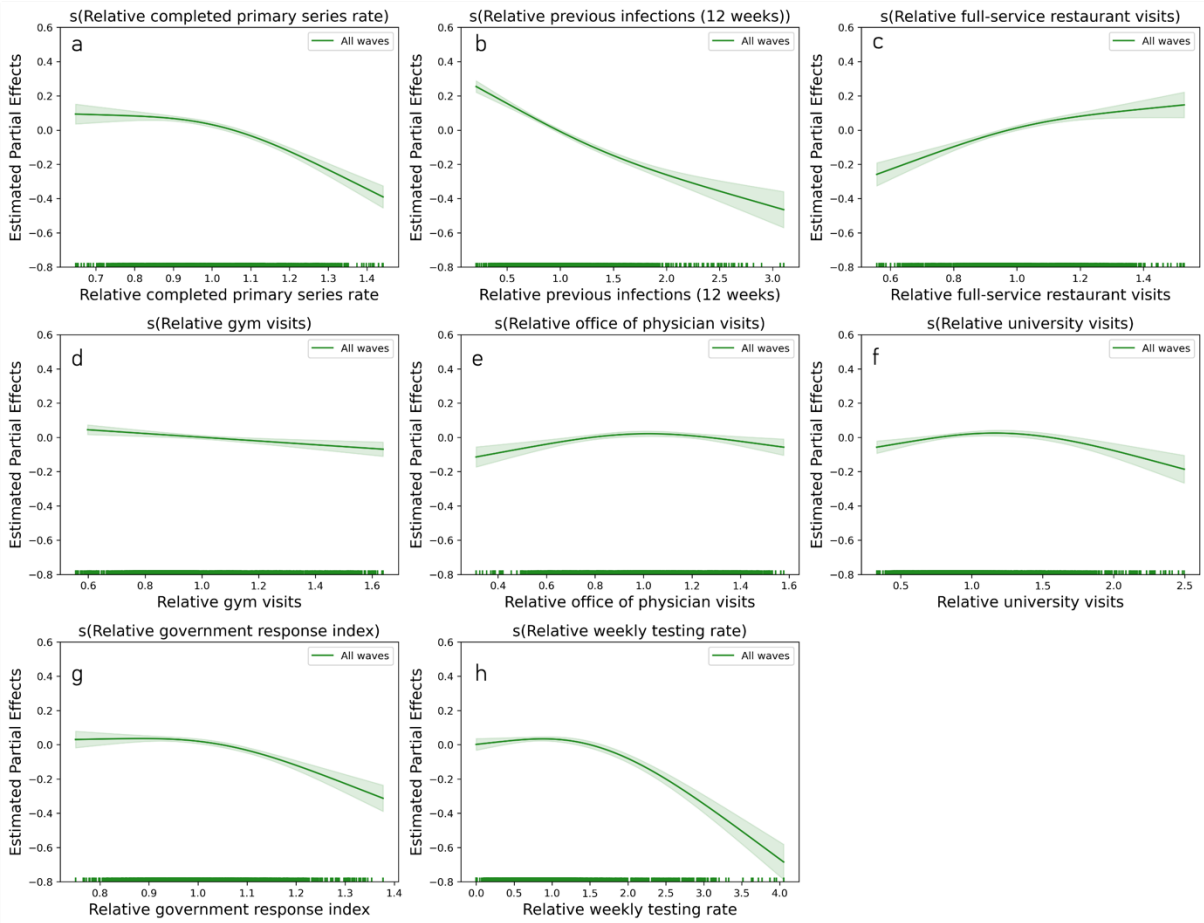
$$RIR_i^t \sim Gamma(\mu, \phi)$$
$$log(\mu) = f_1(RCPSR_i^{t-1}, RBR_i^{t-1}) + f_2(RWTR_i^{t-1}) + f_8(RGP_i^{t-1}) + f_7(RPI_i^t) + f_5(\text{Black proportion}) + f_6(\text{social vulnerability index}) + f_7(\text{adults at high risk}) + f_8(\text{Medicaid spending}) \quad (S4)$$

The results for Models 4-8, with infection rate as the outcome variable, are presented in Appendix section 3.3 and 3.4.

