Supplemental Online Content

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*. Published September 2, 2021. doi:10.1001/jama.2021.15161

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods

Sample selection and serologic testing

eTable 1 displays the blood collection organizations that participated in the study, the month participation began, and the number of specimens each contributed to the study. This study is scheduled to continue until the end of December 2021, but this report includes data on specimens collected July 1, 2020–May 31, 2021.

Blood collection organizations selected blood donations using slightly different methods (eTable 2). Six organizations selected donations every week of the month; the remaining eleven selected donations only during certain weeks of the month. Seven blood collection organizations selected blood donations using a pseudorandomization process (84.6% of all specimens). One collected fewer than 2,000 blood donations a month in total and included all samples (0.6% of all specimens). The remaining organizations selecting a nonrandom proportional number of samples on select days spread throughout the month or throughout certain weeks (14.8% of all specimens).

During September–December 2020 in the Southeastern Texas Region, sampling errors resulted in inaccurate Vitros CoV2T results. Because the blood collection organization for the region was independently testing all blood donation specimens using the Elecsys CoV2T, Elecsys CoV2T-based seroprevalence results were used for this study region for these months.

Statistical Methods

Because personal identifiers were removed we could not analyze repeat blood donors longitudinally, so first-time and repeat blood donors were grouped together in seroprevalence estimates. To evaluate the representativeness of the samples relative to the overall donor pool for each geographic region, the demographic composition of the sample and that of the donor pool were compared monthly, and significant deviations were investigated for sampling issues. Statistically significant differences were

determined by applying a two-sample t-test on the unweighted seroprevalence proportions in both months.

It was anticipated that the full geographic regions from which the blood donation samples originated would vary by month. To ensure the same geographic areas were represented each month and hence would remain directly comparable, the blood collection organization-specific geographic regions, referred to as “blood donor regions” were trimmed to smaller regions that captured a large fraction of the monthly donations but remained fixed across months. ZIP codes in which a larger fraction of the population had made blood donations were expected to remain well represented across months, while ZIP codes in which smaller fractions of the population had donated were more likely to be missing in one or more months. Therefore, blood donor regions were created based on the ZIP code distribution of blood donors for 1–3 months prior to study participation (eTable 1, eFigure 2). The blood donor regions were constructed by selecting ZIP codes with the highest relative representation of blood donors compared to the general population, while also maintaining geographically contiguous regions. The specification of the blood donor region boundaries represents a trade-off between capturing as large a fraction of the sample donations as possible while having the full extent of the blood donor regions well represented in all monthly samples. Across all blood donor regions in the study, the average fraction of donors retained in the smaller regions was 92.8%, with a minimum of 84.1% .

After excluding COVID-19 convalescent plasma donations (whom blood collection organizations specifically recruited to be donors based on documented COVID-19 infections and hence if included would have biased seroprevalence estimates), samples of approximately 2,000 specimens from allogeneic donors were compiled monthly from blood donor regions in each blood organization’s catchment area. If less than 2,000 blood donations were collected in that month from a blood donor region, all blood donations were initially included. Blood donor demographic data from 2019 and early 2020 were analyzed to determine which blood donor regions had the highest number of blood donors

from racial and ethnic minority groups. Approximately 4,000 specimens were compiled monthly from blood donor regions with more blood donors from racial and ethnic minority groups, including blood donor regions located in Georgia, southern California, North Carolina, eastern Michigan, eastern Pennsylvania, Maryland, Washington, D.C., southeastern Texas, southern Florida, northeastern Illinois, New Jersey, and southeastern New York. Increasing the number of specimens only from blood donors of racial and ethnic minority groups was not possible.

Following creation of the blood donor regions, a weighting adjustment was used to account for demographic differences between the monthly sample donors and the general population in each blood donor region. Monthly estimation weights were created for the individual sample donations, so that their over- or underrepresentation in different demographic groups relative to the general population was adjusted. Estimation weights account for the number of people in the general population a blood donor represents. Because the number of blood donations selected from a blood donor region was independent of the underlying population size, monthly estimation weights would generally be larger for blood donors residing in blood donor regions with larger population sizes. This approach is sometimes referred to as a "pseudo-design" and is the most commonly used statistical method to estimate general population characteristics based on nonprobability data.[3] The monthly estimation weights were constructed by raking, using the 2018 American Community Survey (ACS) 5-year estimates for the age, gender, race, and ethnicity composition for the blood donor regions. Raking is a statistical calibration technique commonly used in large-scale surveys.[4] In addition to these estimation weights, monthly sets of 50 replicate weights were created to be used to estimate the variance of the weighted seroprevalence estimates.[5] Following the creation of the estimation and replicate weights, SAS version 9.4 (SAS Institute, Cary, NC) was used to compute estimates.

To the extent that the demographic calibration adjustment is sufficient to account for the differences between blood donors and the general population, the weights enable constructing statistically valid seroprevalence estimates for the general population in the blood donor regions.

