Supplemental Methods Appendix to:
Race-Specific, State-Specific COVID-19 Vaccination Rates Adjusted for Age
Elizabeth Wrigley-Field, Kaitlyn M. Berry, and Govind Persad

This analysis integrates several distinct datasets, summarized and cited in Table S1. The most important dataset we draw on gives the portion of each state's total vaccinations that were given to each racial/ethnic group (dataset 1 in Table S1), collected by the Kaiser Family Foundation (KFF). In order to transform these "vaccine shares" into age-standardized vaccination rates, there are two main steps. First, we transform vaccination shares into raw vaccination rates (i.e., per-capita vaccinations) for each racial/ethnic group in each state. Second, we standardize these raw vaccination rates for age using an "indirect age standardization" method. Here, we explain these two steps in detail.

Throughout, we use "vaccinated" to mean having received two doses of an mRNA vaccine or one dose of a Johnson and Johnson vaccine (i.e., "fully vaccinated" without consideration of booster shots). The time period we cover (with datasets reflecting cumulative vaccinations through early-to-mid October 2021) summarizes a situation in which people aged 12+ had been eligible for vaccination for long enough that appointments were widely available without delay in most places, but nobody younger than age 12 was yet eligible.

Step 1: Constructing raw (crude) vaccination rates

No single dataset gives rates of vaccination by race/ethnicity for all states. In order to construct vaccination rates for each racial/ethnic group in each state, we took aggregate vaccination rates for each state (dataset 2) and applied the vaccine share for each racial/ethnic group (dataset 1) to those state aggregate rates. For example, if 50% of a state is vaccinated (dataset 2) and 10% of its vaccinations are of Hispanic people (dataset 1), then 5% of the state's total population is vaccinated Hispanic people. We divided these by the portion of the state's total population (dataset 3) that is Hispanic (datasets 4 and 5), which yielded the percentage of the state's Hispanic population that is vaccinated. This percentage is what we refer to as a "raw rate" of vaccination for each racial/ethnic group in each state (also called a crude rate).

States vary widely in the apparent quality of their vaccination data (dataset 1), as summarized in Table S2. The data quality measures we report in Table S2 (e.g., percent of vaccinations with unknown race/ethnicity) are reported by KFF; the underlying vaccination data are collected by KFF from state departments of health. We drop Pennsylvania from all analyses because Pennsylvania's reported race/ethnicity-specific vaccination data (dataset 1) exclude Philadelphia, while Pennsylvania's aggregate vaccination rates (dataset 2) and populations (datasets 3-5) include Philadelphia.

Step 2: Age-standardizing vaccination rates

There are two ways to age-standardize rates. "Direct age standardization" involves reweighting age-specific rates for each population. However, age-specific vaccination rates are not available for race/ethnicity-specific, state-specific populations. Therefore, we use "indirect age standardization." This method translates each racial/ethnic state population's vaccination rate into a scale reflecting the degree to which the population is more vaccinated or less vaccinated than would be expected, based on the population's age distribution.

In order to estimate these indirectly age-standardized rates, we begin with a set of age-specific vaccination rates for the United States as a whole (dataset 6). This age pattern is shown in Figure S1 across the Centers for Disease Control and Prevention (CDC)'s vaccine-eligible age categories as of the time of the study (ages 12-15, 16-17, 18-24, 25-39, 40-49, 50-64, 65-74, and 75+), which are used throughout the analysis. For comparison, Figure S2 shows the age distribution across all U.S. states of age for each racial/ethnic group. Note that the CDC vaccination rates for ages 65-74 seem unrealistically high, which may reflect inaccurate population denominators and/or booster shots (perhaps among those not yet eligible) being misreported as new vaccinations. We use a measure of "fully-vaccinated" rather than "any vaccine" (measuring having received at least one dose) in order to limit the latter possibility. Note also that, although nobody younger than age 12 was formally eligible, the CDC data (dataset 6) do show some vaccinations among younger children (0.3% vaccination among children under age 12). Since all racial/ethnic state populations will be normalized to this standard age schedule of vaccination, the comparative analysis will be unbiased as long as any errors in the CDC age schedule of vaccination are proportional across ages but will be biased if the CDC data contain greater error for some ages compared with others.