3. Supplementary Results

3.1 Results for Model 1 (Dyn-Hosp)

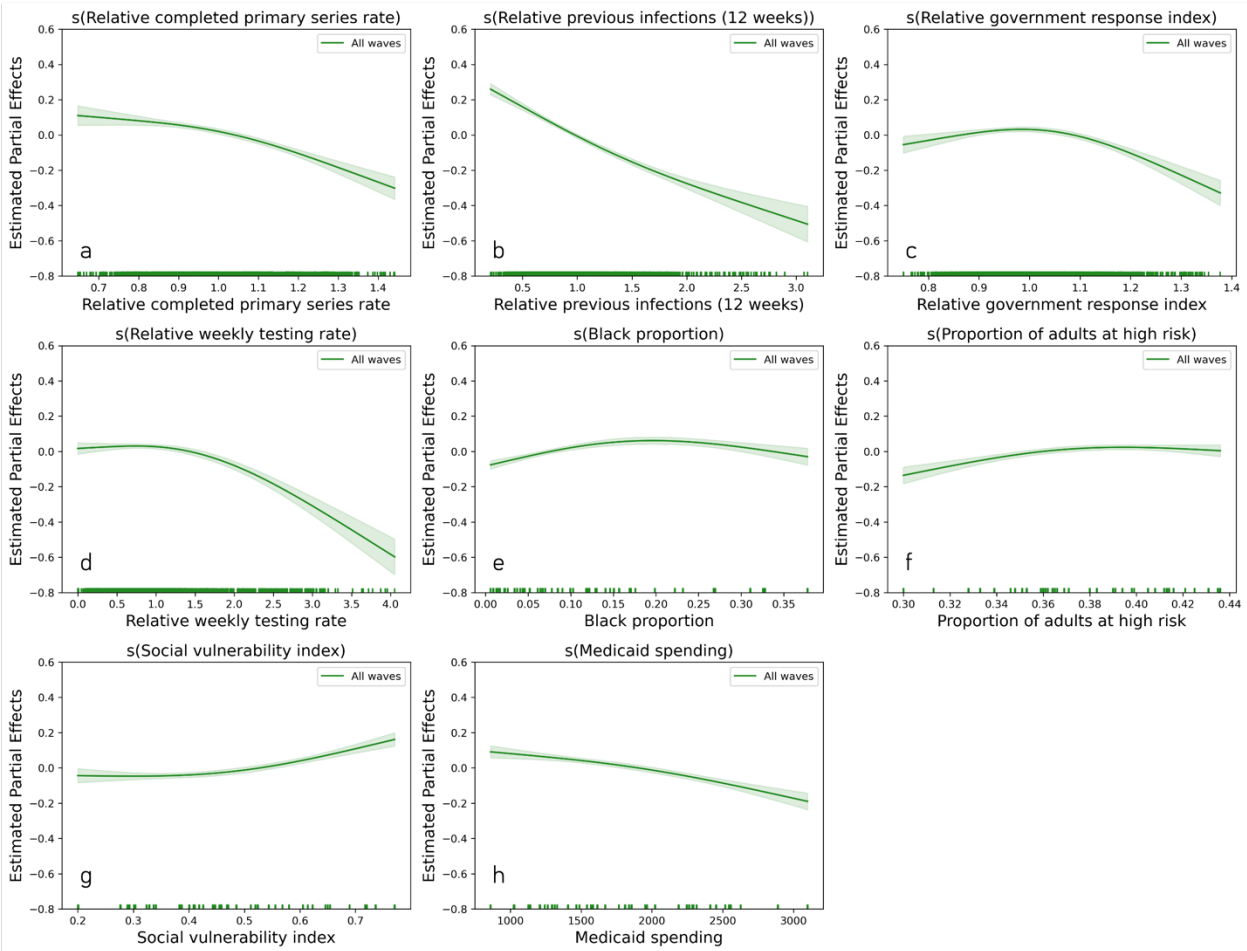
Appendix figure S22 displays the estimated partial effects obtained from the Model 1 (Dyn-Hosp) for all state-week pairs. Each subfigure corresponds to the estimated impact of an independent variable on the outcome variable. The x-axis represents the values of the independent variable, while the y-axis represents its estimated effects. The dots plotted on the x-axis indicate the distribution of the independent variable. The magnitude of the y-axis is uniform across all subplots, allowing for a comparison of the relative impact of each variable on the outcome variable.



Appendix figure S22. Estimated partial effects of dependent variables by all state-week pairs from Dyn-Hosp. The dots on the x-axis represent the distribution of the variables.

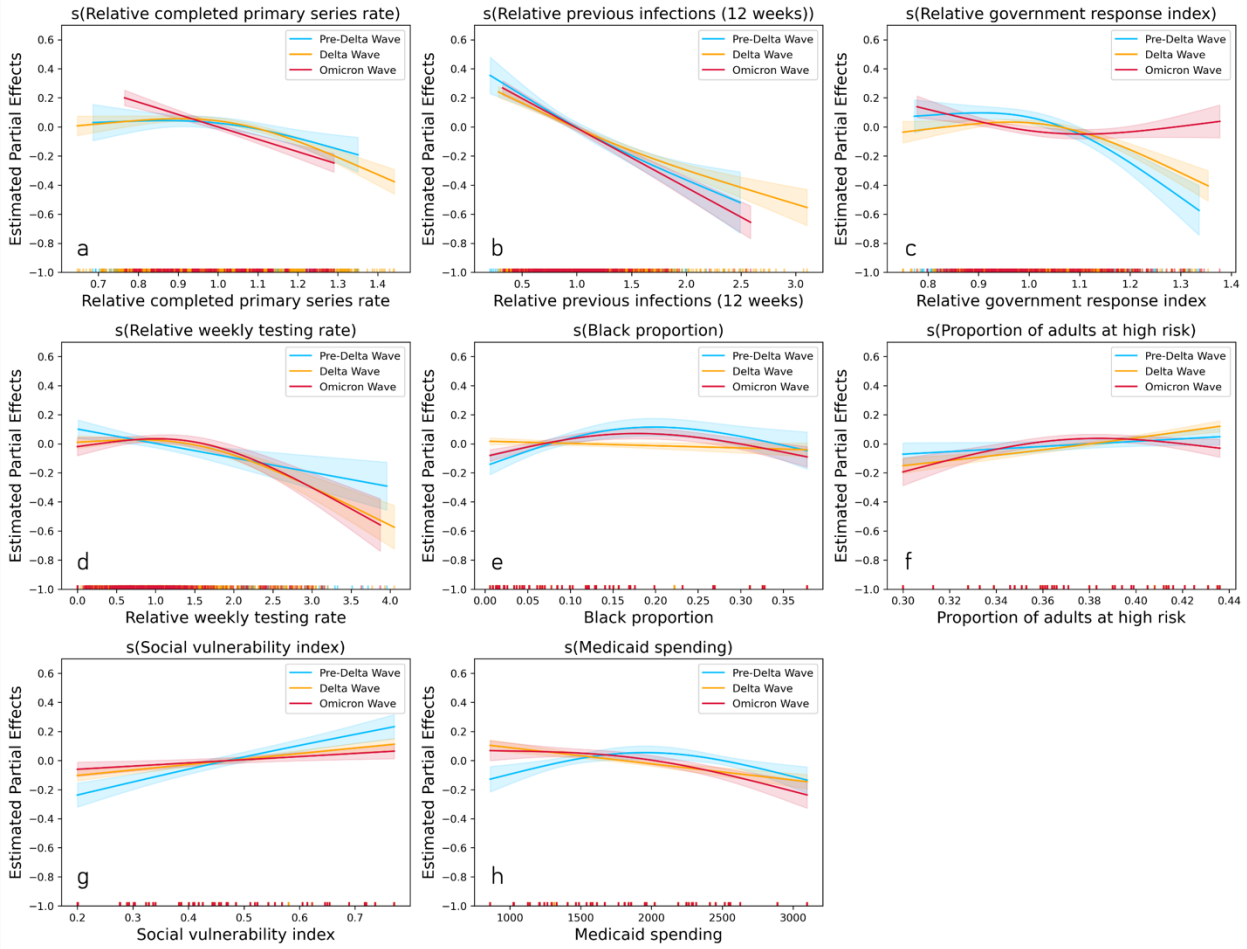
3.2 Results for Model 3 (Int-Hosp)