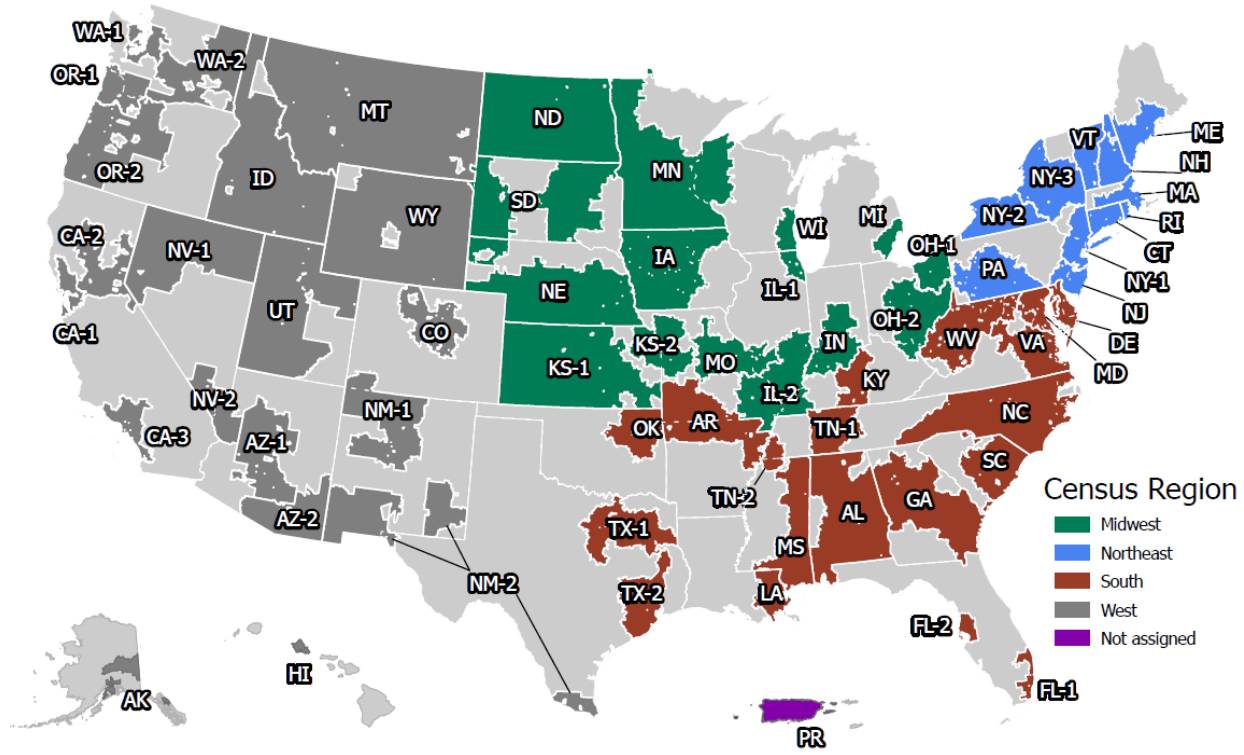
Because the blood donor regions reflect the ZIP codes from which the sampled blood donations originated, these do not necessarily correspond to city or state boundaries, and several blood donor regions overlap. Hence, all ZIP codes covered by the combination of all blood donor regions were divided into 66 nonoverlapping study regions (eFigure 1, eFigure 2, eTable 3). These study region boundaries are based on major metropolitan area or state boundaries. For any study region in which multiple blood donor regions overlap, the weights in the overlapping ZIP codes were adjusted using composite estimation methods for independent samples.[6] This compositing was performed for each overlap area and for each month by applying adjustment factors to the weights proportional to the monthly sample sizes in the contributing blood donor region. The adjustment factors sum to 1 across the contributing blood donor regions for each overlap area.

Donations with detectable S antibodies were presumed to have antibodies because of either infection, vaccination, or both (see eTable 4). The weighted proportion of donations with S antibodies was called the combined seroprevalence. Donations with S and N antibodies were presumed to have antibodies because of infection. Blood donors with S+/N+ blood specimens might have been vaccinated, but this could not be determined without additional information on vaccination history at the time of donations that was not available from all participating blood collection organizations. For blood donations with S antibodies that were not tested for N antibodies, imputation was used to predict the proportion with N antibodies. Among specimens that should have been tested for nucleocapsid antibodies per the study algorithm, 2.6% of specimens had missing nucleocapsid antibody testing results and were imputed.

Weighted S+ (combined) and S+/N+ (infection-induced) seroprevalence rates were estimated for each Census region by combining all study regions within a Census region (eFigure 1, eTable 3). Study regions that crossed Census region borders were assigned to the Census region containing the greatest study region population.