Using this national age pattern of vaccination and the age distribution of each racial/ethnic group population in each state (datasets 4 and 5), we estimate the "expected" vaccination rate of each racial/ethnic group in each state based on its age distribution. This reflects the aggregate vaccination rate that each state would have if it had the national aggregate vaccination rates for each of its age groups.

Finally, we take the ratio of the actual vaccination rate to this expected vaccination rate for each racial/ethnic population in each state. For convenience, in this Appendix, we refer to this quantity as the Indirectly Standardized Rate (ISR) (in a mortality context, it is sometimes called a Comparative Mortality Ratio). When the ISR exceeds 1, it means the population is more vaccinated than would be expected; for example, a ISR of 1.2 means a population whose raw vaccination rate is 120% of what would be expected based on its age distribution. When the ISR is less than 1, the population is less vaccinated than would be expected. This ISR is the outcome for each analysis shown in Figure 1.

Note that the state aggregate ISRs are, on average, slightly below 1 (mean=0.98) and the two largest racial groups (white and Black) are substantially below 1. This may seem puzzling since one might expect that the population as a whole should have the "expected" vaccination levels. The discrepancy is most likely driven by two factors. First, "other race" and race-not-reported observations are included in dataset 2's national data but excluded from dataset 1's race-specific data, which might depress rates for specific racial groups (but not for the state aggregate rates). Second, dataset 1 is state-level data, where all states count equally, while dataset 2 is national-level, population-weighted data, where large states count more. Thus, state aggregate ISRs hover below 1 because larger states tend to have higher vaccination rates; when states are weighted by population, their aggregate ISRs have mean=1.

Racial/ethnic groups

The racial categories we use are white, Black, Asian-American/Asian/Pacific Islander/Native Hawaiian, Native, and Hispanic. The KFF vaccination dataset (dataset 1) distinguishes Asian from native Hawaiian, but because the National Center for Health Statistics (NCHS) population

data (datasets 3-5) use a collapsed Asian/PI category that includes native Hawaiians, we collapse these categories in the vaccination rates as well.

Not all states report each racial/ethnic group. These states are listed, along with other limitations of dataset 1, in Table S2. We use all reported racial/ethnic categories in each state, except for "other race." We omit this category (except for Hispanic individuals in this category, who are treated as Hispanic) out of concern that this group may have a different meaning in state vaccination records than in NCHS population denominators. In addition, the racial shares that we use are shares among vaccinations with known race/ethnicity. In some states, the percent of vaccinations with unknown race/ethnicity is quite high--up to 21% (in Washington, DC).

States differ in their treatment of Hispanic populations. Some states treat "Hispanic" as a racial group. In these states, "White" means non-Hispanic White, and so forth. We refer to these as states with "mutually exclusive" categories. Other states treat Hispanic as a "cross-cutting group." In those states, "White" includes White Hispanic, and the total vaccination shares add up to more than 100% because the Hispanic population shares are double-counted. We rely on KFF data notes to distinguish how each state handled Hispanic populations (checked against whether vaccination shares total to 100%, with rounding error, or more than 100%).

We construct population denominators that match each way of counting the Hispanic population: mutually exclusive (dataset 4) or cross-cutting (dataset 5). We use the appropriate denominator for each state so that vaccinations and populations were constructed in the same way. The handling of Hispanic populations in constructing population denominators can matter substantively.

In Figure 1's Panels B-D, state abbreviations for non-Hispanic populations are denoted with parentheses when the state uses cross-cutting Hispanic populations and without parentheses when the state uses mutually exclusive Hispanic populations. When states report racial groups with cross-cutting Hispanic populations (the groups denoted with parentheses), race-specific divergences from what would be expected based on age may be understated, compared to the divergences of mutually exclusive racial groups. This is because the Hispanic populations (where reported) have vaccinations close to what would be expected based on their age distributions.