Appendix figure S23 displays the estimated partial effects for hospitalization rate obtained from the Model 3 (Int-Hosp) for all state-week pairs. Each subfigure corresponds to the estimated impact of an independent variable on the outcome variable. The x-axis represents the values of the independent variable, while the y-axis represents its estimated effects. The dots plotted on the x-axis indicate the distribution of the independent variable. The magnitude of the y-axis is uniform across all subplots, allowing for a comparison of the relative impact of each variable on the outcome variable (hospitalization rate).



Appendix figure S23. Estimated partial effects of dependent variables by all state-week pairs from Int-Hosp. The dots on the x-axis represent the distribution of the variables.

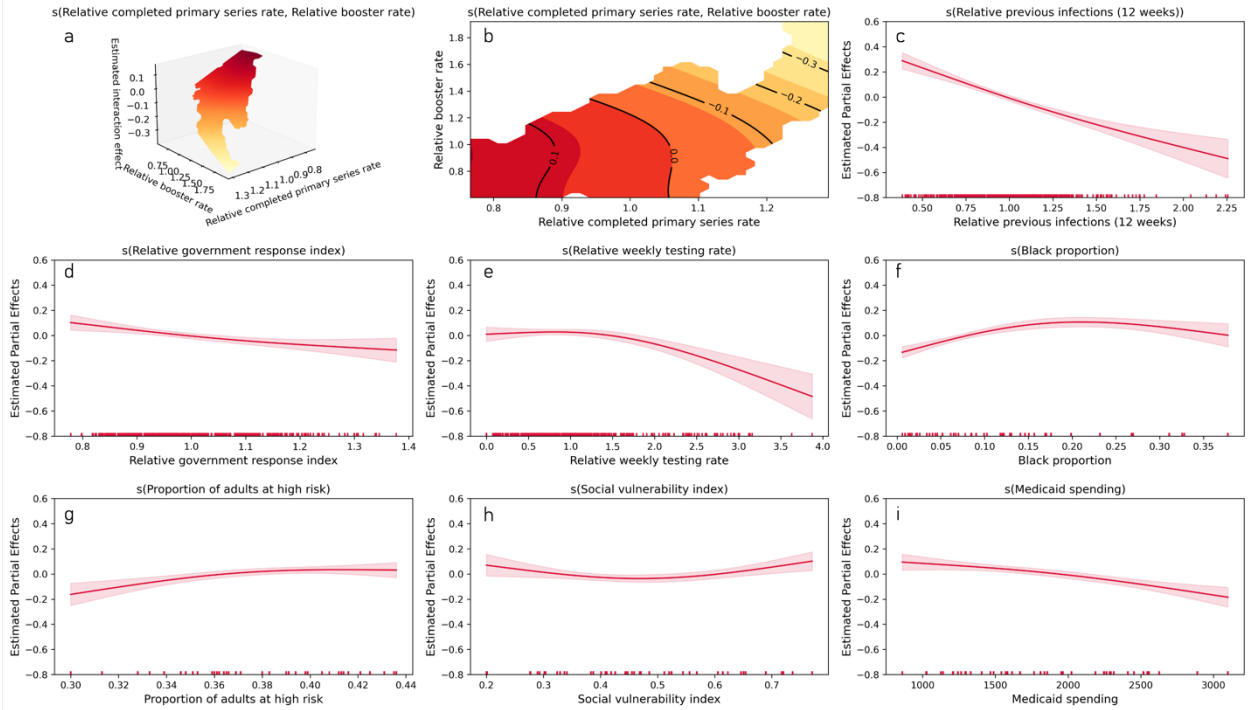
Appendix figure S24 illustrates the results for Model 3 (Int-Hosp) for each variant, namely, the Pre-Delta (Blue), Delta (Orange), and Omicron waves (Red), respectively.



Appendix figure S24. Estimated partial effects of dependent variables by different variant waves from Int-Hosp. The dots on the x-axis represent the distribution of the variables.

