

Supplementary materials belonging to

**Population immunity to SARS-CoV-2 in US states and counties due to infection and vaccination,
January 2020–November 2021**

Authors: Fayette Klaassen, Melanie H. Chitwood, Ted Cohen, Virginia E. Pitzer, Marcus Russi, Nicole A. Swartwood, Joshua A. Salomon and Nicolas A. Menzies

Correspondence to: fklaassen@hsph.harvard.edu

This file contains:

eMethods.

eTable 1.

eTable 2.

eFigure 1.

eFigure 2.

eFigure 3.

eFigure 4.

eFigure 5.

eFigure 6.

eMethods

Imputation of cases and deaths data for Nebraska counties

The time-series of cases and deaths data are truncated for counties in Nebraska after June 30, 2021, due to non-reporting of cases and deaths data at the county level. To produce estimates for Nebraska we imputed cases and deaths for each county based on state-level cases and deaths, assuming each county experienced the same proportional changes in cumulative cases and deaths as reported at the state level.

Imputation and temporal disaggregation of vaccination data

We imputed data for 2,006 counties in the weekly vaccination data (*First dose*) (minimum missing weeks: 1, 567 counties, maximum missing weeks: 35, 8 counties). The states with the highest percentage of counties with missing data are Nebraska and South Dakota. After the temporal disaggregation of the weekly time series to daily data, a remaining 57 counties had missing data at the end of the time-series (minimum missing days: 8, maximum missing days: 211).

We used a three-step process to render a daily timeseries from the (irregular) weekly timeseries. First, we used linear interpolation to impute missing *First dose* and *Fully vaccinated* data on a weekly (7-day) interval from January 10th, 2021. Second, we disaggregated the weekly time series to the daily level, using a second order smoothness constraint. This method fits a smooth line that always passes the observation on the first day of the week and increases in the following days. Third, we imputed the remaining missing data of at the end of the time series using the following approach.

We imputed the missing data of *First dose* at time X using the fraction of people who get fully vaccinated. We assume there is an average lag of Y days between *First dose* and *Fully vaccinated*. For a single-dose vaccine (Johnson, 8% of all fully vaccinated individuals), the status *Fully vaccinated* is achieved on the same day as the *First dose*. For Moderna (37.3% of all fully vaccinated individuals), the recommended time between the first and second dose is four weeks (28 days). For Pfizer-BioNTech (54.6%), the recommended time is three weeks (21 days)^{1,2}. The weighted average of time until fully vaccinated is 22 days.

For each time point, we computed the fraction fully vaccinated. The missing values of *First dose* were calculated from the imputed values and had to meet the constraint that at any time, the total number *First dose* must be smaller than the total population size. We further constrain the time series to be monotonically increasing.

Children under 12 years old were not eligible for COVID-19 vaccination in the time period of analysis. We constrained the fraction of the full population with a *First dose* to never exceed the fraction of the full population of 12 years and older. For example, Santa Cruz, Arizona, reports 99% of its population has received a *First dose*, while only 81% of the population is 12 years and older. In this case, we constrain the fraction vaccinated to be 81%.

Logistic regression rendering state specific odds ratios of vaccination given infection status

We used a logistic regression model to estimate the association between self-reported vaccination status and prior COVID-19 diagnosis. The regression model included fixed effects for week and state. We also included state-level random coefficients for prior COVID-19 diagnosis, to allow for state-level differences in the overlap between vaccination and prior infection. Below is the full regression model, where $X_{(.)}$ is a model matrix for the specified variable. We used a binomial likelihood to model the number people with a self-reported vaccination status (n) out of the total (N).

$$\begin{aligned} \text{logit}(\theta) &= \mathbf{b}_0 + X_{\text{week}} \mathbf{b}_{\text{week}} + X_{\text{state}} \mathbf{b}_{\text{state}} + X_{\text{state} \times \text{covid}} (\mathbf{b}_{\text{covid}} + \mathbf{g}_{\text{r.e.}}) \\ \mathcal{L}(\theta|N, n) &= \theta^n (1 - \theta)^{N-n} \\ g_{\text{r.e.}} &\sim \mathcal{N}(0, \sigma) \\ \sigma &\sim U(0, \infty) \end{aligned}$$

Derivation of joint probability from odds ratio

We have estimates of the probability of ever being infected and (adjusted) data of the probability of ever being vaccinated. In addition, we have a specified a data-driven prior distribution on the odds ratio of vaccination for those with a prior diagnosis versus those without a prior diagnosis. We want to compute the probability of being immune (the probability of ever being vaccinated and/or infected).