In the sensitivity analyses that exclude state-race observations where Hispanic individuals comprise at least 10% of a non-Hispanic racial group, the number of state-race observations that are excluded varies widely by state. For example, in New York, only the Asian/Pacific Islander group is included in the restricted sample, whereas in Ohio, only the Native group is excluded. The portion of observations excluded varies even more widely by race. Out of 24 states that report vaccination with Hispanic identity cross-cutting race, 15 white observations are included, while 9 are excluded; 20 Black observations are included, while 4 are excluded; 21 Asian/Pacific Islander observations are included, while zero are excluded and 3 states did not report vaccinations for this group; and only 2 Native observations are included, while 3 are excluded and 19 states did not report vaccinations for this group. (These included observations are joined, in this alternative measure, by 21 white, 21 Black, 17 Asian/Pacific Islander, and 10 Native observations from states where racial categories exclude Hispanic individuals, as well as 20 Hispanic observations.) In general, Native populations are excluded from the restricted sample at a higher rate because many states that use cross-cutting

racial/ethnic designations have relative many Hispanic-Native individuals, who represent a large portion of typically small Native populations. Because this sensitivity analysis excludes so many observations, we use it merely to verify that the relationship between Trump vote and total and race-specific vaccination is not confounded by differences between states' racial classification systems.

Analyses

An alternative presentation of the main results is presented in Appendix Figure S3. This presentation may aid comparison of racial groups to the distribution of state aggregate rates, at the cost of more visually complex graph panels and lack of labeling of individual states.

The density distributions shown in Figure 1 Panel A (and Figure S3 Panel A) are estimated with Epanechnikov kernels with bandwidth chosen to minimize mean integrated squared error under assumption of a Gaussian distribution and kernel, estimated using Stata. In Panel A, the state aggregate distribution includes four states that had total vaccination data, but no race-specific data, in dataset 1: Montana, Nebraska, North Dakota, and Wyoming.

In our analysis of vaccination rates in relation to 2020 Trump vote share (dataset 7), we treat Trump share as a broad proxy for a wide array of state politics, which this analysis is unable to distinguish. This analysis omits Washington, D.C., which is included elsewhere as a "state" (in Figure 1 Panel A), because its Trump vote share (5%) is far outside the range of all other states (30%-69%). Regression results are reported in Table S3A and Table S4A. These regressions are repeated with the exclusion of race-state observations in which Hispanic individuals are more than 10% of an other-than-Hispanic race, and reported in Table S3B and Table S4B, respectively.

Finally, we note that our results (with data from mid-October, 2021) differ from individual-level polling presented by KFF a few weeks later (KFF COVID-19 Vaccine Monitor: November 2021). The KFF polling data show a smaller discrepancy between white and Black vaccination rates (not adjusted for age) than we find, with 67% of Black and 73% of white respondents reporting having received at least one vaccination shot. Since our age adjustment will tend to decrease this disparity by adjusting for white populations being over-represented at high-vaccine ages, this presents a puzzle. Potential reasons for this discrepancy include: smaller racial disparities in one-shot vs. two-shot measures (e.g., if Black people are disproportionately likely to have received incomplete vaccination); sampling variability (the KFF poll's margin of error is 4 points for white and 6 points for Black populations); under-reporting of white vaccination and/or over-reporting of Black vaccination in polling data due to contextual social desirability biases; and errors in the state vaccine registration systems that we rely on (dataset 1), particularly any errors that lead to underreporting of Black vaccination in state systems.

Repository

Data and Stata code can be accessed at https://osf.io/h4p8t/

REFERENCES TO THE APPENDIX

KFF COVID-19 Vaccine Monitor: November 2021. Retrieved from https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-november-2021/ on January 18, 2022.

Yoon, Paula, Jeffrey Hall, Jennifer Fuld, S. Linda Mattocks, B. Casey Lyons, Roma Bhatkoti, Jane Henley, A. D. McNaghten, Demetre Daskalakis, and Satish K. Pillai. 2021. "Alternative Methods for Grouping Race and Ethnicity to Monitor COVID-19 Outcomes and Vaccination Coverage." *Mortality and Morbidity Weekly Report* 70: 1075-1080.