3.3 Results for Model 4 (Int-Hosp-Booster)

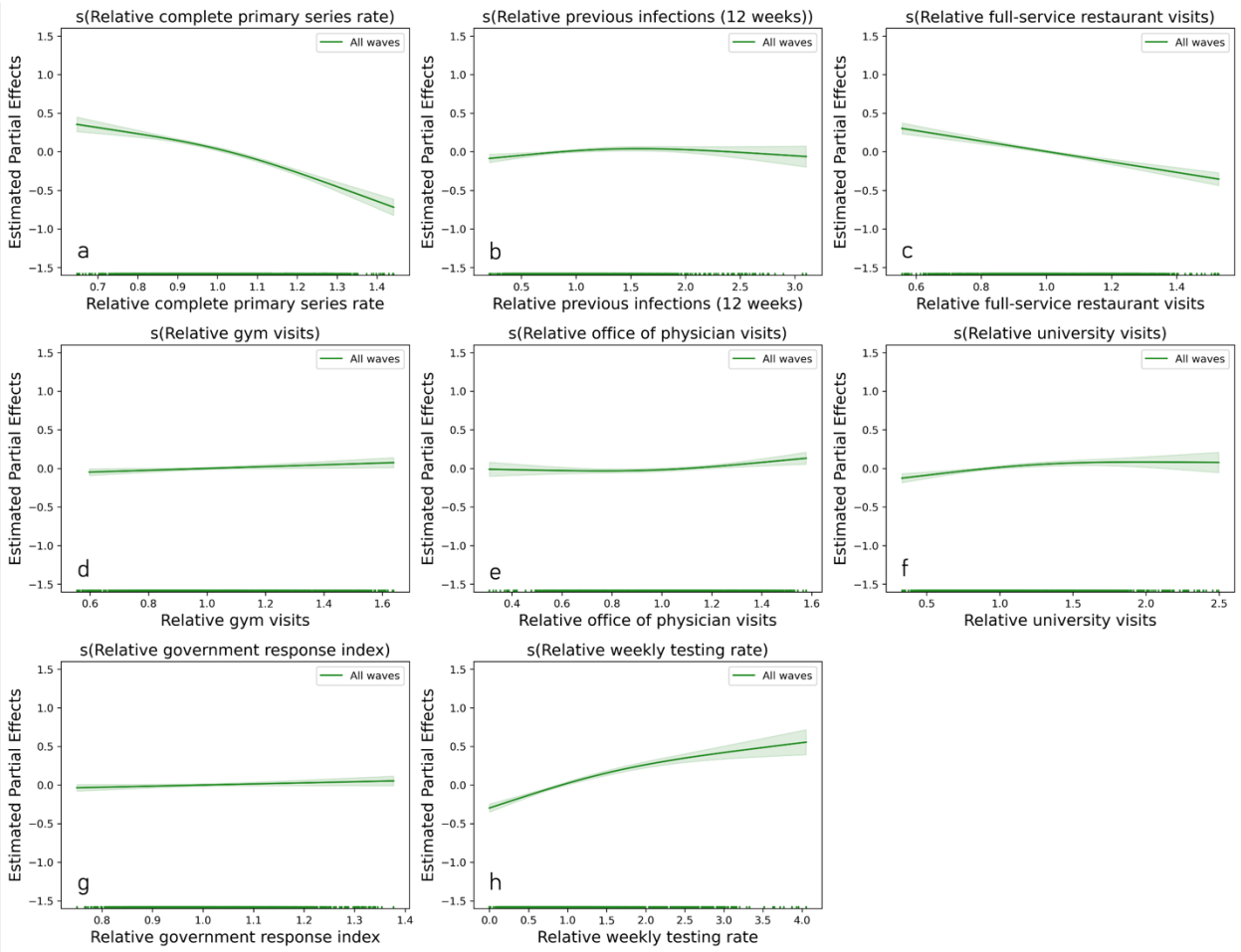
Appendix figure S25 illustrates the results of Model 4 (Int-Hosp-Booster), for just the Omicron wave with the additional inclusion of an interaction effect between the completed primary series rate and the relative booster rate. This interaction effect is presented as a 3-dimensional surface and a 2-dimension contour map in Figure 25.a and 25.b.



Appendix figure S25. Estimated partial effects of dependent variables for the Omicron wave from Int-Hosp-Booster. The figure 25.a shows a three-dimensional partial dependent plot for the interaction and figure 25.b shows the two-dimensional contour plot. The dots on the x-axis represent the distribution of the variables.