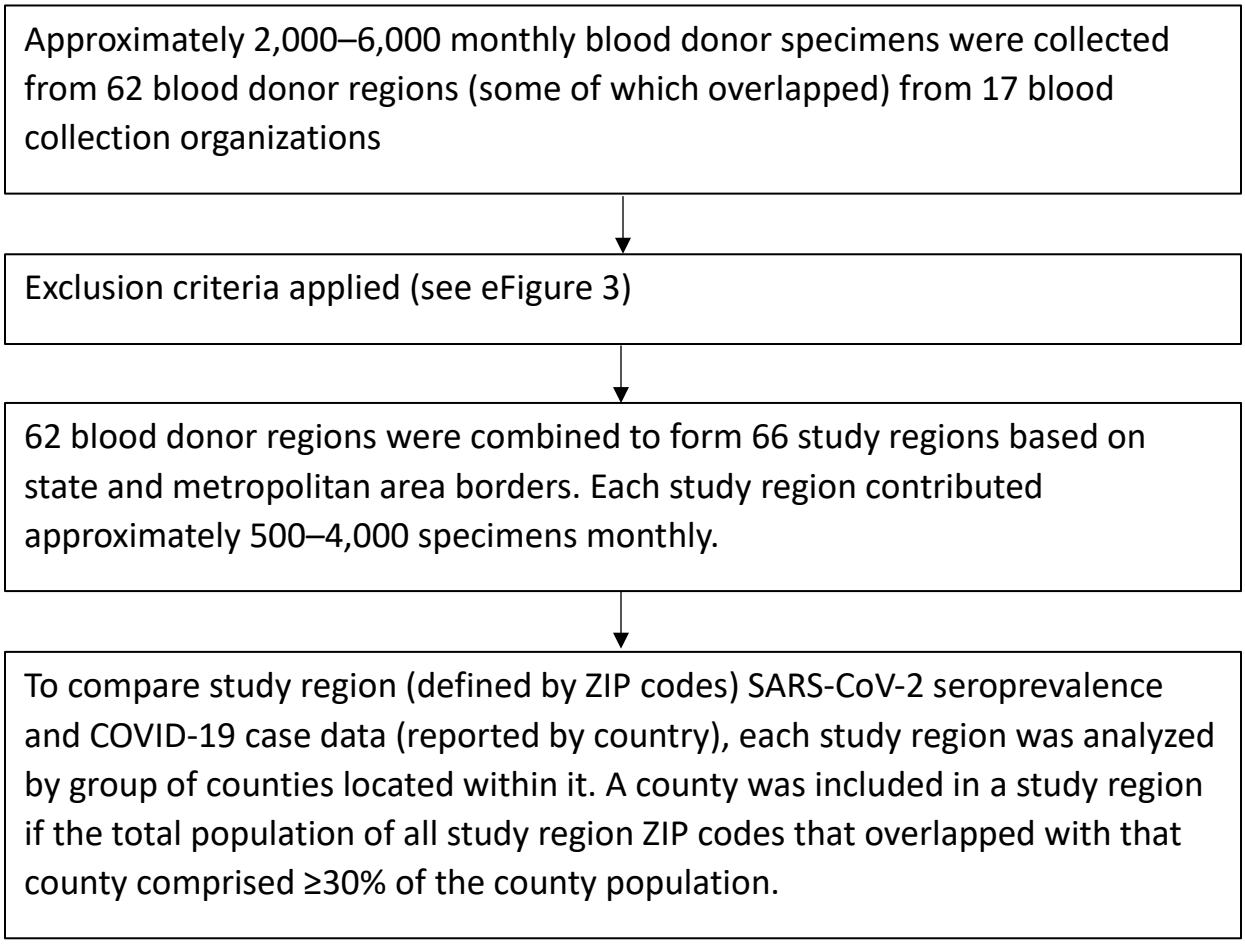
eReferences

1. Centers for Disease Control and Production. U.S. COVID-19 vaccine product information. Accessed March 13, 2021. Available at: https://www.cdc.gov/vaccines/covid-19/info-by-product/index.html.
2. Stone M, Di Germanio C, Wright DJ, et al. Use of U.S. Blood Donors for National Serosurveillance of SARS-CoV-2 Antibodies: Basis for an Expanded National Donor Serosurveillance Program. *Clin Infect Dis*. Jun 10 2021;doi:10.1093/cid/ciab537
3. Elliott MR, Valliant, R. Inference for non-probability samples. *Statistical Science*. 2017;32:249-264.
4. Oh HL, and Scheuren, F.J. . Modified raking ratio estimation. *Survey Methodology*. 1987;13(2):209–219.
5. Rust KF, Rao JN. Variance estimation for complex surveys using replication techniques. *Stat Methods Med Res*. Sep 1996;5(3):283-310. doi:10.1177/096228029600500305
6. Lohr SL, and J. N. K. Rao Inference from dual frame surveys. *Journal of the American Statistical Association*. 2000;95(449):271-280.