Following the rules of probability and our definition of immunity, we can compute the probability of being immune as a function of the probability of vaccinated, infected, and the conditional probability of being vaccinated given prior infection.

$$\begin{aligned}
 p_{\text{immune}} &= p_{\text{vac} \cup \text{inf}} \\
 &= p_{\text{vac}} + p_{\text{inf}} - p_{\text{vac} \cap \text{inf}} \\
 &= p_{\text{vac}} + p_{\text{inf}} - p_{\text{vac}} p_{\text{vac} | \text{inf}} \\
 &= p_{\text{vac}} + p_{\text{inf}} - p_{\text{vac}} \frac{O_{\text{vac} | \text{inf}}}{1 + O_{\text{vac} | \text{inf}}}
 \end{aligned}$$

To compute the probability of being immune, we need an expression for the unknown $O_{\text{vac} | \text{inf}}$. Making use of the fact that:

$$\begin{aligned}
 \frac{O_{\text{vac} | \text{inf}}}{O_{\text{vac} | \text{not inf}}} &= OR \\
 O_{\text{vac} | \text{inf}} &= OR O_{\text{vac} | \text{not inf}}
 \end{aligned}$$

we can express p_{vac} as a function of p_{inf} , OR and $O_{\text{vac} | \text{inf}}$

$$\begin{aligned}
 p_{\text{vac}} &= p_{\text{inf}} p_{\text{vac} | \text{inf}} + p_{\text{not inf}} p_{\text{vac} | \text{not inf}} \\
 &= p_{\text{inf}} \frac{O_{\text{vac} | \text{inf}}}{1 + O_{\text{vac} | \text{inf}}} + p_{\text{not inf}} \frac{O_{\text{vac} | \text{not inf}}}{1 + O_{\text{vac} | \text{not inf}}} \\
 &= p_{\text{inf}} \frac{OR O_{\text{vac} | \text{not inf}}}{1 + OR O_{\text{vac} | \text{not inf}}} + (1 - p_{\text{inf}}) \frac{O_{\text{vac} | \text{not inf}}}{1 + O_{\text{vac} | \text{not inf}}} \\
 &= \frac{p_{\text{inf}} OR O_{\text{vac} | \text{not inf}}}{1 + OR O_{\text{vac} | \text{not inf}}} + \frac{O_{\text{vac} | \text{not inf}} - p_{\text{inf}} O_{\text{vac} | \text{not inf}}}{1 + O_{\text{vac} | \text{not inf}}}
 \end{aligned}$$

For readability, let us define short hands for the above terms.

$$\begin{aligned}
 p_{\text{vac}} &= v \\
 p_{\text{inf}} &= i \\
 O_{\text{vac} | \text{not inf}} &= d \text{ (for odds)} \\
 OR &= r \text{ (for odds ratio)}
 \end{aligned}$$

Using the new labels, rearranging the terms, and equating to zero renders a quadratic equation of d ($O_{\text{vac} | \text{not inf}}$):

$$\begin{aligned}
v &= \frac{ird}{1+rd} + \frac{d-id}{1+d} \\
0 &= v(1+rd)(1+d) - ird(1+d) - (d-id)(1+rd) \\
&= v + vrd + vd + vrd^2 - ird - ird^2 - d + id - rd^2 + ird^2 \\
&= (vr-r)d^2 + (vr+v-ir+i-1)d + v
\end{aligned}$$

Making the following substitutions,

$$\begin{aligned}
(vr-r) &= a \\
(vr+v-ir+i-1) &= b \\
v &= c
\end{aligned}$$

allows us to solve the quadratic formula, rendering a solution for $O_{\text{vac|not inf}}$, which allows us to compute the joint probability of being vaccinated and infected, and finally the probability of being immune.