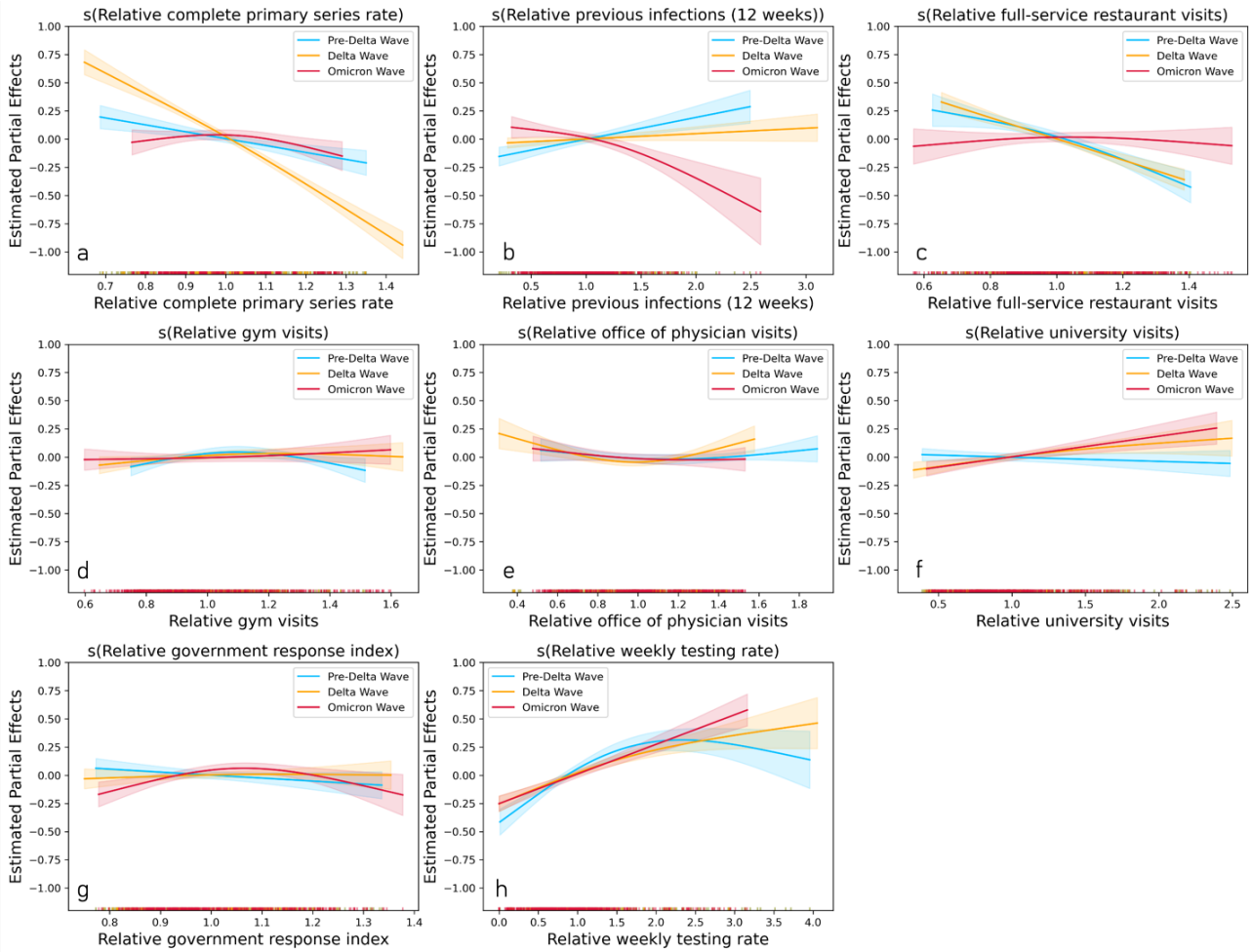
3.4 Results for Model 5 (Dyn-Inf)

Appendix figure S26 displays the estimated partial effects for infection rate obtained from the Model 5 (Dyn-Inf) for all state-week pairs. Each subfigure corresponds to the estimated impact of an independent variable on the outcome variable. The x-axis represents the values of the independent variable, while the y-axis represents its estimated effects. The dots plotted on the x-axis indicate the distribution of the independent variable. The magnitude of the y-axis is uniform across all subplots, allowing for a comparison of the relative impact of each variable on the outcome variable (hospitalization rate).



Appendix figure S26: Estimated partial effects of dependent variables by all state-week pairs from Dyn-Inf. The dots on the x-axis represent the distribution of the variables.

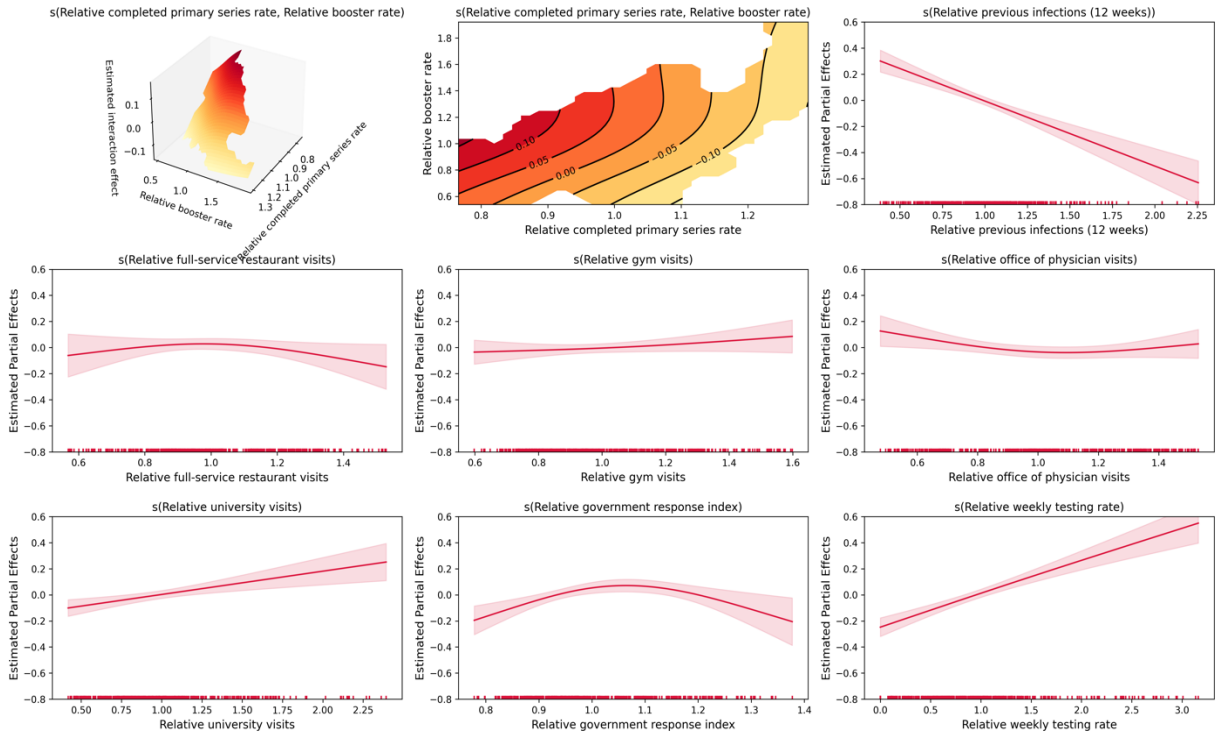
Appendix figure S27 illustrates the results for Model 5 (Dyn-Inf) for each variant, namely, the Pre-Delta (Blue), Delta (Orange), and Omicron waves (Red), respectively.