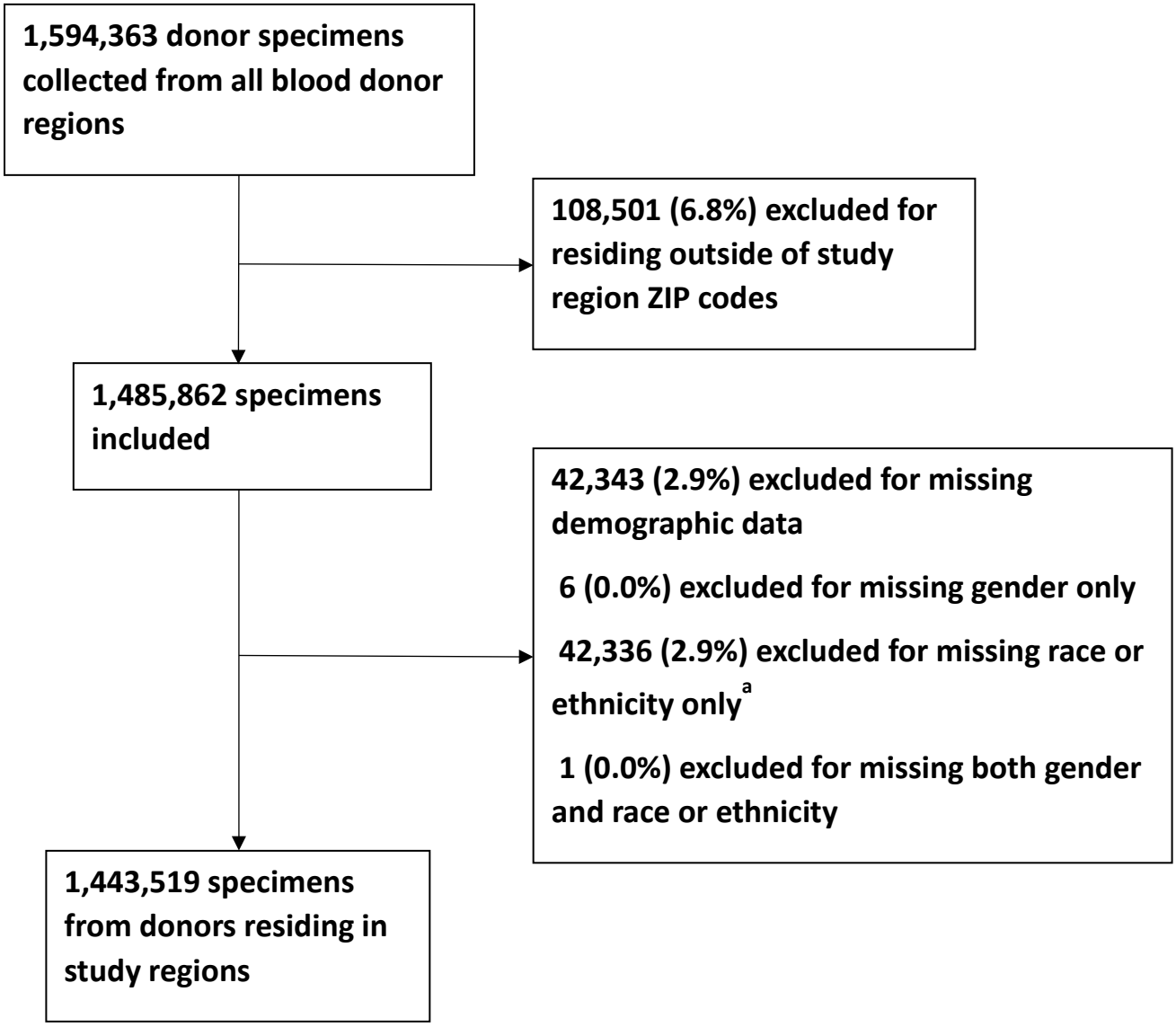


eFigure 1. Map of 66 U.S. study regions. See eTable 3 for the full U.S. study region names. Study regions are geographic regions uniquely defined for this study based on state and metropolitan borders that blood donors consistently reside in. The study regions were constructed by selecting ZIP codes with the highest relative representation of blood donors compared to the general population, while also maintaining geographically contiguous regions.

eFigure 2. Summary flow diagram of methods used to form study regions and compare study region SARS-CoV-2 seroprevalence to county-based COVID-19 case rates.



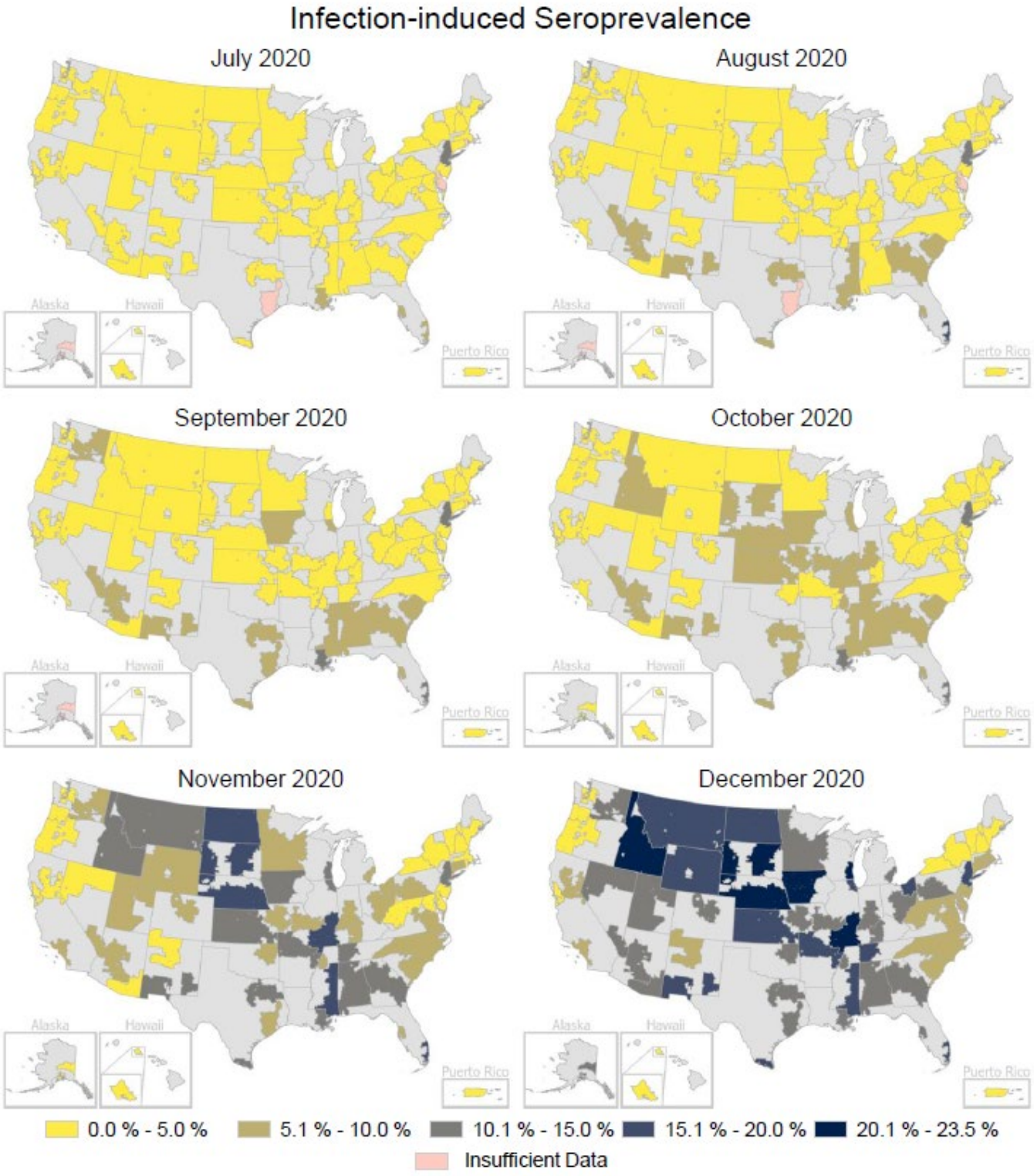
eFigure 3. Flow diagram of study exclusion and inclusion.



