Supplementary Appendix

Supplement to: Lin D-Y, Gu Y, Wheeler B, et al. Effectiveness of Covid-19 vaccines over a 9-month period in North Carolina. N Engl J Med 2022;386:933-41. DOI: 10.1056/NEJMoa2117128

This appendix has been provided by the authors to give readers additional information about the work.

Table of Contents

SUPPLEMENTARY METHODS

Data Linkage

Data from the state's Covid-19 vaccination management system include a person-level view of all doses received from state-affiliated providers. Data from federal pharmacy providers are provided at the dose level. Individuals are free to participate in both vaccination programs and may be represented in both views. These data are refreshed weekly.

Records from the state's lab and case surveillance system are matched to vaccination data weekly. To capture post vaccination cases, positive lab results collected after December 11, 2020 -- the date of first vaccination in North Carolina – are selected for matching to vaccination data. Lab records contain name elements, date of birth, and geographic information for linkage purposes, demographic covariates, and Covid-19 outcomes including hospitalization and death.

Matching Model

Data from the vaccination registry are transformed prior to matching. State-administered records are already normalized at the person level. A person-normalized view of the federal pharmacy data is generated by probabilistically deduplicating dose-level data on name elements, date of birth, and geography prior to matching using Link Plus 3.0[1] (CDC).

Positive lab results are then matched to person-level records from the state and federal pharmacy vaccine registry across two linkage models. Records in each are probabilistically linked on name elements, dates of birth, and geographic territory in Link Plus, generating a scored list of potential record matches. Results of matching models above a minimum score threshold are selected for manual review.

After manual validation, matching records are added to a cumulative linkage table, capturing record IDs from each system and the date of the linkage. The matching process is continuous; IDs not previously matched are selected in subsequent rounds of linkage. Cumulative linkage tables are used to join the lab and vaccination data to summarize vaccination outcomes.

Statistical Methods

Let S denote the time when an individual is injected with the first dose of a vaccine, and T denote the time when the individual experiences an event of interest (e.g., Covid-19, hospitalization, death); both times are measured in days from the start of the vaccine rollout, i.e., December 11, 2020. We consider three vaccine products: BNT162b2, mRNA-1273, and Ad26.COV2.S, and we assume that each individual can receive only one product. For an mRNA vaccine, we treat those who have been vaccinated with one dose only and those who have been fully vaccinated as two different groups, and we estimate the vaccine effectiveness of the one-dose and two-dose regimens separately. Thus, there are a total of five vaccine types, i.e., BNT162b2 one dose, BNT162b2 two doses, mRNA-1273 one dose, mRNA-1273 two doses, and Ad26.COV2.S single dose. We number them from 1 to 5 and use V to denote the vaccine type the individual has received by the end of the study (i.e., the date when the surveillance data are locked). In addition, let X denote baseline risk factors (i.e., age group, sex, race, ethnicity, geographic region, and county-level vaccination rate).

We specify that the hazard function of T is related to S, V, and X through the Cox regression model

$$\lambda(t|\, S, V, X) = \lambda_0(t) \exp\left\{\beta^{\mathrm{T}} X + \sum_{k=1}^{5} \eta_k(t - S) A_k(t)\right\}, \tag{1}$$

where $\lambda_0(\cdot)$ is an arbitrary baseline hazard function, β is a set of regression parameters representing the effects of baseline risk factors, $\eta_k(\cdot)$ is a function characterizing the time-varying effect of the kth vaccine type, $A_k(t) = I(S < t, V = k)$, and $I(\cdot)$ is the indicator function. This is a generalization of a Cox model previously used for a single vaccine.[2] All baseline risk factors are categorical variables, and the categories of each variable are compared to its first category. Under this formulation, the baseline hazard function varies over the calendar time, and the effect of a vaccine on the risk of disease depends on the time elapsed since the first dose.

Let τ denote the end of the study. We observe $\tilde{T} = \min(T, \tau)$ and $\Delta = I(T \leq \tau)$. The data consist of $\left(\tilde{T}_i, \Delta_i, S_i, V_i, X_i\right)$ $(i = 1, \dots, n)$, where n is the total number of individuals in the population. We let the five log hazard ratios, $\eta_1(\cdot), \dots, \eta_5(\cdot)$, be piecewise linear functions of time since the first dose, with potentially different numbers of pieces and different change points. That is, for $k = 1, \dots, 5$,

$$\eta_k(t) = \sum_{l=0}^{m_k} \gamma_{kl}\, B_{kl}(t) = \gamma_{k0} t + \gamma_{k1}(t - t_{k1})_+ + \cdots + \gamma_{km_k}\left(t - t_{km_k}\right)_+,$$

where $t_+ = t$ if $t > 0$ and 0 otherwise, $t_{k1}, \dots, t_{km_k}$ are the m_k change points, and $\gamma_{k0}, \dots, \gamma_{km_k}$ are the unknown parameters pertaining to the slope of each piece. In addition, for a two-dose regimen, we assume that the time between the first and second doses is roughly the same for every individual and that the difference in the hazard ratio between those who take one dose only and those who take two doses can only occur after the injection of the second dose. Thus, we impose the constraints that $\gamma_{10} = \gamma_{20}$ and $\gamma_{30} = \gamma_{40}$. For $k = 1, \dots, 5$, write $\gamma_k = \left(\gamma_{k0}, \dots, \gamma_{km_k}\right)^{\mathrm{T}}$, and $Z_{ik}(t) = \Big(B_{k0}(t - S_i)A_{ik}(t), \dots, B_{km_k}(t - S_i)A_{ik}(t)\Big)^{\mathrm{T}}$, where $A_{ik}(t) = I(S_i < t, V_i = k)$. We further write $\gamma = \left(\gamma_1^{\mathrm{T}}, \dots, \gamma_5^{\mathrm{T}}\right)^{\mathrm{T}}$ and $Z_i(t) = (Z_{i1}(t)^{\mathrm{T}}, \dots, Z_{i5}(t)^{\mathrm{T}})^{\mathrm{T}}$. Then the partial likelihood for β and γ takes the form

$$L(\beta, \gamma) = \prod_{i=1}^{n} \left\{\frac{e^{\beta^{\mathrm{T}} X_i + \gamma^{\mathrm{T}} Z_i(\tilde{T}_i)}}{S^{(0)}\left(\beta, \gamma; \tilde{T}_i\right)}\right\}^{\Delta_i},$$

where $S^{(0)}(\beta, \gamma; t) = \sum_{j=1}^{n} I\left(\tilde{T}_j \geq t\right) e^{\beta^{\mathrm{T}} X_j + \gamma^{\mathrm{T}} Z_j(t)}$.

The analysis dataset contains the information of all the individuals who have been vaccinated or have developed Covid-19 by the end of the study. For each stratum defined by the combination of the baseline risk factors, we can determine the number of individuals who have neither been vaccinated nor developed Covid-19 by the end of the study by subtracting the number of individuals in the analysis dataset from the total number of individuals in the census data. Suppose that there are $\tilde{n}$ individuals in the analysis dataset and that there are n_g individuals in the gth stratum $(g = 1, \dots, G)$ out of those who are not in the analysis dataset. Note that the individuals in the same stratum have the same value of X. Note also that, for the individuals not in the analysis dataset, $\tilde{T} = \tau$, $\Delta = 0$, and

$S > \tau$. Thus, $S^{(0)}(\beta, \gamma; t) = \sum_{j=1}^{\tilde{n}} I(\tilde{T}_j \geq t) e^{\beta^{\mathrm{T}} X_j + \gamma^{\mathrm{T}} Z_j(t)} + \sum_{g=1}^{G} n_g e^{\beta^{\mathrm{T}} X_g}$, where X_g is the value of X for the gth stratum. The same strategy is used for hospitalization and death.

We maximize $L(\beta, \gamma)$ through the Newton-Raphson algorithm. The resulting maximum partial likelihood estimator $(\hat{\beta}, \hat{\gamma})$ is approximately normal with mean (β, γ) and covariance matrix $\mathbf{I^{-1}}(\hat{\beta}, \hat{\gamma})$, where $\mathbf{I}(\beta, \gamma)$ is the negative second-derivative matrix of $\log L(\beta, \gamma)$. Given $\hat{\gamma}_k$, we can estimate the vaccine effectiveness in reducing the hazard rate for the kth vaccine type by

$$\widehat{VE}_k(t) = 1 - e^{\hat{\eta}_k(t)} = 1 - \exp\left\{\sum_{l=0}^{m_k} \hat{\gamma}_{kl} B_{kl}(t)\right\}$$

and also construct the 95% confidence interval.

For each of the three clinical outcomes of interest (i.e., Covid-19, hospitalization, and death), we fit model (1) for the whole population, adjusting for the effects of age, race/ethnicity, sex, geographic region, and county-level vaccination rate. In addition, we perform subgroup analysis for one demographic variable at a time, adjusting for all the other variables. In the piecewise linear approximation of the log hazard ratio $\eta_k(\cdot)$, $k = 1, \ldots, 5$, we place a change point at every month, until the month prior to the last observed event time among all individuals of the kth vaccine type. In addition, to allow the curve of vaccine effectiveness to depend on the date of vaccination, we let the log hazard ratio $\eta_k(\cdot)$ take a different value for individuals who received the first dose during a different calendar period.

References

1. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 2020. (https://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm.)
2. Lin DY, Gu Y, Zeng D, Janes HE, Gilbert PB. Evaluating vaccine efficacy against SARS-CoV-2 infection. Clin Infect Dis 2021; ciab630, https://doi.org/10.1093/cid/ciab630.