APPENDIX FIGURES

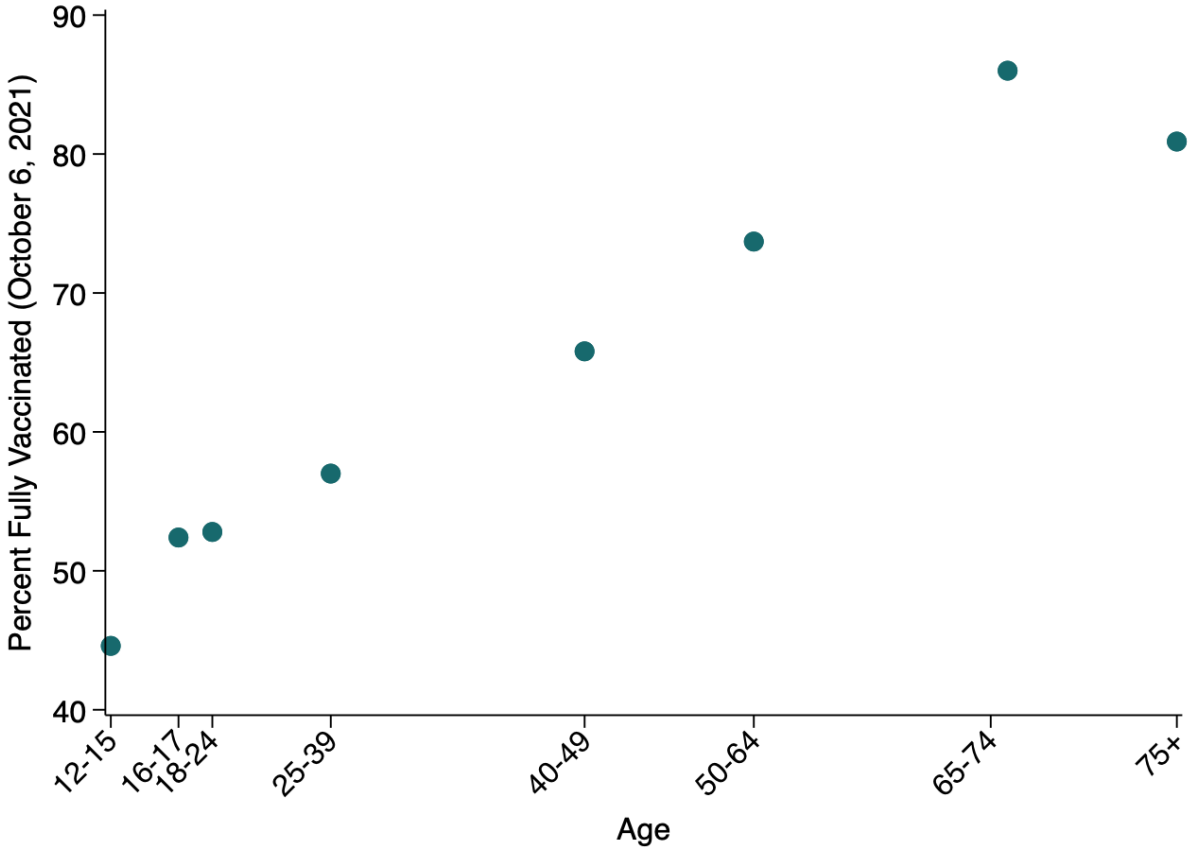
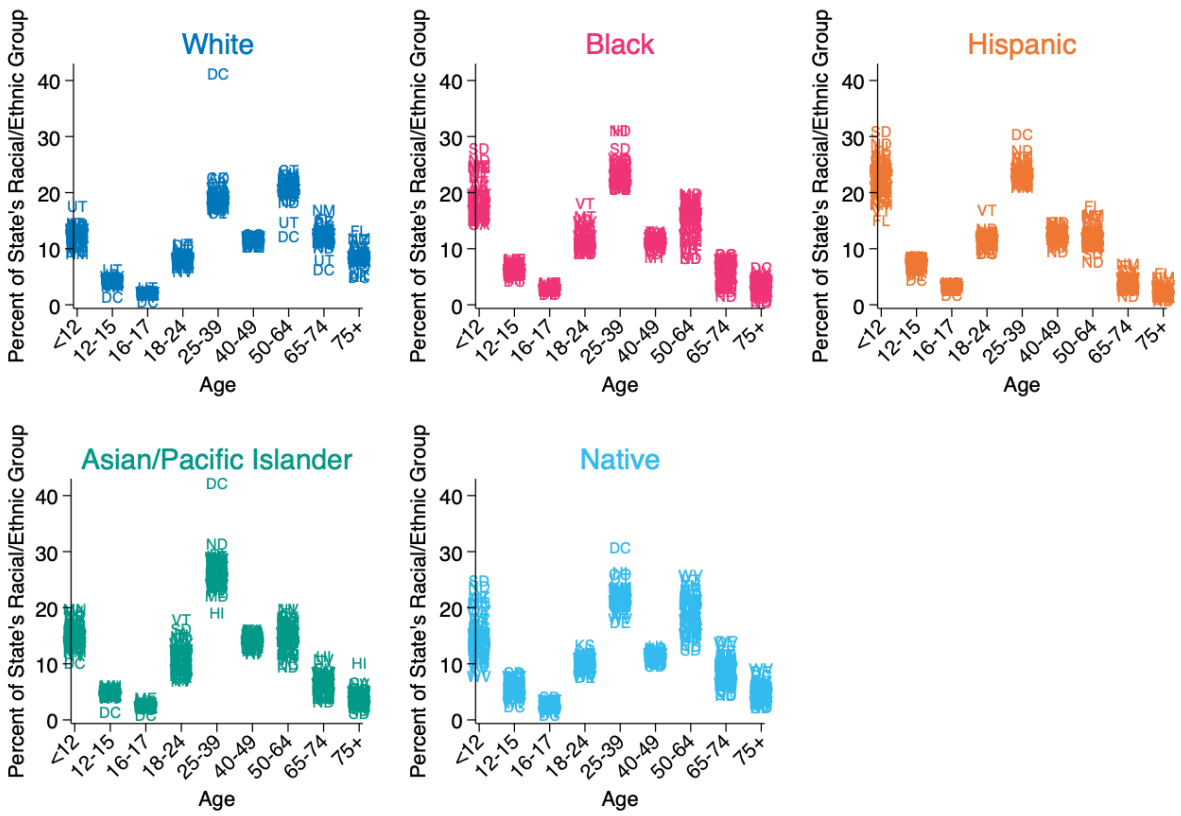