[a] The Puerto Rico study region did not collect donor race or ethnicity data, but the donations for this study region were not excluded. Analyses involving race and ethnicity excluded donations from Puerto Rico.

eFigure 4. Weighted spike antibody SARS-CoV-2 seroprevalence by study region, United States, July–December 2020

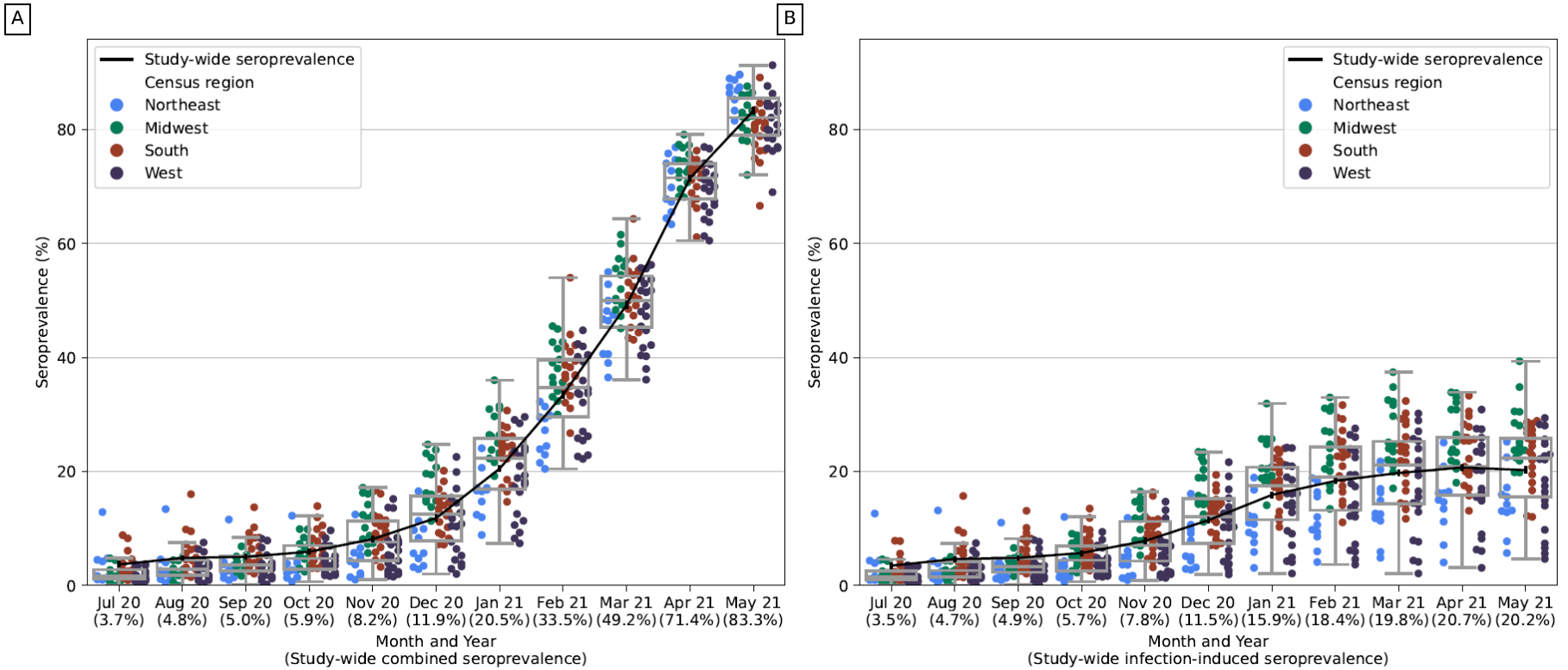
Figure displays data for blood donation specimens collected July—December 2020 from 17 blood collection organizations. Because the number of participating blood collection organizations increased, the number of blood donations included increased from 116,513 in July 2020 to 135,030 in December 2020 (see eTable 1 and eTable 5). Study regions are geographic regions uniquely defined for this study based on state and metropolitan borders that blood donors consistently reside in. The monthly median number of specimens from an individual study region was 1,877 (25th–75th percentile: 1,386–2,251). See eTable 8 in Supplement 4 for detailed data. The maps display the weighted spike antibody seroprevalence by study region and by month, representing the proportion of the population with antibodies from infection, vaccination, or both (labeled combined seroprevalence). During September–December 2020 for the TX-2 study region (see eFigure 1), sampling issues prevented successful spike antibody testing. The blood collection organization for that study region had been performing nucleocapsid antibody testing on all blood specimens, and nucleocapsid seroprevalence is displayed for this study region for these months.