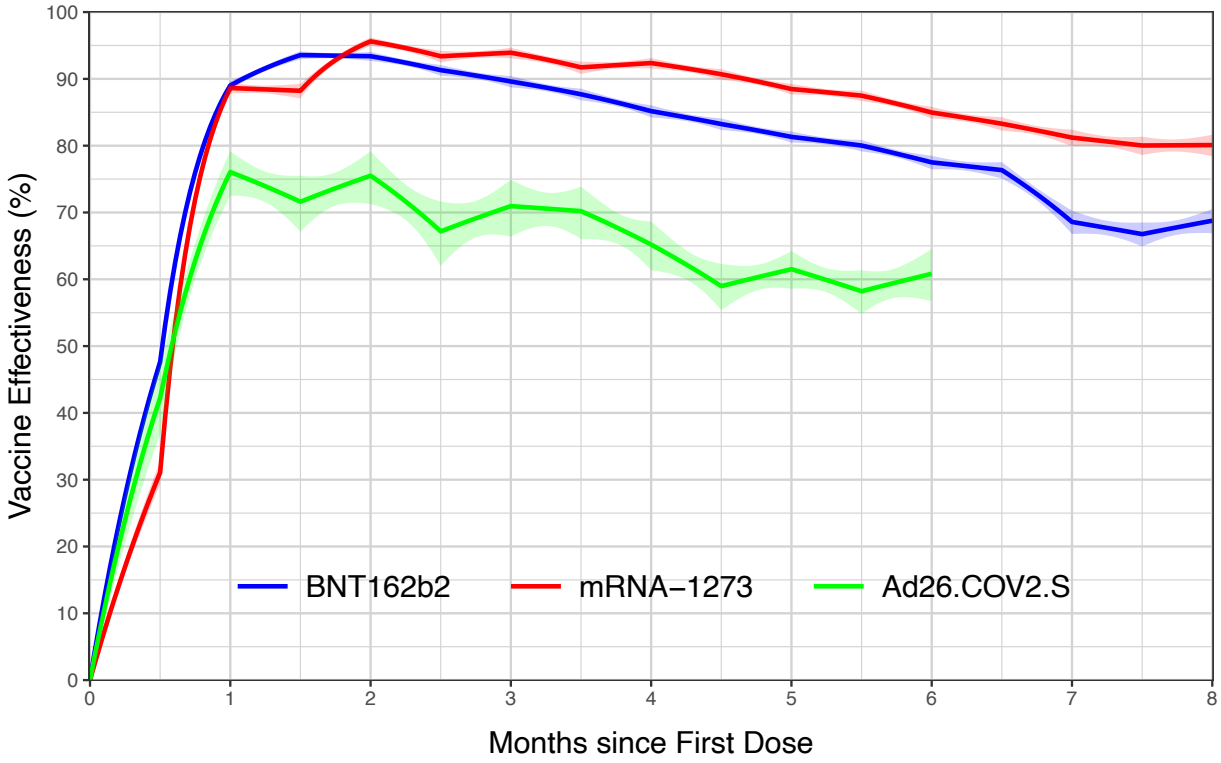


Figure S1. Vaccine effectiveness in reducing the risk of Covid-19 with change points placed at every 2 weeks. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands.

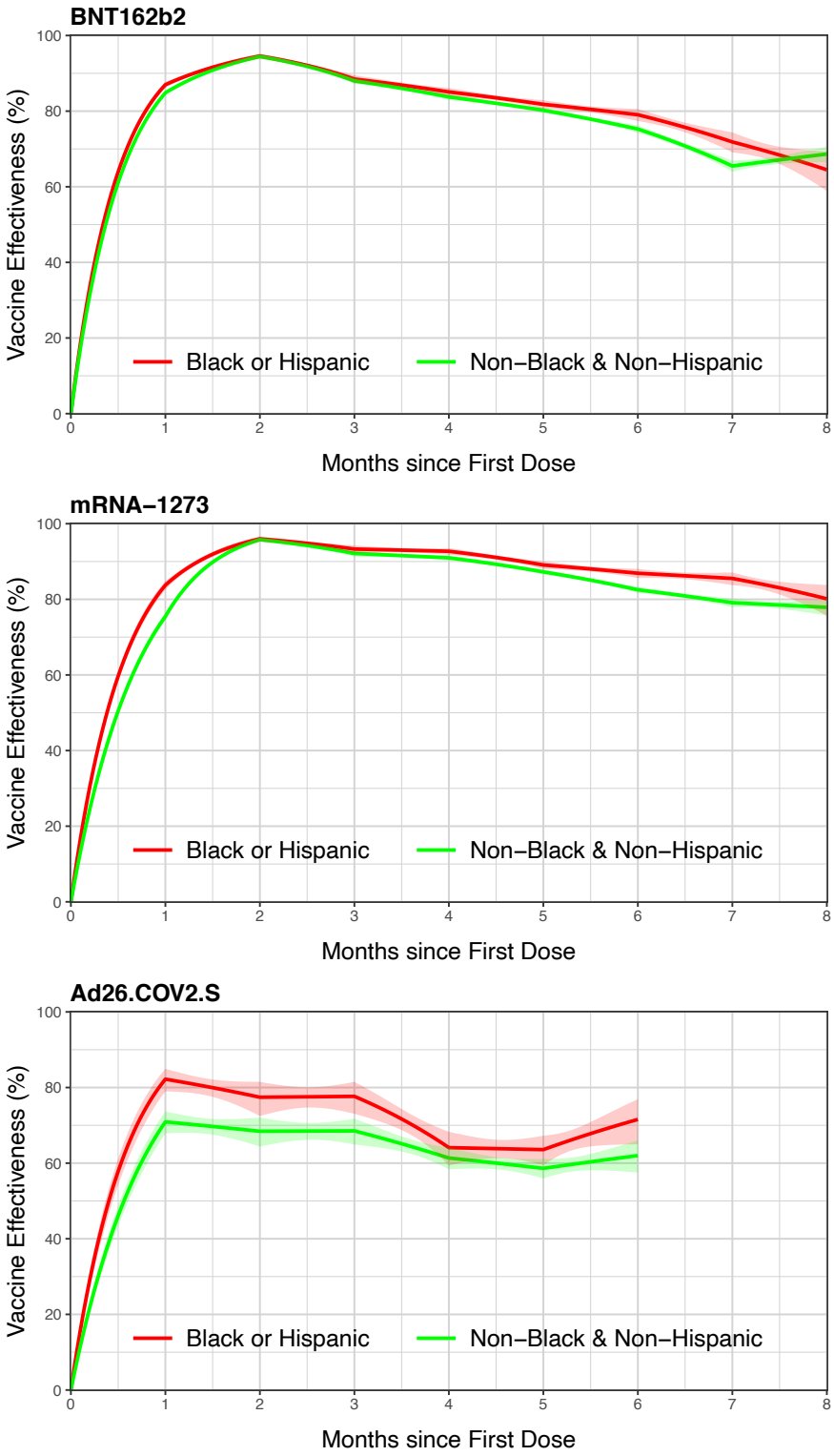


Figure S2. Vaccine effectiveness in reducing the risk of Covid-19 by race/ethnicity. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands.

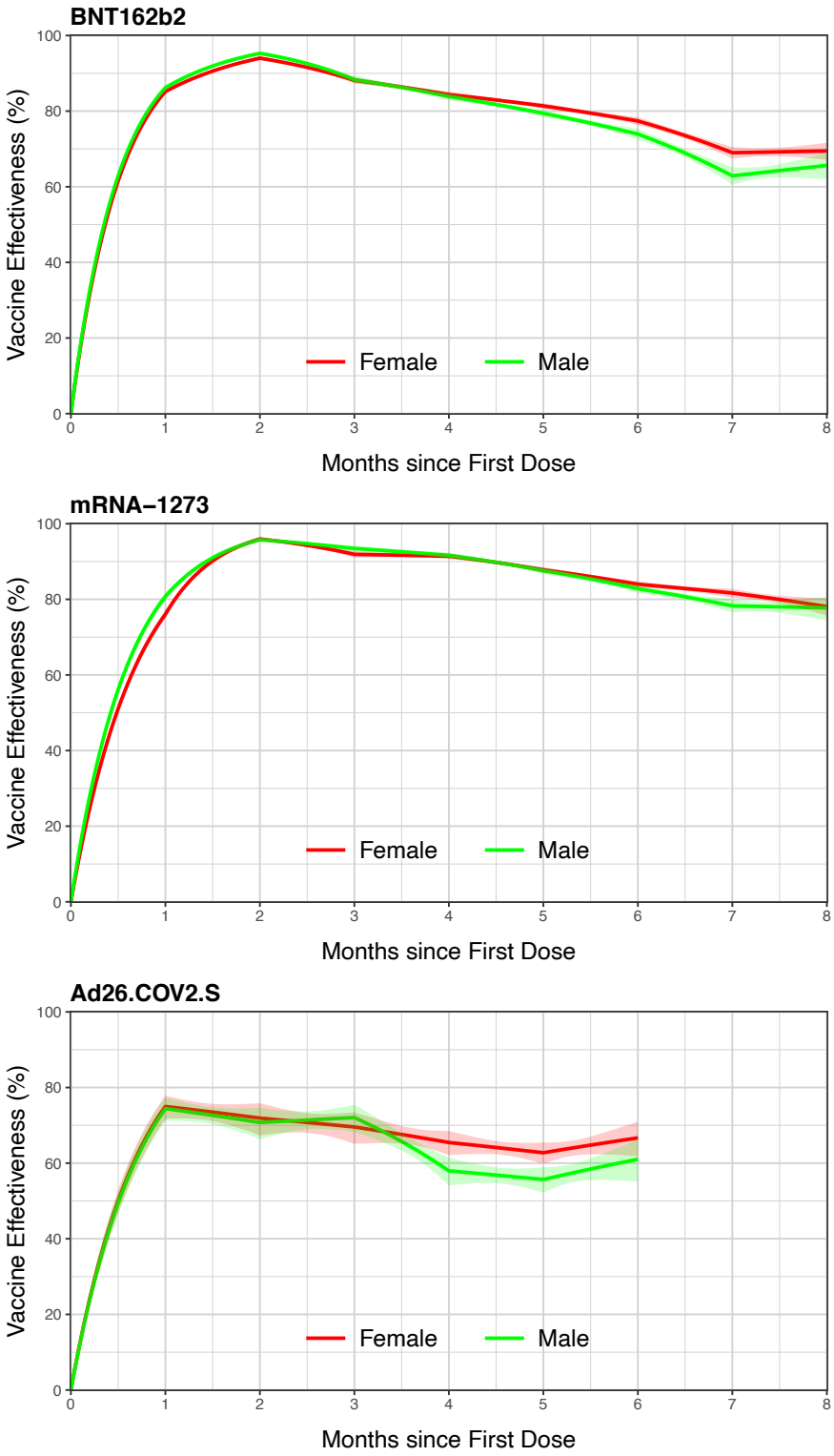


Figure S3. Vaccine effectiveness in reducing the risk of Covid-19 by sex. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands.