Figure S1. Vaccination status by age. Data: CDC (Table S1's "dataset 6").



State values are jittered to make distributions visible

Figure S2. Age distributions by race and state. Data: NCHS (Table S1's "dataset 4"). All racial groups besides Hispanic are limited to non-Hispanic individuals.

State Distributions of Age–Standardized COVID–19 Vaccination by Race/Ethnicity

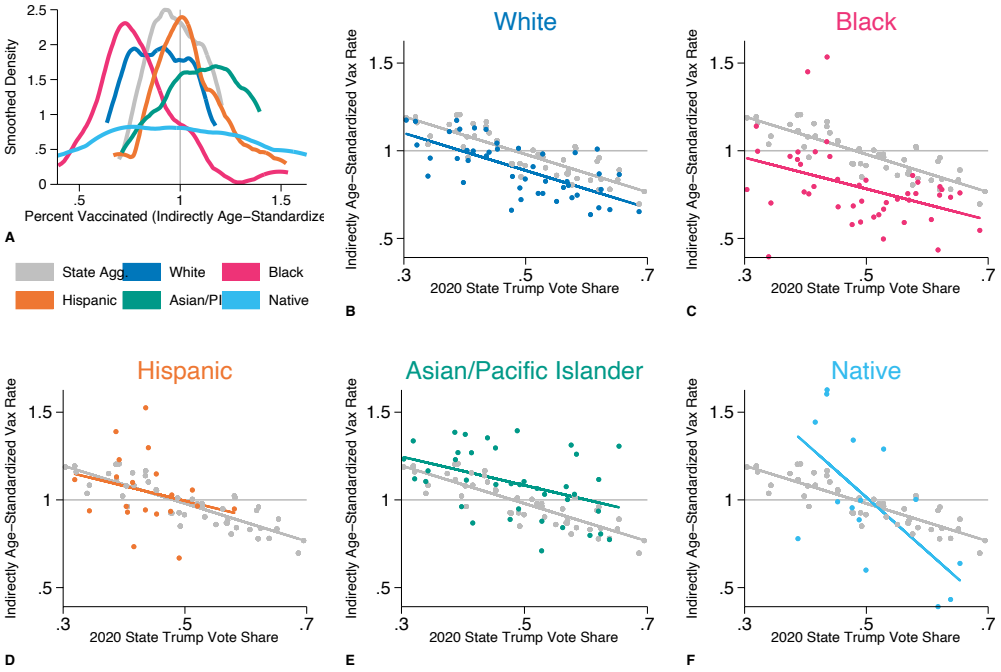


Figure S3. Indirectly age-standardized vaccination rates by state and race/ethnicity. When these rates exceed 1, the population displayed has vaccination rates higher than would be expected based on its age distribution; when the rates are less than 1, the population has vaccination rates lower than would be expected. Panel A shows the distribution across states of age-standardized vaccination rates for each racial/ethnic group, smoothed using the Epanechnikov kernel. Panels B-F show the state-specific, race/ethnicity-specific age-standardized vaccination rates as a function of states' overall Trump vote share. Distributions in grey are state aggregate vaccination rates.

States differ in their handling of Hispanic populations. Some exclude Hispanic individuals from other racial groups; others include Hispanic individuals in other racial groups. The latter states are labeled in Panels B-F with parentheses surrounding their state abbreviation for racial groups other than Hispanic, indicating that those racial groups include Hispanic individuals. Population denominators match states' reported handling of Hispanic individuals in vaccination data.

APPENDIX TABLES

Table S1. Datasets used in analysis

#	Description	Source	Date Downloaded	Citation
1	Vaccine share for each racial/ethnic group	KFF	10/18/2021	https://github.com/KFFData/COVID-19-Data/tree/kff_master/Race%20Ethnicity%20COVID-19%20Data/Vaccines
2	Aggregate vaccination rates for each state	CDC	10/21/2021	https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-rate-total
3	State populations	NCHS	11/3/2021	https://wonder.cdc.gov/controller/saved/D178/D245F115
4	Race-specific, age-specific state populations where Hispanic is a mutually exclusive category	NCHS	10/22/2021	https://wonder.cdc.gov/controller/saved/D178/D241F020