eFigure 5. Weighted SARS-CoV-2 seroprevalence by U.S. Census region and by study region, United States, July 2020–May 2021.

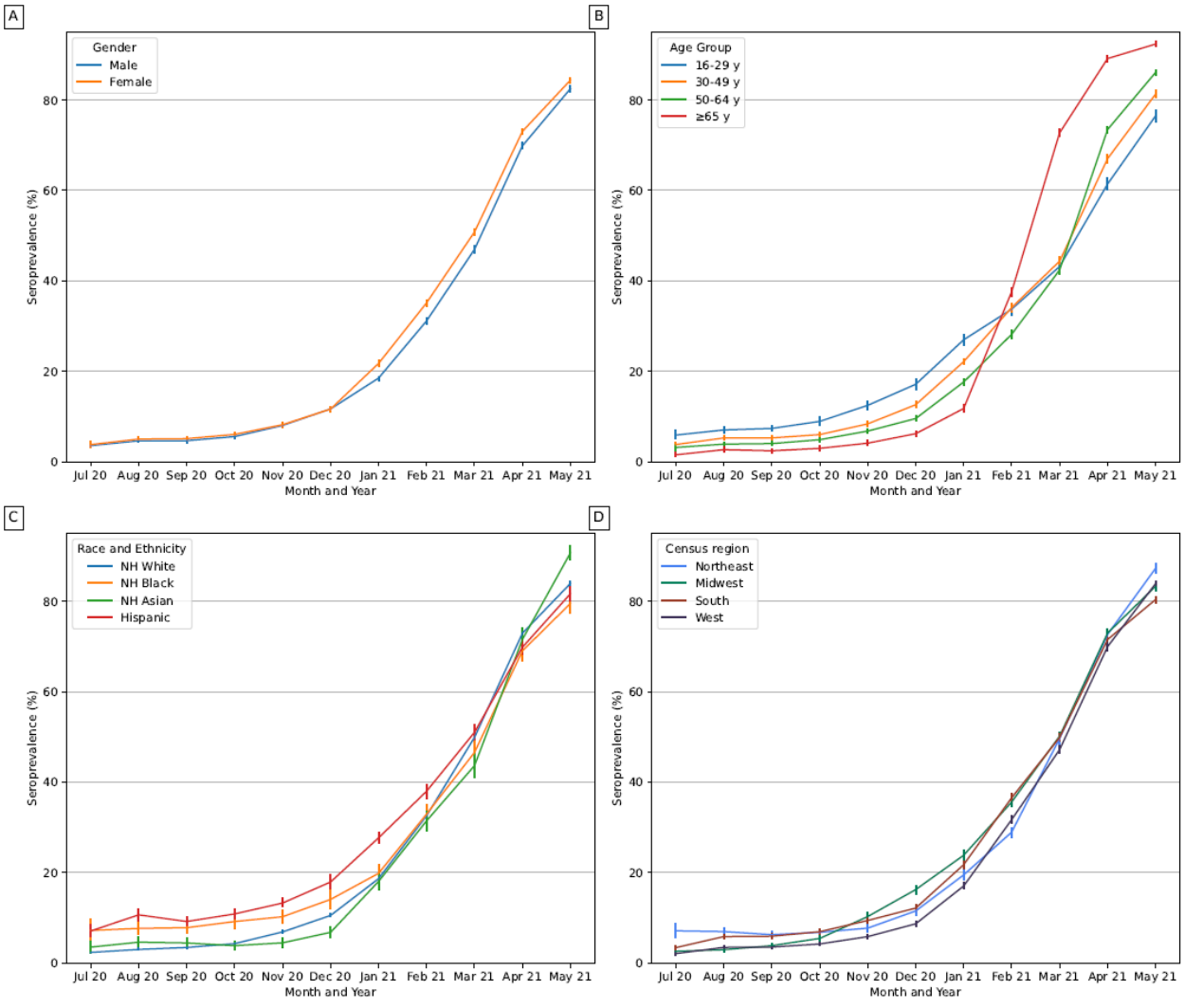
Because the number of participating blood collection organizations increased, the number of blood donations included increased from 116,513 in July 2020 to 135,174 in May 2021 (see eTable 1 and eTable 5). The monthly median number of specimens from an individual study region was 1,878 (25th–75th percentile: 1,393–2,296). See eTable 7 in Supplement 3 and eTable 8 in Supplement 4 for detailed data. Panel A displays the weighted spike antibody seroprevalence by study region and by month, representing the proportion of the population with antibodies from infection, vaccination, or both (combined seroprevalence). Panel B displays the weighted seroprevalence of the population with both spike and nucleocapsid antibodies by study region and by month, representing the proportion of the population with antibodies from infection (infection-induced seroprevalence). The monthly seroprevalence for each study region is represented by an individual dot. The color of the dot represents the Census region in which the study region resides (see eTable 3). The solid black line represents the average seroprevalence estimate of all study regions combined.