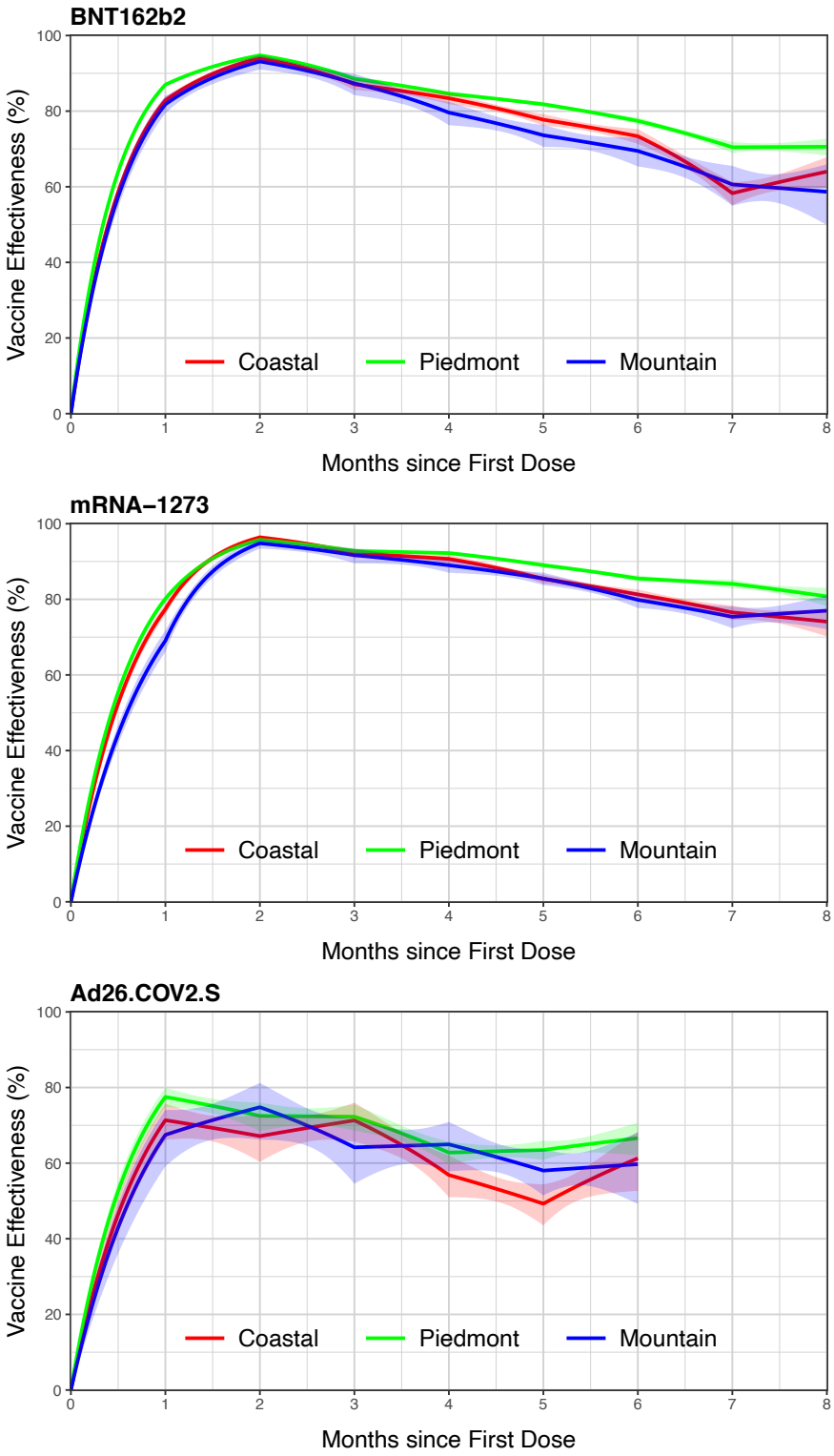


Figure S4. Vaccine effectiveness in reducing the risk of Covid-19 by geographic region. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands.

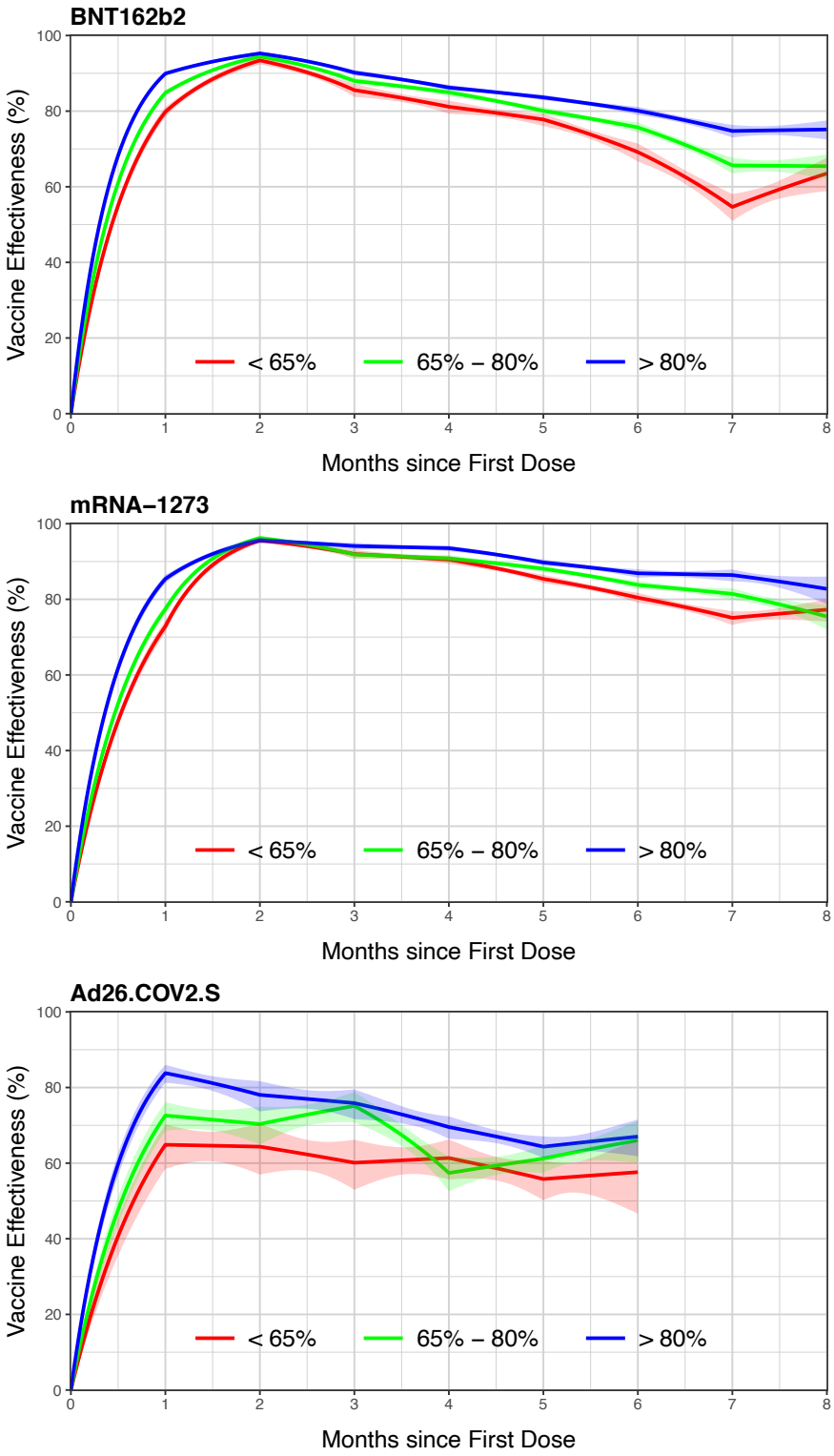


Figure S5. Vaccine effectiveness in reducing the risk of Covid-19 by county-level vaccination rate. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands.