5	Race-specific, age-specific state populations where Hispanic crosscuts racial categories	NCHS	10/22/2021	https://wonder.cdc.gov/controller/saved/D178/D241F020
6	Age-specific national vaccination rates	CDC	10/15/2021	https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-and-Case-Trends-by-Age-Group-/gxj9-t96f
7	State 2020 Trump vote share	MEDSL	9/28/2021	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX

KFF = Kaiser Family Foundation
CDC = Centers for Disease Control and Prevention
NCHS = National Center for Health Statistics
MEDSL = MIT Election Data and Science Lab
Dataset #4 and Dataset #5 are derived from the same underlying NCHS dataset, as cited above, by the research team.

Table S2. Data limitations in state vaccination data (Dataset 1)

State	Any race-specific vaccine data	Racial groups with at least 1% total vaccination share	Hispanic crosscuts racial categories	Percent of vaccinations attributed to "other race"	Percent of vaccinations with unknown race
Alabama	Yes	white, Asian, Black, Hispanic	Yes	11%	11%
Alaska	Yes	white, Asian, Hawaiian/Pacific Islander, Black, Hispanic, Native	Yes	12%	8%
Arizona	Yes	white, Asian, Black, Hispanic, Native	No	18%	11%
Arkansas	Yes	white, Asian, Black, Hispanic	Yes	8%	7%
California	Yes	white, Asian, Hawaiian/Pacific Islander, Black, Hispanic	No	11%	5%
Colorado	Yes	white, Asian, Black, Hispanic, Native	No	4%	10%
Connecticut	Yes	white, Asian, Black, Hispanic	No	7%	5%
Delaware	Yes	white, Asian, Black, Hispanic	Yes	21%	3%
District of Columbia	Yes	white, Asian, Black, Hispanic	Yes	.	21%
Florida	Yes	white, Black, Hispanic	No	13%	13%
Georgia	Yes	white, Asian, Black, Hispanic	Yes	14%	5%
Hawaii	Yes	white, Asian, Hawaiian/Pacific Islander, Black	Yes	3%	19%
Idaho	Yes	white, Asian, Black, Hispanic, Native	Yes	15%	18%
Illinois	Yes	white, Asian, Black, Hispanic	No	4%	4%
Indiana	Yes	white, Asian, Black, Hispanic	Yes	11%	4%
Iowa	Yes	white, Asian, Black, Hispanic	Yes	2%	10%
Kansas	Yes	white, Asian, Black, Hispanic	Yes	19%	9%
Kentucky	Yes	white, Asian, Black	Yes	10%	6%
Louisiana	Yes	white, Asian, Black, Hispanic	Yes	7%	14%
Maine	Yes	white, Asian, Black, Hispanic, Native	No	16%	6%