$$\begin{aligned}
\frac{-b \pm \sqrt{b^2 + 4ac}}{2a} &= d = O_{\text{vac|not inf}} \\
O_{\text{vac|inf}} &= O_{\text{vac|not inf}} \text{ OR} \\
p_{\text{vac|inf}} &= \frac{O_{\text{vac|inf}}}{1 + O_{\text{vac|inf}}} \\
p_{\text{vac} \cap \text{inf}} &= p_{\text{vac}} p_{\text{vac|inf}}
\end{aligned}$$

Waning functions

We propose four simplistic functions of waning of immunity (eFigure 1), under three scenarios: a baseline scenario, used in the analyses in the manuscript, and an additional optimistic and pessimistic scenario (dotted lines in eFigure 1) for sensitivity analyses.

Baseline scenario:

- Protection against infection conferred by vaccination or natural infection is 80% for the first two months after natural infection or vaccination, then declines to 50% in month 4, and then declines more slowly over the next five months, such that the protection nine months after infection or vaccination is 25% and constant thereafter.
- Protection against infection conferred by the combination of vaccination and natural infection is 90% and does not decline.
- Protection against severe outcomes conferred by vaccination or natural infection is 95% for the first six months after infection or complete vaccination, and then declines by 10 percent points every six months.
- Protection against severe outcomes conferred by the combination of natural infection and vaccination is 95% and does not decline.

Optimistic scenario:

- Protection against infection conferred by vaccination or natural infection is 90% for the first two months after natural infection or vaccination, then declines to 75% over the next two months, and then declines further over the next five months, such that the protection nine months after infection or vaccination is 50% and constant thereafter.
- Protection against infection conferred by the combination of vaccination and natural infection is 95% and does not decline.
- Protection against severe outcomes conferred by vaccination or natural infection is 100% for the first six months after the infection or complete vaccination, and then declines by 5 percent points every six months.
- Protection against severe outcomes accrued by the combination of natural infection and vaccination is 100% and does not decline.

Pessimistic scenario:

- Protection against infection conferred by vaccination or natural infection is 75% for the first two months after natural infection or vaccination, then declines to 45% over the next two months, and then declines further over the next five months, such that the protection nine months after infection or vaccination is 20% and constant thereafter.
- Protection against infection conferred by the combination of vaccination and natural infection is 80% for the first six months after being previously infected and vaccinated, and then declines by 10 percent points every six months.
- Protection against severe outcomes conferred by vaccination or natural infection is 100% for the first six months after the infection or complete vaccination, and then declines by 20 percent points every six months.
- Protection against severe outcomes conferred by the combination of natural infection and vaccination is 90% for the first six months after being previously infected and vaccinated, and then declines by 10 percent points every six months.

Validation of Household Pulse Survey data

We fit a logistic regression model on the Axios-Ipsos Coronavirus Survey data³, similar to the logistic regression we used for the Household Pulse Survey data, that was used in the main analyses. The Axios-Ipsos survey data does not have state-specific estimates, rendering a simplified regression equation:

$$\begin{aligned} \text{logit}(\theta) &= b_0 + X_{\text{week}} \mathbf{b}_{\text{week}} + X_{\text{covid}} b_{\text{covid}} \\ \mathcal{L}(\theta|W, w) &= \theta^w (1 - \theta)^{W-w} \\ b_{(\cdot)} &\sim \mathcal{N}(0, 1000) \end{aligned}$$

eTable 1: Means and standard deviations of the log-normal prior distributions for odds ratio of vaccination given infection status for each US state.