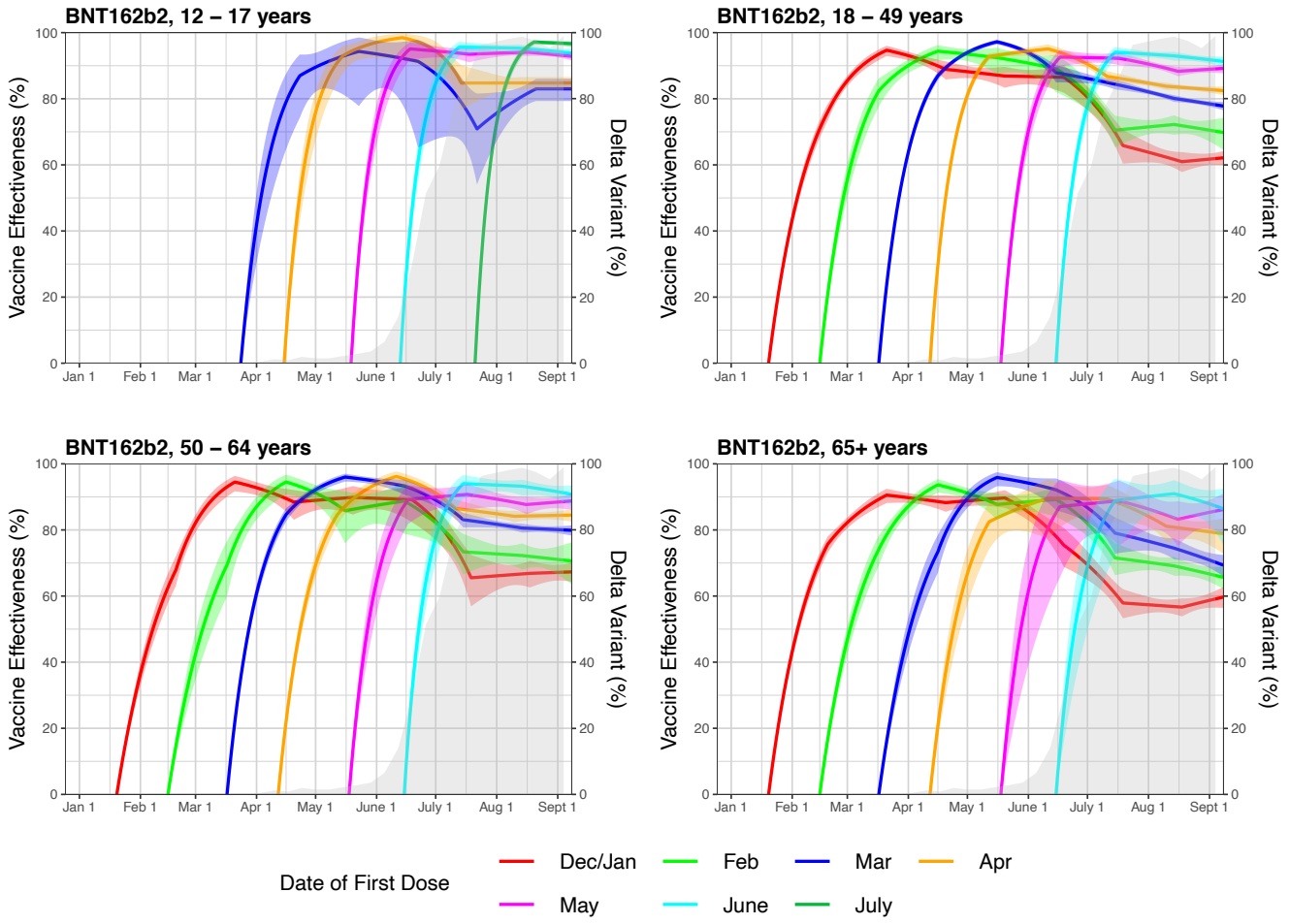


Figure S6. Vaccine effectiveness in reducing the risk of Covid-19 by age group and date of first dose. Estimates of VE effectiveness are shown by solid curves, 95% confidence intervals are shown by shaded bands, and the prevalence of the Delta variant is shaded.

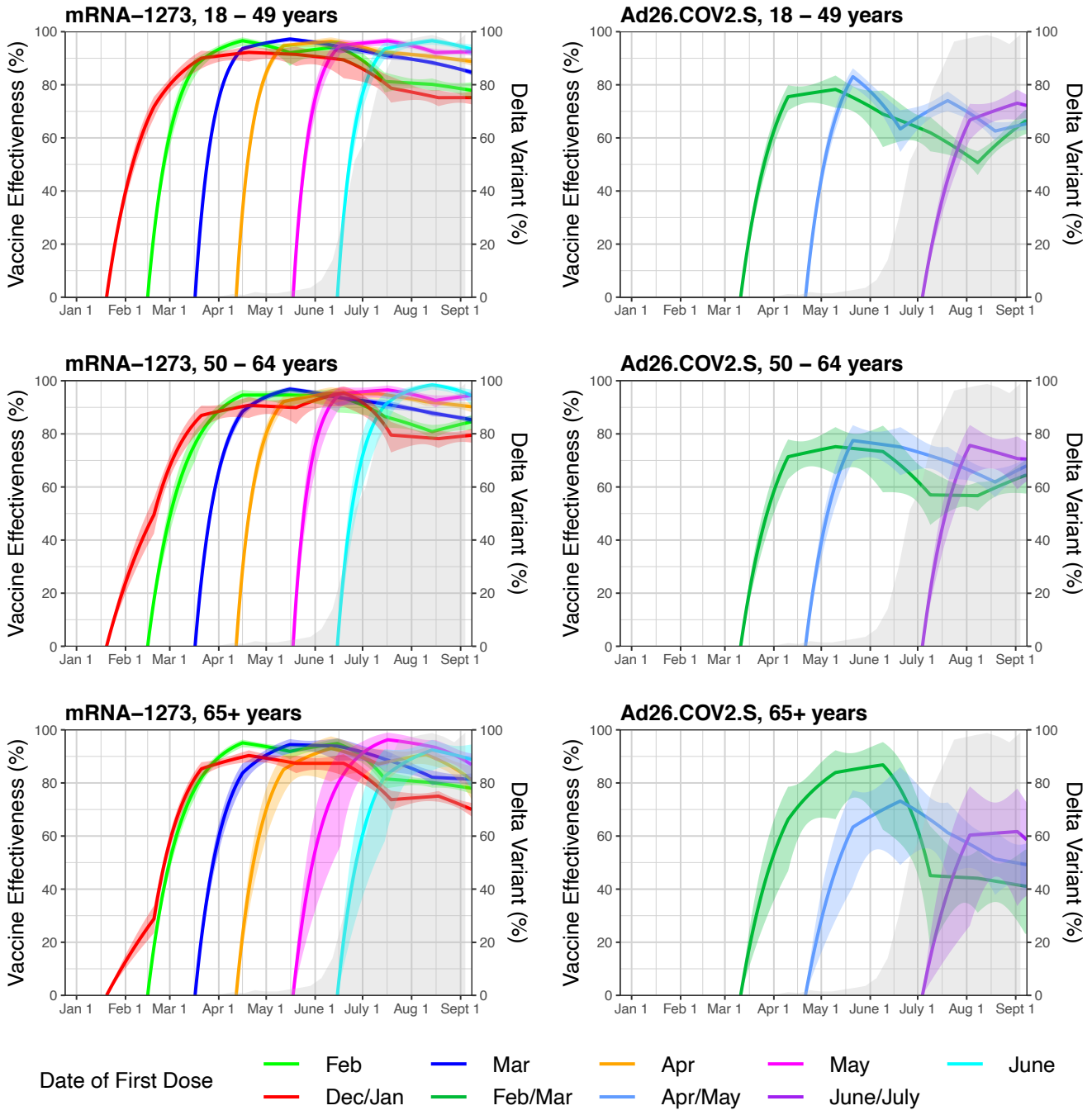


Figure S6 (cont.)

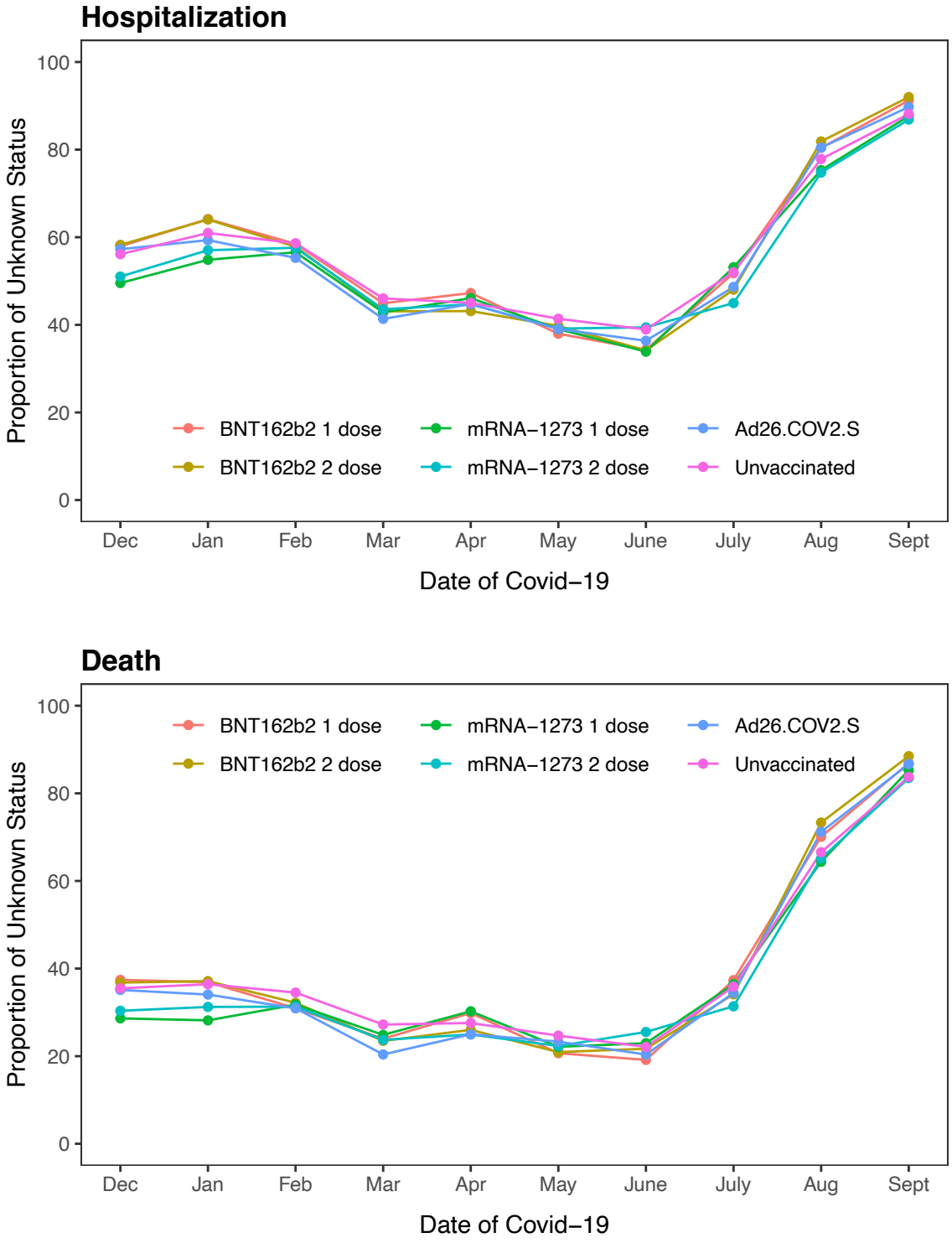


Figure S7. Proportion of unknown hospitalization status or unknown survival status according to date of Covid-19 and vaccine type.