Maryland	Yes	white, Asian, Black, Hispanic, Native	Yes	10%	5%
Massachusetts	Yes	white, Asian, Black, Hispanic	No	3%	NR
Michigan	Yes	white, Asian, Black, Hispanic, Native	No	9%	16%
Minnesota	Yes	white, Asian, Black, Hispanic, Native	No	1%	7%
Mississippi	Yes	white, Asian, Black, Hispanic	Yes	3%	NR
Missouri	Yes	white, Asian, Black, Hispanic	Yes	.	NR
Montana	No		No	.	NR
Nebraska	No		Yes	.	NR
Nevada	Yes	white, Asian, Black, Hispanic, Native	No	19%	NR
New Hampshire	Yes	white, Asian, Black, Hispanic	No	4%	8%
New Jersey	Yes	white, Asian, Black, Hispanic	No	13%	8%
New Mexico	Yes	white, Asian, Black, Hispanic, Native	No	4%	11%
New York	Yes	white, Asian, Black, Hispanic	Yes	2%	NR
North Carolina	Yes	white, Asian, Black, Hispanic, Native	Yes	8%	6%
North Dakota	No		Yes	.	NR
Ohio	Yes	white, Asian, Black, Hispanic	Yes	9%	4%
Oklahoma	Yes	white, Asian, Black, Hispanic, Native	Yes	5%	19%
Oregon	Yes	white, Asian, Hawaiian/Pacific Islander, Black, Hispanic, Native	No	4%	7%
Pennsylvania	Yes	white, Asian, Black, Hispanic	Yes	14%	10%
Rhode Island	Yes	white, Asian, Black, Hispanic	No	.	NR
South Carolina	Yes	white, Black, Hispanic	Yes	13%	7%
South Dakota	Yes	white, Black, Native	No	3%	9%
Tennessee	Yes	white, Asian, Black, Hispanic	Yes	24%	4%
Texas	Yes	white, Asian, Black, Hispanic	No	15%	6%
Utah	Yes	white, Asian, Hawaiian/Pacific Islander, Black, Hispanic, Native	No	3%	13%

Vermont	Yes	white, Asian, Black, Hispanic	Yes	2%	3%
Virginia	Yes	white, Asian, Black, Hispanic	No	6%	8%
Washington	Yes	white, Asian, Hawaiian/Pacific Islander, Black, Hispanic, Native	No	11%	7%
West Virginia	Yes	white, Black	Yes	8%	6%
Wisconsin	Yes	white, Asian, Black, Hispanic, Native	Yes	9%	5%
Wyoming	No		No	.	NR

NR=Not reported. Source for all data: Kaiser Family Foundation (Table S1's "dataset 1").