1



eFigure 6. Weighted SARS-CoV-2 seroprevalence by gender, by age group, by race and ethnicity, and by Census region, United States, July 2020–May 2021, after restricting to specimens from blood collection organizations that began participation in July 2020.

In this figure, results are restricted to the 10 blood collection organizations that began participation in July 2020, with approximately 116,500 blood donation samples per month combined. Displayed error bars are 95% confidence intervals. The weighted spike antibody seroprevalence, representing the proportion of the population with antibodies from infection, vaccination, or both (combined seroprevalence), is displayed. All blood donation specimens missing race and ethnicity data were excluded except for specimens collected in the Puerto Rico study region, in which donor race and ethnicity data was not collected.



eTable 1. Participating blood collection organizations, month participation began, and total number of specimens contributed to study, United States, July 2020–May 2021.

Blood collection organization	Month and year participation began	Number of specimens contributed to study [a]	Laboratory performing serology [b]	
			Spike	Nucleocapsid
American Red Cross	Jul-20	813,768	CTS	CTS
Banco de Sangre de Servicios Mutuos	Jul-20	21,014	CTS	CTS
Blood Bank of Alaska	Oct-20	11,797	VRI	CTS/TBC
Blood Bank of Delmarva	Sep-20	17,238	RIBC	CTS/TBC
Blood Bank of Hawaii	Jul-20	21,000	CTS	CTS
Bloodworks Northwest	Jul-20	22,244	VRI	CTS/TBC
Carter BloodCare	Jul-20	22,000	CTS	GCBC
Community Blood Center of Kansas City	Sep-20	17,877	VRI	CTS/TBC
Gulf Coast Regional Blood Center	Sep-20	35,997	VRI	GCBC
LifeServe	Sep-20	18,195	VRI	CTS/TBC
LifeSouth Community Blood Centers	Sep-20	18,000	LSO	CTS/TBC
New York Blood Center	Jul-20	40,120	RIBC	CTS/TBC
OneBlood	Jul-20	66,000	CTS	CTS
Rhode Island Blood Center	Aug-20	19,735	RIBC	CTS/TBC
The Blood Center	Jul-20	21,876	VRI	TBC
Versiti	Jul-20	42,464	VRI	VST/CTS/TBC
Vitalant	Jul-20	385,038	CTS	CTS

[a] Prior to applying exclusion criteria (see eFigure 3).

[b] CTS= Creative Testing Systems, VRI = Vitalant Research Institute, TBC = The Blood Centers, GCBC = Gulf Coast Regional Blood Center, LSO = LifeSouth Community Blood Centers, RIBC = Rhode Island Blood Center, VST = Versiti.

eTable 2. Method of blood donation specimen selection by blood collection organization.

Blood collection organization	Sampling Dates	Sampling Frequency	Donations Sampled[a]	Selection Process[b]
American Red Cross	First 3 weeks	Monthly	74,000	Random
Banco de Sangre de Servicios Mutuos	First 3 weeks	Su, M, Tu, W	200	Random
Blood Bank of Alaska	All days	Daily	All	Exhaustive
Blood Bank of Delmarva	First 3 weeks	Tu, W, Th	150–200	Convenience
Blood Bank of Hawaii	First 3 weeks	Su, M, Tu, W	200	Random
Blood Center of Kansas City	First 3 weeks	Weekday	~100	Convenience
Bloodworks Northwest	All days	Daily	N/A	Convenience
Carter Blood Care	First 3 weeks	Su, M, Tu, W	200	Random
Gulf Coast Blood Center	Weeks 2 and 4	Weekday	400	Convenience
LifeServe	First 3 weeks	Weekday	~130	Convenience
LifeSouth	All days	Weekly	500	Random
New York Blood Center	All days	Daily	200	Convenience
OneBlood	All days	Monthly	6,000	Random
Rhode Island Blood Center	All days	Daily	100	Convenience
The Blood Center	First 3 weeks	Weekday	N/A	Convenience
Versiti	First 3 weeks	M, Tu, W	150–200	Convenience
Vitalant	First 3 weeks	Monthly	36,000	Random

N/A = Not applicable

[a] Per sampling frequency

[b] Random selection: pseudorandom number generator-based algorithmic selection. Convenience sample selection: selection of proportional samples among daily or weekly eligible donations. Exhaustive sampling: selection of all eligible donations because of insufficient availability of specimens (when the number of donations is lower than the target sampling number).

eTable 3. Full descriptive names of each study region[a], United States, July 2020–May 2021.