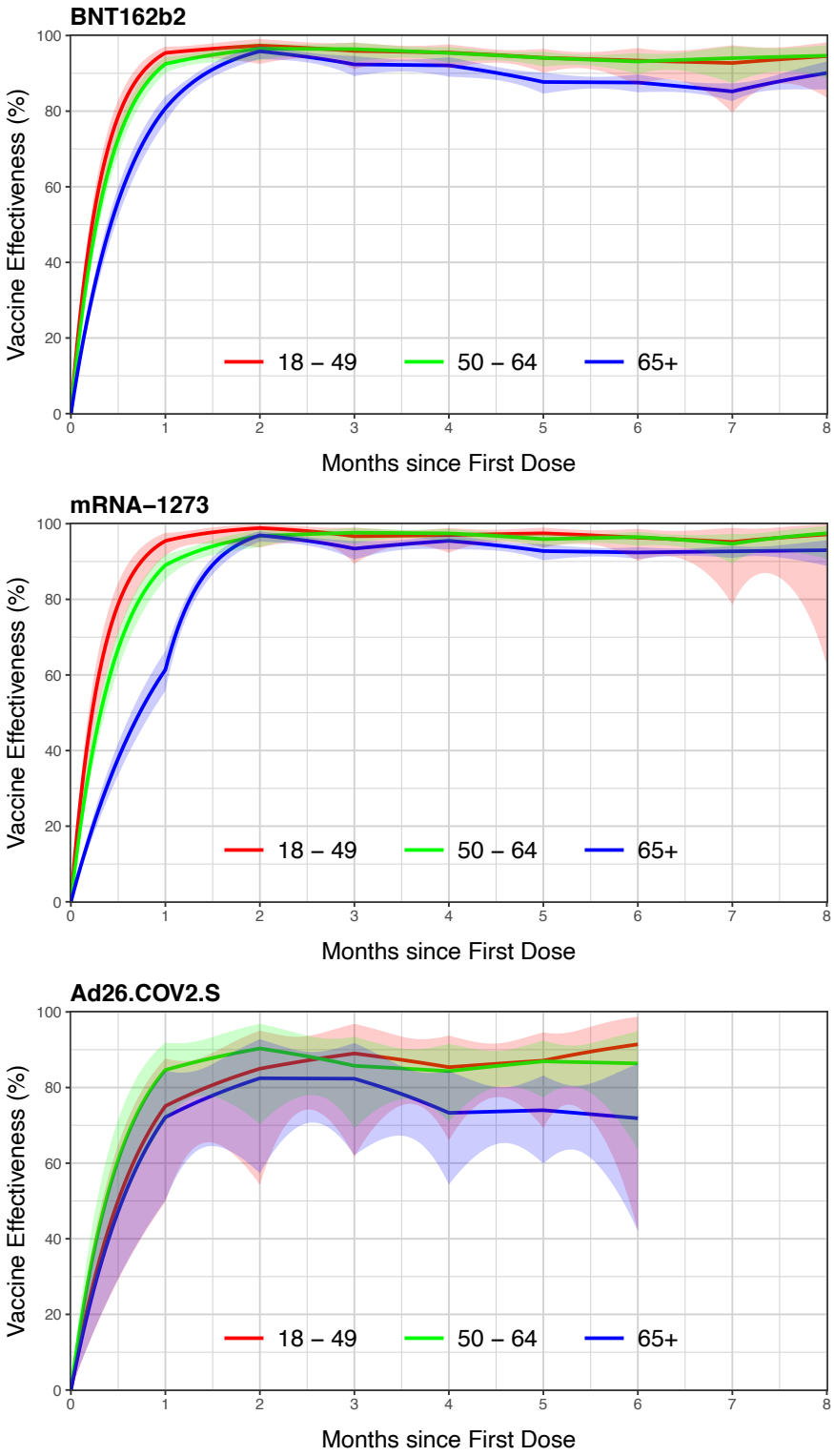


Figure S8. Vaccine effectiveness in reducing the risk of hospitalization by age group. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 50 imputations.

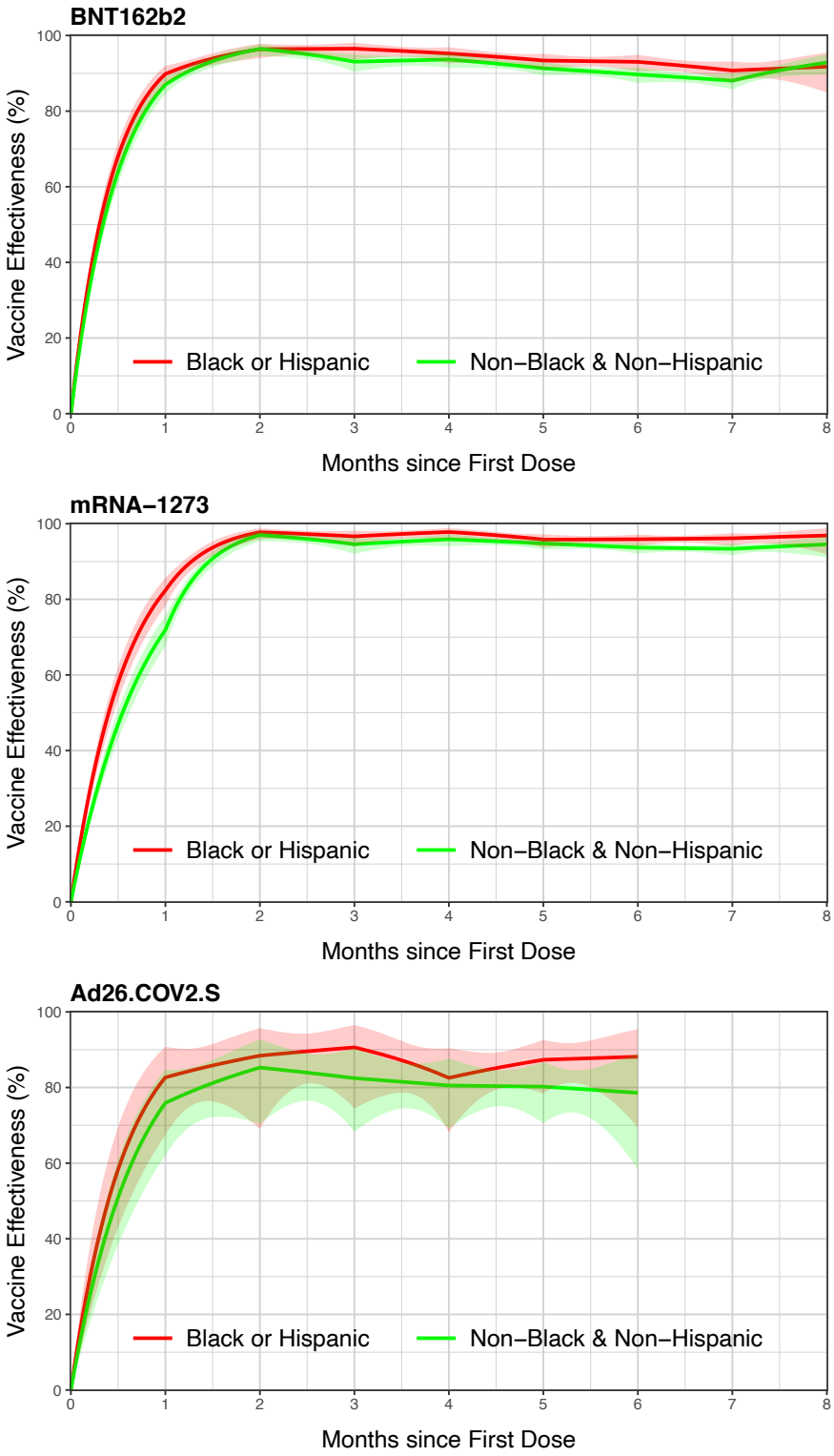


Figure S9. Vaccine effectiveness in reducing the risk of hospitalization by race/ethnicity. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 50 imputations.

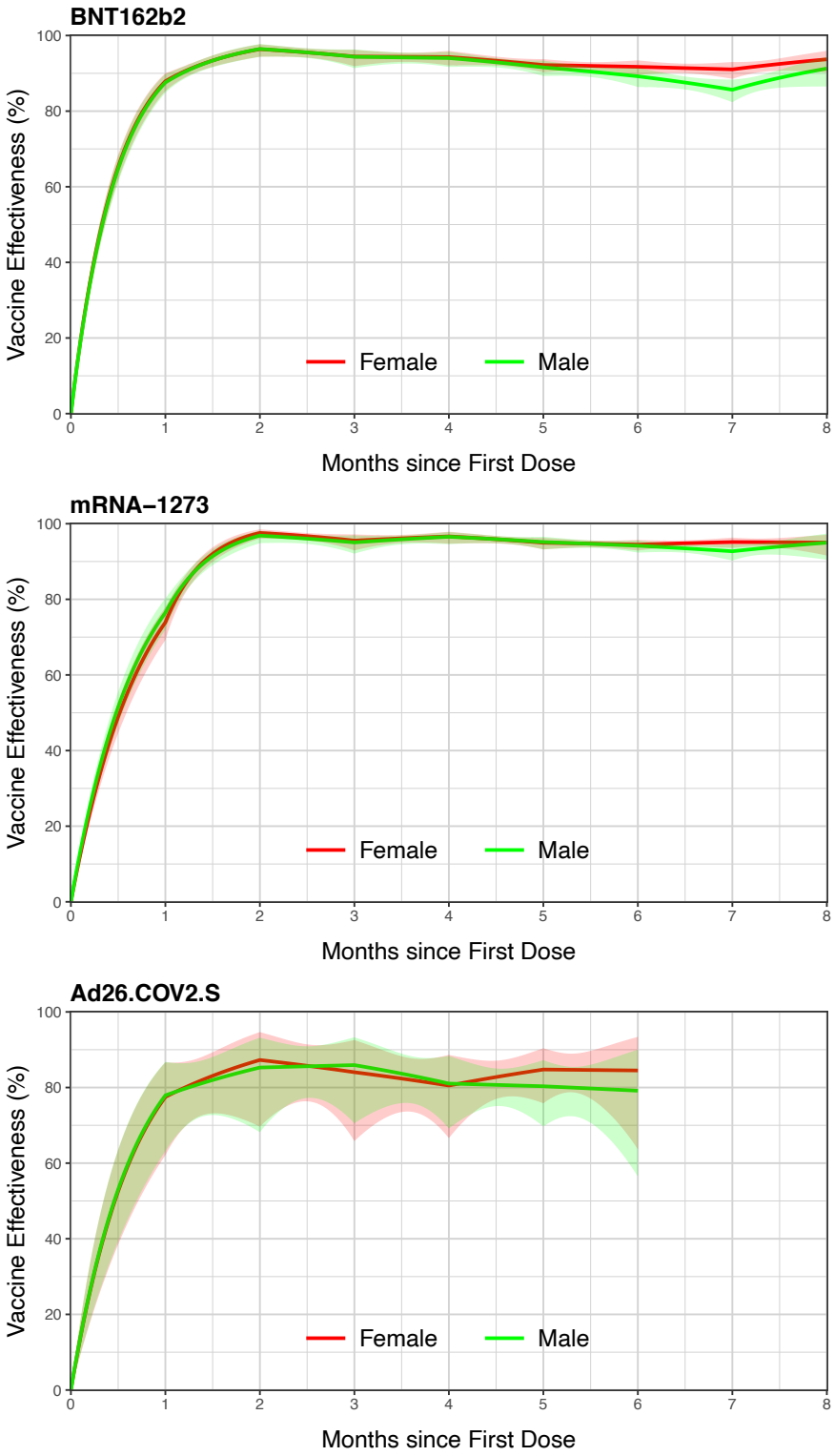


Figure S10. Vaccine effectiveness in reducing the risk of hospitalization by sex. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 50 imputations.