Table S3A. State age-standardized vaccination rates (aggregated over race) as a function of Trump vote share

	Main Effects Model			Interaction Model		
	Coef.	**95% CI**	**p**	**Coef.**	**95% CI**	**p**
Trump share	-1.07	(-1.47 to -0.67)	<0.001	-1.07	(-1.41 to -0.72)	<0.001
Race/ethnicity						
white	Ref	-	-	Ref	-	-
Asian/Pacific Islander	0.19	(0.13 to 0.25)	<0.001	0.07	(-0.16 to 0.31)	0.546
Black	-0.11	(-0.17 to -0.05)	0.001	-0.19	(-0.50 to 0.12)	0.221
Hispanic	0.09	(0.00 to 0.18)	0.045	-0.00	(-0.44 to 0.43)	0.993
Native	0.12	(-0.07 to 0.30)	0.203	1.14	(0.03 to 2.24)	0.044
Interaction term						
white x trump share				Ref	-	-
Asian/PI x trump share				0.25	(-0.27 to 0.77)	0.336
Black x trump share				0.18	(-0.42 to 0.77)	0.557
Hispanic x trump share				0.22	(-0.71 to 1.15)	0.639
Native x trump share				-2.02	(-4.05 to 0.01)	0.051

Standard errors are clustered for state. States are unweighted.

Table S3B. State age-standardized vaccination rates (aggregated over race) as a function of Trump vote share, limited to state-race observations where under 10% of racial groups (other than Hispanic) are Hispanic individuals

	Main Effects Model			Interaction Model		
	Coef.	**95% CI**	**p**	**Coef.**	**95% CI**	**p**
Trump share	-1.14	(-1.51 to -0.77)	<0.001	-1.24	(-1.60 to -0.87)	<0.001
Race/ethnicity						
white	Ref	-	-	Ref	-	-
Asian/Pacific Islander	0.18	(0.11 to 0.24)	<0.001	-0.03	(-0.33 to 0.27)	0.852
Black	-0.11	(-0.18 to -0.04)	0.002	-0.18	(-0.52 to 0.16)	0.298
Hispanic	0.08	(-0.02 to 0.17)	0.100	-0.10	(-0.51 to 0.31)	0.626
Native	0.18	(-0.01 to 0.38)	0.067	0.96	(-0.35 to 2.26)	0.146
Interaction term						
white x trump share				Ref	-	-
Asian/PI x trump share				0.42	(-0.23 to 1.07)	0.199
Black x trump share				-0.14	(-0.53 to 0.81)	0.672
Hispanic x trump share				0.39	(-0.51 to 1.29)	0.386
Native x trump share				-1.56	(-4.05 to 0.93)	0.213

Standard errors are clustered for state. States are unweighted. The calculation of the proportion of each state-race observation's population that is also Hispanic is limited to the vaccine-eligible population at the time of data collection (ages 12+).

Table S4A. Race/ethnicity-specific state age-standardized vaccination rates as a function of Trump vote share

		Coefficient for Trump Share		
Model	**Included race/ethnicity**	**Coef.**	**95% CI**	**p-value**
1	white	-1.07	(-1.41 to -0.72)	<0.001
2	Asian/Pacific Islander	-0.82	(-1.40 to -0.23)	0.008
3	Black	-0.89	(-1.50 to -0.28)	0.005
4	Hispanic	-0.85	(-2.43 to 0.74)	0.276
5	Native	-3.09	(-5.30 to -0.87)	0.010

Each model shows state age-standardized vaccination rates as a function of Trump share for a single racial/ethnic group. For example, Model 1 only includes the age-standardized vaccination rates for the white population in each state.

Table S4B. Race/ethnicity-specific state age-standardized vaccination rates as a function of Trump vote share, limited to state-race observations where under 10% of racial groups (other than Hispanic) are Hispanic individuals

		Coefficient for Trump Share		
Model	**Included race/ethnicity**	**Coef.**	**95% CI**	**p-value**
1	white	-1.24	(-1.63 to -0.84)	<0.001
2	Asian/Pacific Islander	-0.82	(-1.40 to -0.23)	0.008
3	Black	-1.10	(-1.72 to -0.47)	0.001
4	Hispanic	-0.85	(-2.43 to 0.74)	0.276
5	Native	-2.80	(-5.38 to -0.22)	0.036

Each model shows state age-standardized vaccination rates as a function of Trump share for a single racial/ethnic group. For example, Model 1 only includes the age-standardized vaccination rates for the white population in each state. The calculation of the proportion of each state-race observation's population that is also Hispanic is limited to the vaccine-eligible population at the time of data collection (ages 12+).