State	Odds ratio	Lower bound 95% CrI	Upper bound 95% CrI
Florida	0.400	0.361	0.444
Michigan	0.455	0.390	0.529
Oregon	0.469	0.397	0.555
Washington	0.47	0.401	0.551
North Carolina	0.475	0.424	0.532
Pennsylvania	0.484	0.428	0.547
Alaska	0.488	0.407	0.586
Illinois	0.495	0.441	0.557
New Mexico	0.501	0.403	0.623
Maryland	0.503	0.425	0.596
New Jersey	0.504	0.439	0.578
Arizona	0.505	0.444	0.574
New York	0.505	0.46	0.556
Vermont	0.507	0.406	0.633
Hawaii	0.508	0.417	0.618
Delaware	0.511	0.421	0.621
Indiana	0.512	0.449	0.584
West Virginia	0.512	0.431	0.608
Connecticut	0.516	0.44	0.605
North Dakota	0.516	0.431	0.617
Idaho	0.516	0.434	0.614
Colorado	0.516	0.447	0.596
Nevada	0.517	0.431	0.621
Massachusetts	0.518	0.441	0.607
Ohio	0.518	0.466	0.575
Wyoming	0.519	0.427	0.631
Rhode Island	0.519	0.432	0.625
Kansas	0.521	0.441	0.616
Minnesota	0.523	0.446	0.612
Maine	0.525	0.442	0.623
Mississippi	0.526	0.445	0.621
District of Columbia	0.53	0.438	0.641
Georgia	0.53	0.472	0.595
Tennessee	0.532	0.457	0.62
New Hampshire	0.534	0.435	0.656

Montana	0.535	0.439	0.651
Oklahoma	0.535	0.456	0.627
Missouri	0.535	0.467	0.614
Utah	0.536	0.451	0.637
Nebraska	0.537	0.435	0.663
Virginia	0.538	0.459	0.629
South Dakota	0.538	0.444	0.653
Arkansas	0.541	0.456	0.643
South Carolina	0.541	0.464	0.632
California	0.545	0.502	0.591
Iowa	0.554	0.472	0.652
Wisconsin	0.555	0.483	0.638
Alabama	0.56	0.484	0.649
Louisiana	0.567	0.487	0.659
Kentucky	0.569	0.486	0.665
Texas	0.581	0.532	0.634

eTable 2: Effective protection outcomes for each US state on October 31, 2021 under base-case, pessimistic and optimistic scenarios.

State	<i>Effective protection against infection; base-case scenario (pessimistic; optimistic)</i>	<i>Effective protection against severe disease; base-case scenario (pessimistic; optimistic)</i>
Wisconsin	42.3% (32.4%-58.1%)	72.5% (56.6%-80.7%)
District of Columbia	48.7% (38%-65.8%)	78.5% (61.3%-87.7%)
Maryland	47.7% (36.8%-64.8%)	78.1% (61%-87.7%)
Louisiana	50.5% (37.7%-64.6%)	76.7% (59.9%-84.6%)
West Virginia	37.2% (21.4%-42.6%)	62.9% (45.1%-66.1%)
Illinois	47.1% (37.7%-65.4%)	77% (61.3%-86.8%)
New Mexico	57.4% (42.5%-71.3%)	83% (64%-90.2%)
Arizona	55.2% (41.1%-69.8%)	81.6% (63.3%-89.1%)
Texas	53% (38.5%-66.8%)	79.5% (60.8%-86.7%)
Oklahoma	57.5% (38.8%-67.4%)	81.6% (60.3%-87.4%)
South Dakota	49.8% (36.9%-64.6%)	76.8% (59.5%-85.3%)
Iowa	44.4% (34.4%-60%)	73.5% (58.4%-81.9%)
North Carolina	43.9% (32.5%-57.8%)	71.8% (56.2%-79.9%)
Arkansas	48.5% (33.8%-60.4%)	75.1% (57.1%-81.9%)
Tennessee	48.3% (33.9%-60.3%)	75.4% (57.6%-82%)
Kentucky	49.9% (34.3%-60.8%)	76% (56.8%-82.1%)