Appendix figure S27: Estimated partial effects of dependent variables by different variant waves from Dyn-Inf. The dots on the x-axis represent the distribution of the variables.

3.5 Results for Model 6 (Dyn-Inf-Booster)

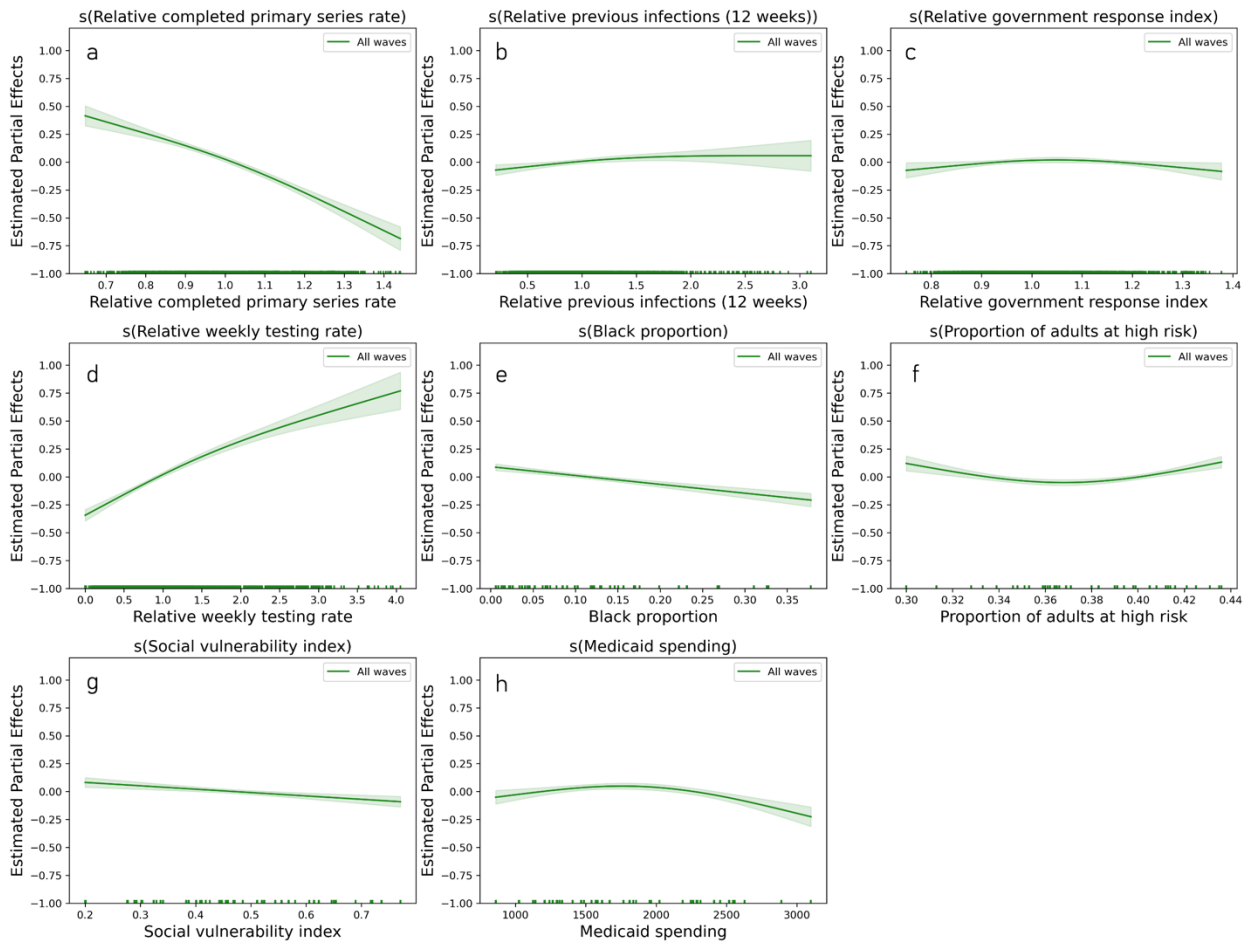
Appendix figure S28 illustrates the results of Model 6 (Dyn-Inf-Booster), for just the Omicron wave with the additional inclusion of an interaction effect between the completed primary series rate and the relative booster rate. This interaction effect is presented as a 3-dimensional surface and a 2-dimension contour map in Figure 28.a and 28.b.



Appendix figure S28. Estimated partial effects of dependent variables for the Omicron wave from Dyn-Inf-Booster. The figure 28.a shows a three-dimensional partial dependent plot for the interaction and figure 28.b shows the two-dimensional contour plot. The dots on the x-axis represent the distribution of the variables.