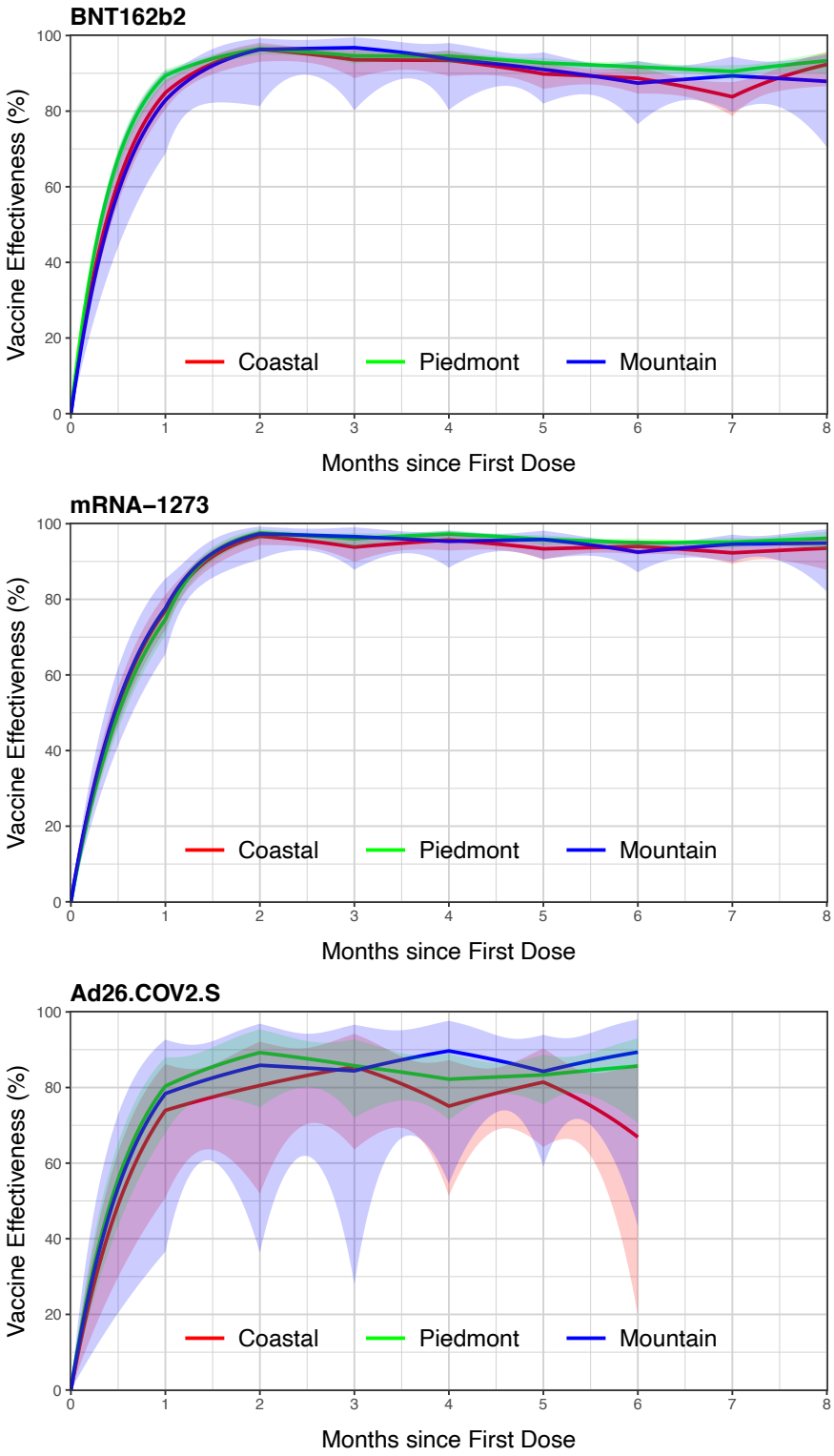


Figure S11. Vaccine effectiveness in reducing the risk of hospitalization by geographic region. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 50 imputations.

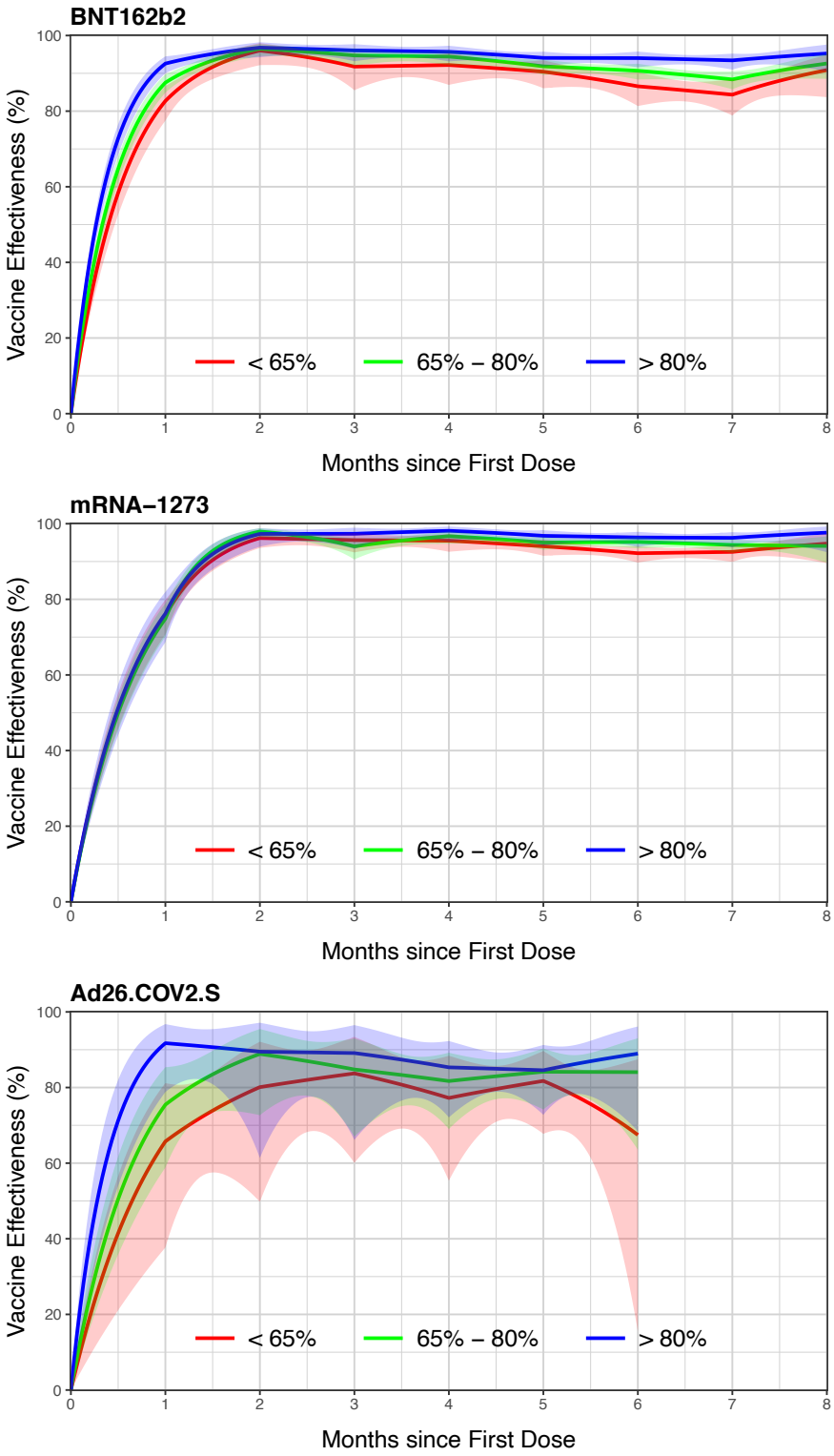


Figure S12. Vaccine effectiveness in reducing the risk of hospitalization by county-level vaccination rate. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 50 imputations.