Virginia	42.9% (31.2%-58.2%)	72% (53.2%-80.4%)
Kansas	47.8% (35.3%-61.8%)	75.4% (58.5%-83.2%)
Michigan	45.7% (33.3%-57.6%)	73.8% (56.4%-79.5%)
Missouri	41.4% (30%-54.2%)	69.3% (53.5%-76.3%)
Alabama	53.1% (36.1%-64.1%)	78.8% (59%-85.1%)
Indiana	42.7% (31.8%-56.9%)	70.9% (55.1%-78.7%)
Nevada	53.7% (38.7%-66.8%)	79.6% (60.8%-86.9%)
California	51% (39.2%-67.7%)	79.6% (62.1%-88.3%)
Montana	48.8% (33.1%-59.3%)	74.8% (56.3%-81.2%)
New Hampshire	46% (30.9%-58.2%)	76.1% (54.9%-83.2%)
Ohio	40.6% (30.8%-55.8%)	69.2% (54.8%-78%)
Hawaii	41.6% (31.7%-60.5%)	78% (61%-89.3%)
Idaho	52.5% (33.2%-60.3%)	76% (55.5%-81.9%)
Wyoming	53.4% (34%-61.6%)	77.8% (56.4%-83.4%)
Massachusetts	49.1% (37.6%-66.3%)	79.3% (59.5%-87.9%)
Georgia	51.8% (36.4%-63.5%)	77.9% (59.3%-84.6%)
New Jersey	52.4% (41.4%-70.5%)	80.3% (62.4%-89.9%)
Pennsylvania	52.1% (37.8%-66.7%)	80.2% (60.9%-88.6%)

Washington	45.5% (31.5%-59.8%)	74.8% (53.3%-82.8%)
Maine	44.3% (31.2%-58.6%)	76.5% (57.9%-84.6%)
North Dakota	47% (35.3%-61.1%)	75% (60.1%-82.8%)
Colorado	56.9% (42.8%-72.1%)	83.3% (65.4%-92.3%)
Nebraska	45.2% (31.5%-57.7%)	72.1% (53.7%-79.2%)
Minnesota	43.2% (33.2%-59.2%)	72.6% (57.4%-81.7%)
Oregon	45.5% (31.2%-58.1%)	74.2% (55.8%-82.1%)
New York	53.8% (41.4%-69.3%)	81% (63.2%-89.6%)
Delaware	51.7% (38.7%-66.3%)	79% (61.2%-87%)
Alaska	54.1% (35.2%-63%)	77.4% (57.1%-83.9%)
South Carolina	49.7% (34.4%-61.1%)	75.5% (57.2%-82.2%)
Vermont	37.2% (26.9%-51.3%)	67.8% (51.9%-75.7%)
Florida	60.9% (40.2%-71.1%)	83.8% (59.7%-89.8%)
Rhode Island	55.3% (41.3%-69.9%)	81.8% (63%-89.7%)
Mississippi	50.8% (36.6%-62.9%)	76.9% (59.6%-83.7%)
Connecticut	51.1% (39.1%-68%)	81.4% (63%-90.6%)
Utah	45.3% (32.6%-57.8%)	71.7% (55.7%-79.3%)

eFigure 1: Assumed waning curves for protection against infection and severe disease, from natural infection and/or vaccination.

Footnote: Dashed lines indicate the optimistic and pessimistic scenario.

eFigure 2: County-level estimates of the fraction of the population with *effective protection* at four time-points between December 31, 2020 and October 31, 2021.

eFigure 3: Contribution of prior infection and vaccination to the fraction of the population *immunologically exposed* for each US state over time.

Footnote: Purple: percent ever infected and not vaccinated. Green: percent vaccinated and not prior infected. Blended color: percent ever infected and vaccinated.

eFigure 4: Contribution of effective protection from prior infection and vaccination or both to the total effective protection for each US state over time.

Footnote: Purple: effective protection from those infected and not vaccinated. Green: effective protection from those vaccinated and not prior infected. Blended color: effective protection from those infected and vaccinated.

eFigure 5: Comparison of immunity estimates for each US state, with credible intervals, with blood donor estimates of immunity⁴.

Footnote: red lines indicate the blood donor estimates

1. Centers for Disease Control and Prevention. COVID data tracker. <https://covid.cdc.gov/covid-data-tracker>. Published 2020. Accessed October 22, 2021.
2. Centers for Disease Control and Prevention. Vaccines for COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/index.html>. Published 2021. Accessed October 22, 2021.
3. Ipsos. Axios-Ipsos Coronavirus Index, waves 34-48. Ann, Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], doi:10.3886/E144903V1
4. Jones JM, Stone M, Sulaeman H, et al. Estimated US Infection- and Vaccine-Induced SARS-CoV-2 Seroprevalence Based on Blood Donations, July 2020-May 2021. *JAMA*. 2021. <https://doi.org/10.1001/jama.2021.15161>.