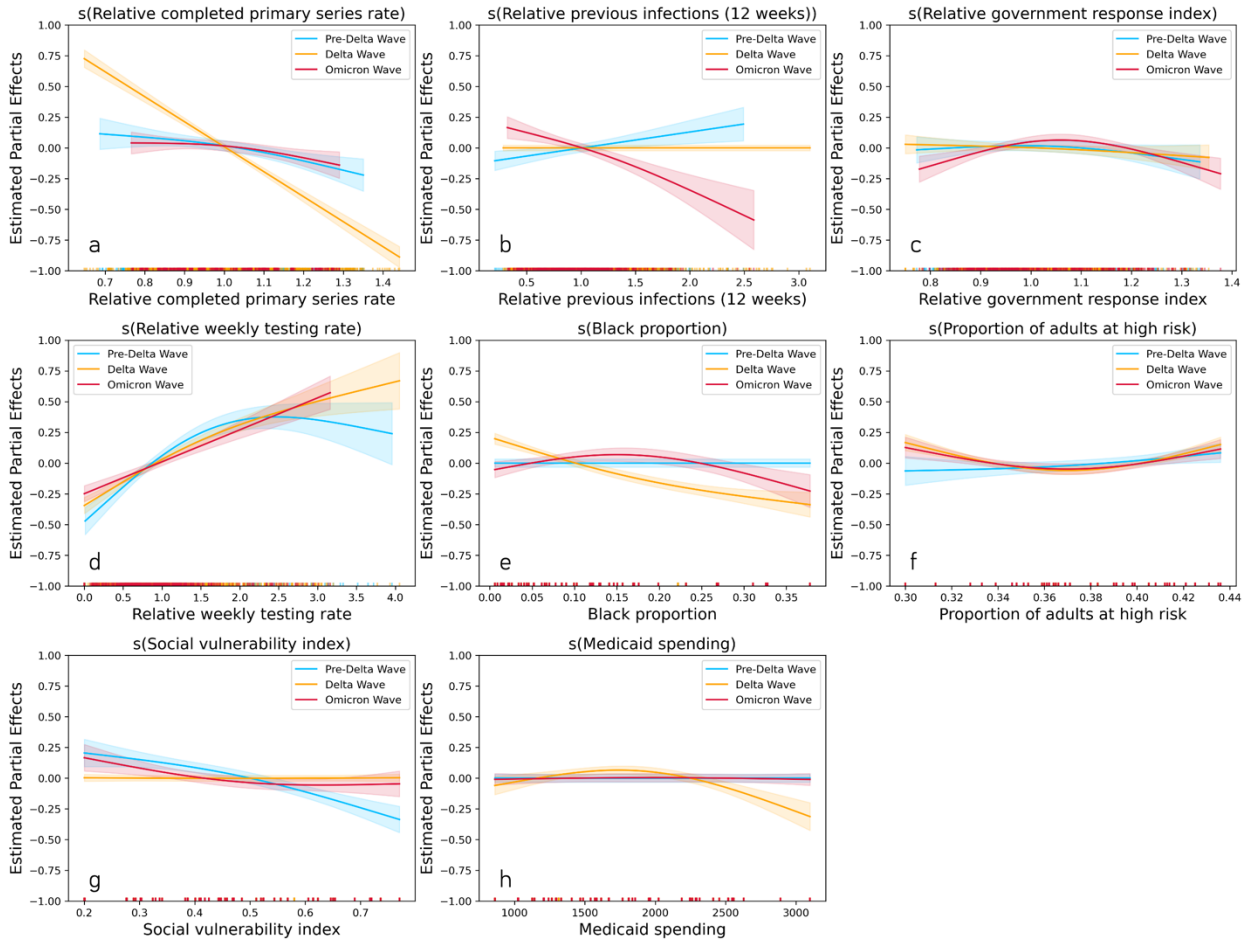
3.6 Results for Model 7 (Int-Inf)

Appendix figure S29 displays the estimated partial effects for infection rate obtained from the Model 7 (Int-Inf) for all state-week pairs. Each subfigure corresponds to the estimated impact of an independent variable on the outcome variable. The x-axis represents the values of the independent variable, while the y-axis represents its estimated effects. The dots plotted on the x-axis indicate the distribution of the independent variable. The magnitude of the y-axis is uniform across all subplots, allowing for a comparison of the relative impact of each variable on the outcome variable (infection rate).



Appendix figure S29. Estimated partial effects of dependent variables by all state-week pairs from Int-Inf. The dots on the x-axis represent the distribution of the variables.

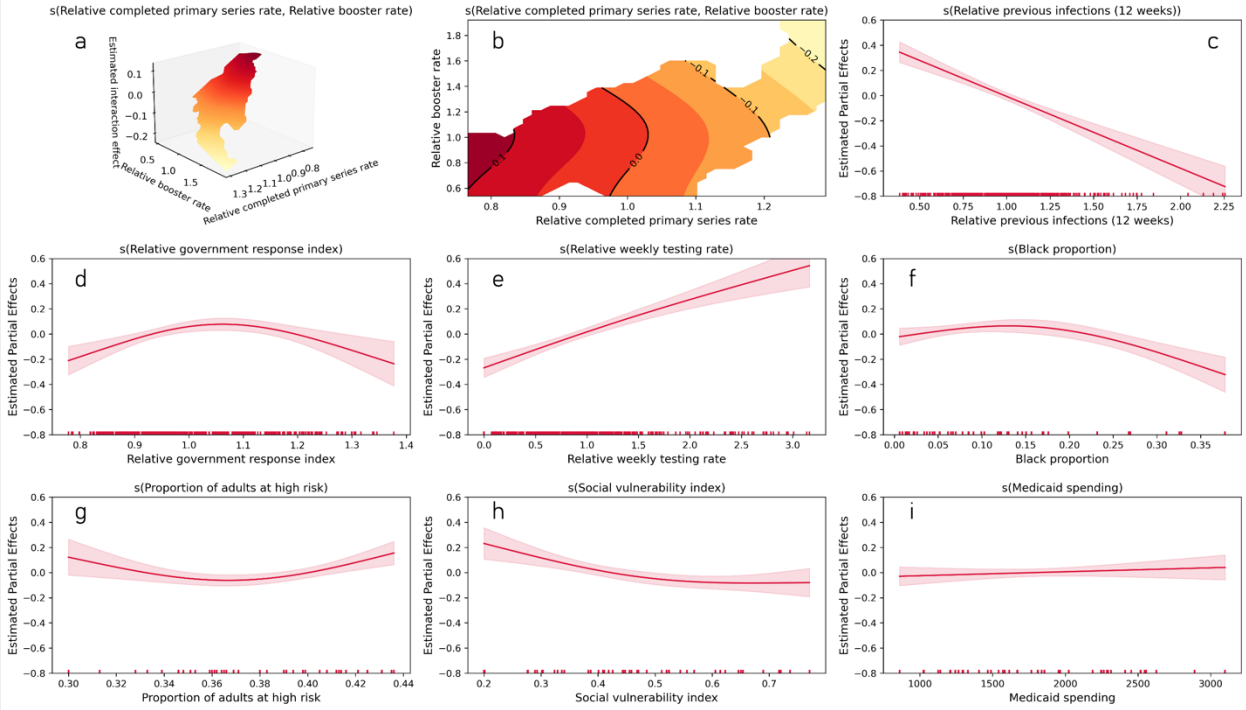
Appendix figure S30 illustrates the results for Model 7 (Int-Inf) for each variant, namely, the Pre-Delta (Blue), Delta (Orange), and Omicron waves (Red), respectively.