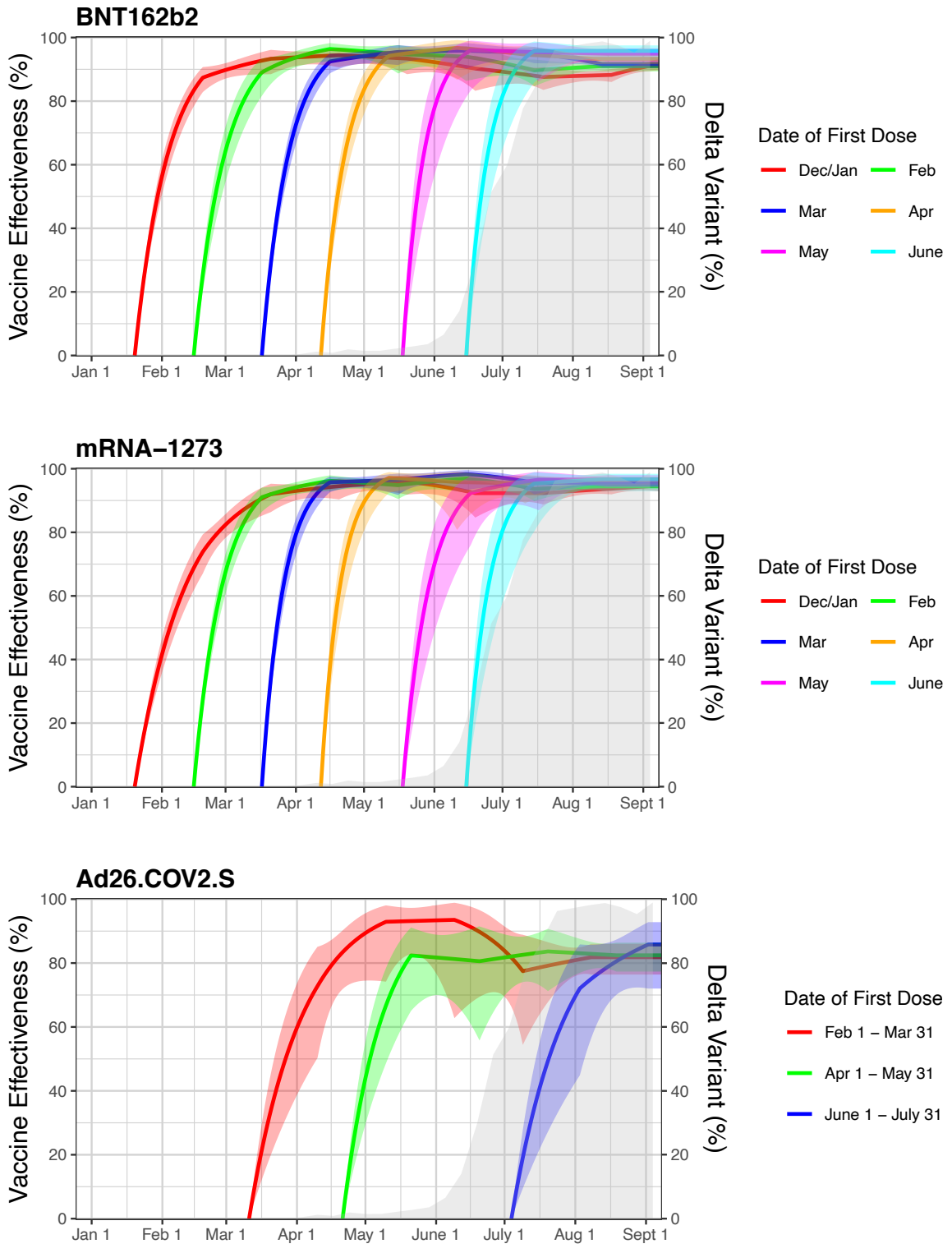


Figure S13. Vaccine effectiveness in reducing the risk of hospitalization by date of first dose. Estimates of VE effectiveness are shown by solid curves, 95% confidence intervals are shown by shaded bands, and the prevalence of the Delta variant is shaded. Results are based on 50 imputations.

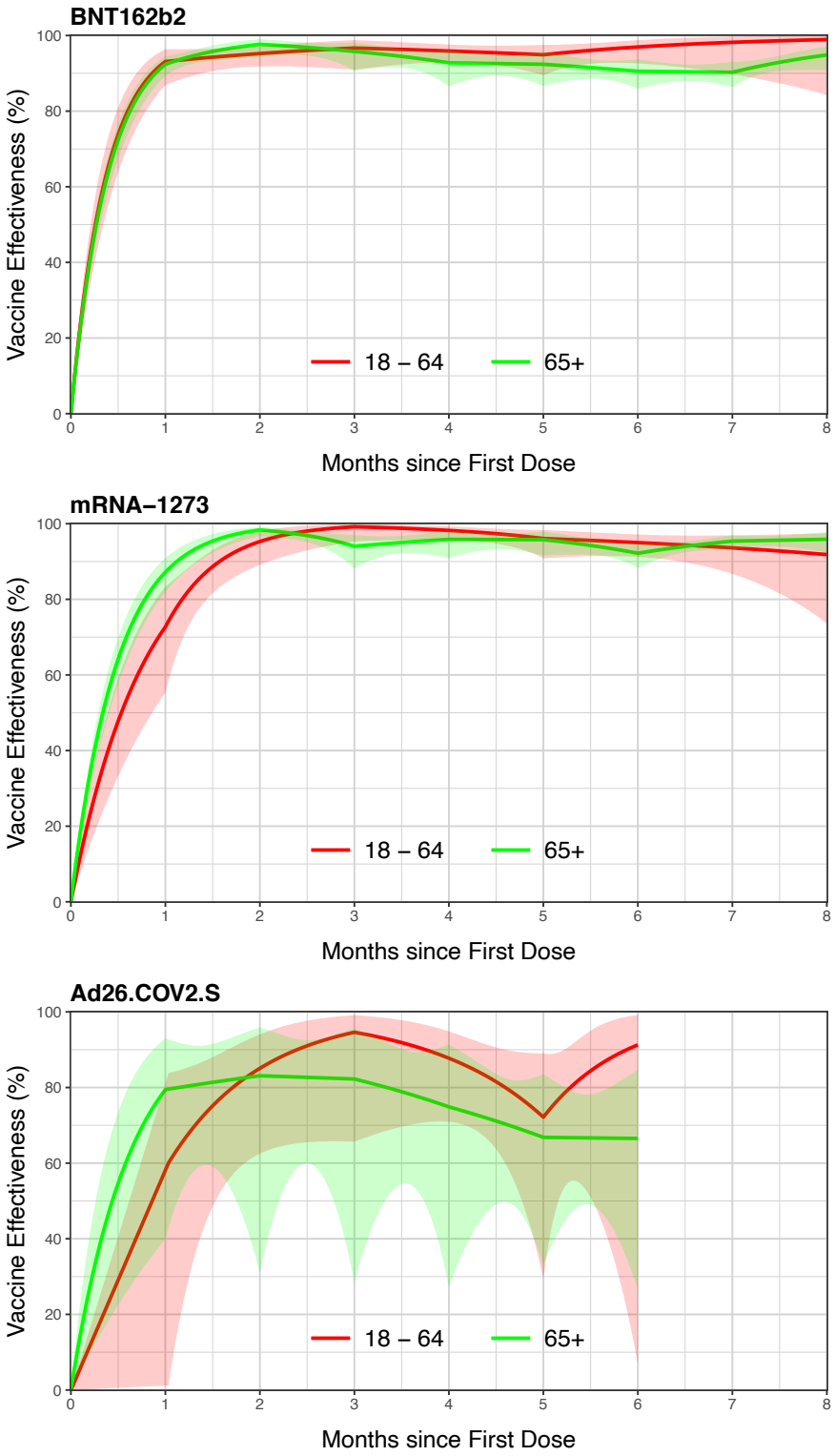
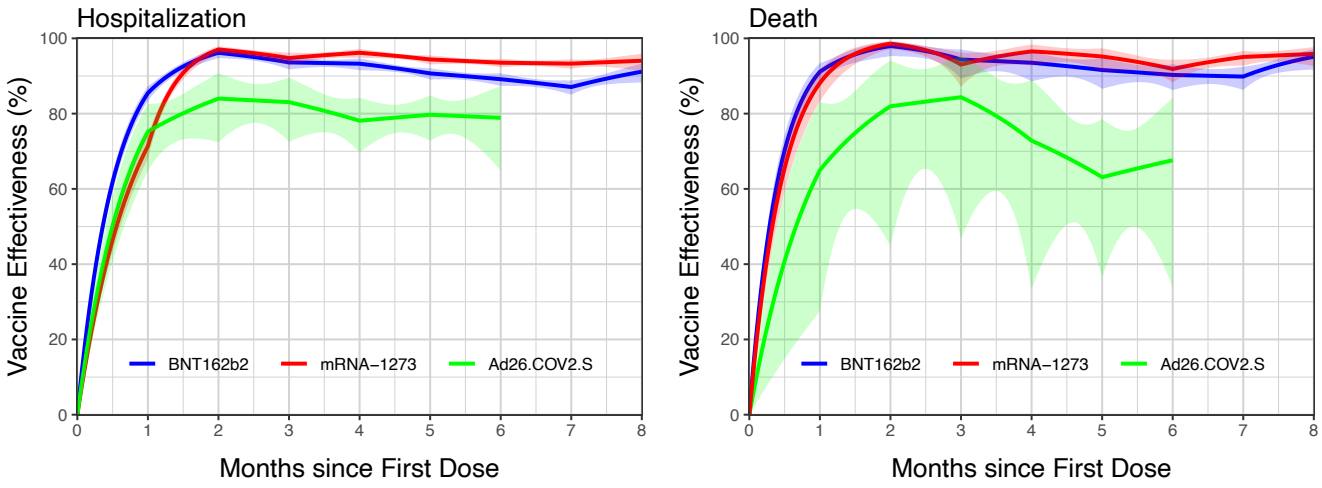


Figure S14. Vaccine effectiveness in reducing the risk of death by age group. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 50 imputations.

A. Imputing missing hospitalization and survival status by 1.2 fold of the observed event rates in vaccinated individuals



B. Imputing missing hospitalization and survival status by 1.5 fold of the observed event rates in vaccinated individuals

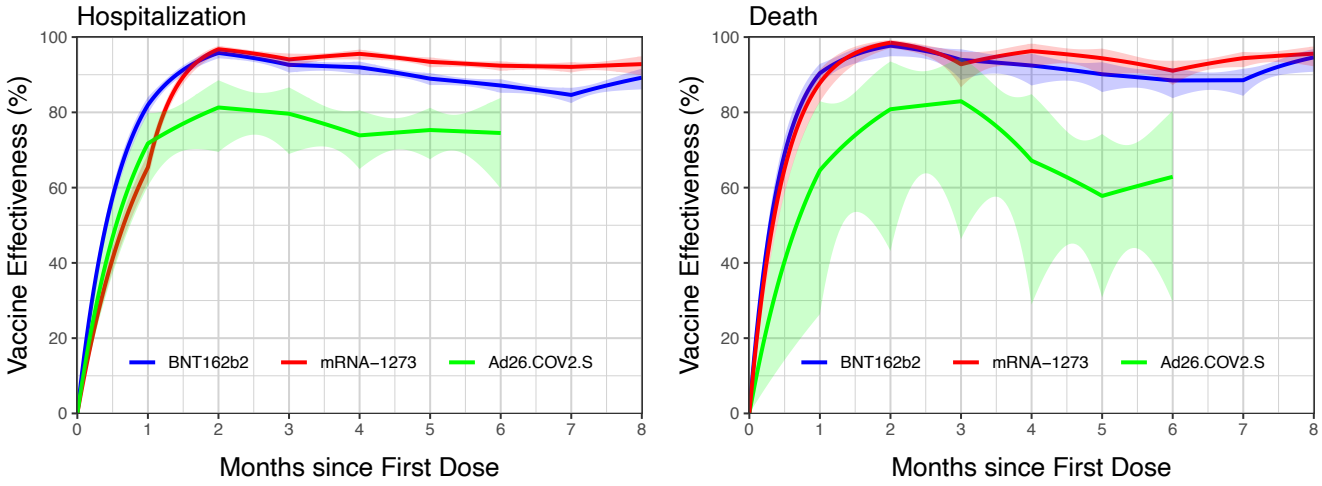


Figure S15. Vaccine effectiveness in reducing the risk of hospitalization or death with different ways of handling missing data. Estimates are shown by solid curves, and 95% confidence intervals are shown by shaded bands. Results are based on 100 imputations.