Appendix figure S30. Estimated partial effects of dependent variables by different variant waves from Int-Inf. The dots on the x-axis represent the distribution of the variables.

3.7 Results for Model 8 (Int-Inf-Booster)

Appendix figure S31 illustrates the results of Model 8 (Int-Inf-Booster), for just the Omicron wave with the additional inclusion of an interaction effect between the completed primary series rate and the relative booster rate. This interaction effect is presented as a 3-dimensional surface and a 2-dimension contour map in Figure 32.a and 31.b.



Appendix figure S31. Estimated partial effects of dependent variables for the Omicron wave from Int-Inf-Booster. The figure 31.a shows a three-dimensional partial dependent plot for the interaction and figure 31.b shows the two-dimensional contour plot. The dots on the x-axis represent the distribution of the variables.

Reference

1 Khare S, Gurry C, Freitas L, *et al.* GISAID's role in pandemic response. *China CDC Wkly* 2021; **3**: 1049.

2 Faes C, Abrams S, Van Beckhoven D, *et al.* Time between symptom onset, hospitalisation and recovery or death: statistical analysis of Belgian COVID-19 patients. *Int J Environ Res Public Health* 2020; **17**: 7560.

3 Du H, Dong E, Badr HS, Petrone ME, Grubaugh ND, Gardner LM. Incorporating variant frequencies data into short-term forecasting for COVID-19 cases and deaths in the USA: a deep learning approach. *eBioMedicine* 2023; **89**: 104482.

4 Places Data Curated for Accurate Geospatial Analytics | SafeGraph. https://www.safegraph.com (accessed Jan 13, 2023).

5 NAICS & SIC Identification Tools. NAICS Assoc. https://www.naics.com/search/ (accessed April 6, 2023).

6 Altarawneh HN, Chemaitelly H, Ayoub HH, *et al.* Effects of Previous Infection and Vaccination on Symptomatic Omicron Infections. *N Engl J Med* 2022; **387**: 21–34.

7 Goldberg Y, Mandel M, Bar-On YM, *et al.* Protection and Waning of Natural and Hybrid Immunity to SARS-CoV-2. *N Engl J Med* 2022; **386**: 2201–12.

8 Bobrovitz N, Ware H, Ma X, *et al.* Protective effectiveness of previous SARS-CoV-2 infection and hybrid immunity against the omicron variant and severe disease: a systematic review and meta-regression. *Lancet Infect Dis* 2023; **23**: 556–67.

9 Abrahim SA, Tessema M, Defar A, *et al.* Time to recovery and its predictors among adults hospitalized with COVID-19: A prospective cohort study in Ethiopia. *PLOS ONE* 2020; **15**: e0244269.

10 Haug N, Geyrhofer L, Londei A, *et al.* Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav* 2020; **4**: 1303–12.

11 Hale T, Webster S, Petherick A, Phillips T, Kira B. Oxford COVID-19 government response tracker (OxCGRT). *Last Updat* 2020; **8**: 30.

12 Bollyky TJ, Castro E, Aravkin AY, *et al.* Assessing COVID-19 pandemic policies and behaviours and their economic and educational trade-offs across US states from Jan 1, 2020, to July 31, 2022: an observational analysis. *The Lancet* 2023; : S0140673623004610.

13 Bureau UC. State Population by Characteristics: 2010-2019. Census.gov. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-detail.html (accessed April 9, 2023).

14 Total Medicaid Spending. KFF. 2022; published online Aug 22. https://www.kff.org/medicaid/state-indicator/total-medicaid-spending/ (accessed Feb 13, 2023).

15 Health Care Expenditures by State of Residence (in millions). KFF. 2022; published online Sept 13. https://www.kff.org/other/state-indicator/health-care-expenditures-by-state-of-residence-in-millions/ (accessed Feb 13, 2023).

16 Poverty Rate by Race/Ethnicity. KFF. 2022; published online Oct 28. https://www.kff.org/other/state-indicator/poverty-rate-by-raceethnicity/ (accessed Feb 13, 2023).

17 CDC/ATSDR SVI Data and Documentation Download | Place and Health | ATSDR. 2022; published online Dec 22. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html (accessed Feb 13, 2023).

18 Global Burden of Disease (GBD). Inst. Health Metr. Eval. 2014; published online March 29. https://www.healthdata.org/gbd (accessed April 9, 2023).

19 Harvard Dataverse. Harv. Libr. https://library.harvard.edu/services-tools/harvard-dataverse (accessed April 9, 2023).

20 Adults at Higher Risk of Serious Illness if Infected with Coronavirus. KFF. 2020; published online May 13. https://www.kff.org/other/state-indicator/adults-at-higher-risk-of-serious-illness-if-infected-with-coronavirus/ (accessed Feb 13, 2023).