Table S1. Effectiveness of the BNT162b2 two-dose vaccine, mRNA-1273 two-dose vaccine, and Ad26.COV2.S vaccine in reducing the risk of Covid-19, as a function of time (in months) elapsed since the first dose, by age group*

Age groups	Months							
	1	2	3	4	5	6	7	8
BNT162b2								
12 – 17	95.2% (94.5, 95.7)	95.3% (94.3, 96.1)	94.2% (93.4, 95.0)	87.3% (85.2, 89.1)	83.4% (80.1, 86.2)	85.2% (77.6, 90.2)		
18 – 34	90.3% (89.6, 90.9)	94.3% (93.5, 95.0)	88.0% (86.8, 89.1)	85.6% (84.5, 86.5)	80.0% (78.8, 81.1)	76.8% (74.8, 78.7)	60.3% (55.9, 64.3)	65.3% (60.5, 69.5)
35 – 49	86.7% (85.8, 87.6)	95.4% (94.6, 96.0)	88.7% (87.5, 89.9)	83.8% (82.6, 85.0)	80.7% (79.6, 81.8)	72.8% (70.8, 74.6)	60.0% (56.1, 63.6)	59.9% (55.2, 64.0)
50 – 64	82.3% (81.1, 83.5)	95.1% (94.3, 95.8)	88.5% (87.0, 89.8)	84.8% (83.6, 85.9)	81.3% (80.1, 82.3)	76.2% (74.4, 77.9)	65.0% (61.1, 68.5)	67.6% (63.0, 71.6)
65+	72.2% (70.2, 74.1)	92.7% (91.5, 93.8)	87.6% (85.6, 89.2)	85.2% (83.2, 87.0)	74.3% (72.0, 76.4)	66.7% (64.6, 68.6)	57.4% (55.0, 59.7)	60.1% (55.3, 64.4)
mRNA-1273								
18 – 34	92.1% (91.2, 92.8)	96.1% (95.2, 96.8)	92.4% (91.1, 93.5)	91.7% (90.6, 92.6)	87.7% (86.4, 88.8)	82.8% (80.8, 84.5)	77.3% (73.2, 80.7)	77.8% (71.6, 82.7)
35 – 49	89.1% (88.0, 90.0)	96.7% (95.9, 97.3)	92.7% (91.4, 93.8)	91.0% (89.9, 92.0)	88.3% (87.3, 89.3)	81.6% (79.9, 83.2)	73.5% (70.0, 76.6)	73.1% (67.7, 77.6)
50 – 64	80.7% (79.2, 82.0)	96.0% (95.2, 96.7)	93.4% (92.1, 94.4)	92.3% (91.3, 93.1)	88.2% (87.2, 89.1)	83.0% (81.5, 84.4)	79.7% (77.1, 82.0)	79.1% (74.7, 82.7)
65+	34.1% (30.5, 37.4)	93.6% (92.6, 94.5)	90.2% (88.5, 91.6)	89.8% (88.1, 91.3)	83.0% (81.2, 84.7)	79.5% (78.0, 80.8)	75.4% (73.8, 77.0)	67.0% (62.6, 70.8)

Ad26.COV2.S						
18 – 34	77.2% (73.6, 80.3)	74.8% (69.6, 79.1)	73.2% (68.2, 77.3)	66.8% (62.9, 70.3)	57.1% (52.9, 60.9)	60.6% (52.2, 67.5)
35 – 49	74.1% (69.7, 77.9)	66.1% (59.5, 71.6)	71.0% (65.6, 75.6)	60.1% (55.0, 64.5)	59.6% (55.2, 63.6)	62.8% (55.2, 69.1)
50 – 64	74.2% (69.5, 78.1)	73.1% (67.3, 77.9)	69.6% (64.0, 74.4)	61.9% (57.1, 66.2)	61.7% (57.7, 65.4)	61.3% (53.9, 67.5)
65+	61.8% (49.6, 71.1)	73.1% (61.6, 81.2)	63.4% (50.7, 72.9)	51.9% (40.7, 61.0)	44.5% (34.4, 53.1)	43.3% (25.6, 56.8)

** 95% confidence intervals are given in parentheses below each estimate.*

Table S2. Effectiveness of the BNT162b2 two-dose vaccine, mRNA-1273 two-dose vaccine, and Ad26.COV2.S vaccine in reducing the risk of hospitalization, as a function of time (in months) elapsed since the first dose, by age group*

Age groups	Months							
	1	2	3	4	5	6	7	8
BNT162b2								
18 – 49	95.4% (92.8, 97.1)	97.3% (92.5, 99.0)	95.9% (91.1, 98.1)	95.5% (91.3, 97.6)	94.1% (90.1, 96.5)	93.4% (86.8, 96.7)	92.7% (79.5, 97.4)	94.5% (83.6, 98.2)
50 – 64	92.5% (90.1, 94.3)	96.5% (93.4, 98.2)	96.4% (93.0, 98.1)	95.4% (92.9, 97.0)	94.0% (91.8, 95.7)	93.1% (89.8, 95.4)	94.0% (87.4, 97.2)	94.7% (88.7, 97.5)
65+	80.7% (77.0, 83.9)	95.8% (93.8, 97.2)	92.4% (89.2, 94.6)	92.1% (89.1, 94.3)	87.7% (84.6, 90.2)	87.6% (85.0, 89.7)	85.2% (82.6, 87.4)	90.1% (85.7, 93.1)
mRNA-1273								
18 – 49	95.5% (91.5, 97.6)	98.8% (93.7, 99.8)	96.7% (89.3, 99.0)	97.0% (92.3, 98.8)	97.4% (93.7, 99.0)	96.3% (90.1, 98.6)	95.2% (78.6, 98.9)	97.2% (62.6, 99.8)
50 – 64	89.0% (85.2, 91.9)	96.7% (93.8, 98.3)	97.6% (94.7, 98.9)	97.4% (95.2, 98.6)	95.9% (93.8, 97.3)	96.5% (94.1, 97.9)	94.7% (89.5, 97.3)	97.5% (90.7, 99.3)
65+	61.4% (55.8, 66.3)	96.9% (95.3, 98.0)	93.4% (90.5, 95.4)	95.5% (93.2, 97.0)	92.8% (90.3, 94.6)	92.4% (90.7, 93.8)	92.7% (90.9, 94.1)	93.0% (88.8, 95.6)
Ad26.COV2.S								
18 – 49	75.1% (49.8, 87.6)	85.0% (54.2, 95.1)	89.0% (61.7, 96.8)	85.4% (66.0, 93.7)	87.1% (69.3, 94.6)	91.4% (42.2, 98.7)		
50 – 64	84.6% (70.9, 91.9)	90.3% (70.3, 96.9)	85.7% (69.2, 93.4)	84.3% (70.9, 91.6)	86.9% (77.4, 92.4)	86.4% (63.5, 94.9)		
65+	72.1% (50.1, 84.4)	82.4% (57.3, 92.8)	82.3% (62.1, 91.8)	73.3% (54.3, 84.4)	74.0% (59.8, 83.2)	71.9% (42.1, 86.3)		

** 95% confidence intervals are given in parentheses below each estimate. Each entry is based on 50 imputations.*

Table S3. Effectiveness of the BNT162b2 two-dose vaccine, mRNA-1273 two-dose vaccine, and Ad26.COV2.S vaccine in reducing the risk of death, as a function of time (in months) elapsed since the first dose, by age group*

Age groups	Months							
	1	2	3	4	5	6	7	8
BNT162b2								
18 – 64	93.1% (86.9, 96.4)	95.2% (91.8, 97.2)	96.7% (90.9, 98.8)	95.9% (92.9, 97.6)	94.9% (89.5, 97.5)	97.0% (92.6, 98.8)	98.2% (89.7, 99.7)	98.9% (84.3, 99.9)
65+	92.4% (89.3, 94.5)	97.7% (94.2, 99.1)	95.8% (90.6, 98.1)	92.8% (86.6, 96.1)	92.4% (86.6, 95.7)	90.5% (85.9, 93.7)	90.3% (86.4, 93.0)	94.9% (90.9, 97.1)
mRNA-1273								
18 – 64	72.7% (55.3, 83.3)	95.3% (89.1, 98.0)	99.2% (95.3, 99.9)	98.2% (95.7, 99.3)	96.1% (90.9, 98.3)	95.0% (91.2, 97.1)	93.6% (86.7, 96.9)	91.8% (73.6, 97.5)
65+	87.2% (81.9, 90.9)[#]	98.4% (96.7, 99.2)	94.0% (88.2, 97.0)	95.9% (90.9, 98.1)	95.8% (91.6, 97.9)	92.2% (88.3, 94.8)	95.4% (93.1, 96.9)	95.9% (92.4, 97.7)
Ad26.COV2.S								
18 – 64	58.7% (-4.6, 83.7)	85.1% (62.5, 94.1)	94.6% (65.7, 99.2)	87.7% (70.9, 94.8)	72.2% (29.4, 89.0)	91.3% (6.9, 99.2)		
65+	79.4% (39.9, 93.0)	83.1% (30.9, 95.9)	82.3% (28.3, 95.6)	74.9% (27.2, 91.3)	66.8% (33.0, 83.6)	66.5% (26.6, 84.7)		

** 95% confidence intervals are given in parentheses below each estimate. Each entry is based on 50 imputations.*

[#] Extrapolated from the estimates at months 0 and 2.