Abbreviation	Full descriptive name	Census region
AK	Alaska Region	West
AL	Alabama Region	South
AR	Southern Missouri and Northern Arkansas Region	South
AZ-1	Central and Northern Arizona Region	West
AZ-2	Southern Arizona Region	West
CA-1	Northern California–Bay Region	West
CA-2	Northern California–Sacramento Region	West
CA-3	Southern California Region	West
CO	Colorado Region	West
CT	Connecticut Region	Northeast
DE	Delaware and Maryland Eastern Shore Region	South
FL-1	Miami Region	South
FL-2	Tampa Bay Region	South
GA	Georgia Region	South
HI	Hawaii Region	West
IA	Iowa Region	Midwest
ID	Idaho Region	West
IL-2	Southern Illinois, Southeastern Missouri, and Western Kentucky Region	Midwest
IN	Central and Southern Indiana Region	Midwest
KS-1	Kansas Region	Midwest
KS-2	Northeastern Kansas and Northwestern Missouri Region	Midwest
KY	Central Kentucky Region	South
LA	Southeastern Louisiana Region	South
MA	Massachusetts Region	Northeast
MD	Central Maryland and National Capital Region	South
ME	Maine Region	Northeast
MI	Detroit Region	Midwest
MN	Minnesota Region	Midwest
MO	Central Missouri and St. Louis Region	Midwest
MS	Eastern Mississippi Region	South
MT	Montana Region	West
NC	Central and Western North Carolina Region	South
ND	North Dakota Region	Midwest
NE	Nebraska Region	Midwest
NH	New Hampshire Region	Northeast
NJ	Southern New Jersey and Southeastern Pennsylvania Region	Northeast
NM-1	Northern New Mexico Region	West
NM-2	Southern New Mexico, El Paso, and Southern Texas Region	West
NV-1	Northern Nevada Region	West
NV-2	Southern Nevada and Northwestern Arizona Region	West

NY-1	Southeastern New York and Northern New Jersey Region	Northeast
NY-2	Western New York Region	Northeast
NY-3	Northern New York Region	Northeast
OH-1	Northeastern Ohio Region	Midwest
OH-2	Central and Southern Ohio Region	Midwest
OK	Northeastern Oklahoma Region	South
OR-1	Northwestern Oregon and Southwestern Washington Region	West
OR-2	Central and Western Oregon Region	West
PA	Central and Western Pennsylvania Region	Northeast
PR	Puerto Rico Region	Not applicable
RI	Rhode Island Region	Northeast
SC	Central and Eastern South Carolina Region	South
SD	South Dakota Region	Midwest
TN-1	Central Tennessee Region	South
TN-2	Memphis Region	South
TX-1	Northeastern Texas Region	South
TX-2	Southeastern Texas Region	South
UT	Utah Region	West
VA	Eastern Virginia Region	South
VT	Vermont Region	Northeast
WA-1	Western Washington Region	West
WA-2	Eastern Washington Region	West
WI	Southeastern Wisconsin Region	Midwest
WV	West Virginia, Western Maryland, and Northern Virginia Region	South
WY	Wyoming Region	West

[a] Study regions are geographic regions uniquely defined for this study based on state and metropolitan borders that blood donors consistently reside in.

eTable 4. Interpretation of antibody results when COVID-19 vaccination status is unknown

Spike antibody	Nucleocapsid antibody	Interpretation
+	+	Previously infected, unknown if vaccinated
+	-	Vaccinated with no previous infection
-	n/a[a]	Not previously vaccinated or infected

n/a=not applicable

[a]Blood donations without detectable S antibodies were not tested for nucleocapsid antibodies.

eTable 5. Number of blood donations included and median collection dates of included blood specimens by month, July 2020–May 2021.

Month and year	No. blood donations included	Median collection date
July 2020	116,513	07/15/20
August 2020	123,200	08/12/20
September 2020	131,052	09/14/20
October 2020	131,731	10/13/20
November 2020	134,061	11/13/20
December 2020	135,026	12/12/20
January 2021	134,632	01/14/21
February 2021	133,288	02/12/21
March 2021	134,557	03/15/21
April 2021	134,510	04/14/21
May 2021	134,949	05/14/21

eTable 6. Number of projected cumulative SARS-CoV-2 infections with detectable antibodies based on infection-induced seroprevalence per reported COVID-19 case by study region, United States, July 2020–May 2021. [a]

[a] See spreadsheet in Supplement 2. Study regions are geographic regions uniquely defined for this study based on state and metropolitan borders that blood donors consistently reside in. Number of projected cumulative SARS-CoV-2 infections per reported case is the infection-induced seroprevalence estimate divided by the cumulative reported cases per 100 persons. The 95% CI is the 95% CI of the seroprevalence estimate divided by the cumulative reported cases per 100 persons. Infection-induced seroprevalence is the seroprevalence of the population with both spike and nucleocapsid antibodies.

CI: Confidence interval

eTable 7. Weighted infection-induced SARS-CoV-2 seroprevalence by study region, and within study region, by age group, by gender, and by race and ethnicity, United States, July 2020–May 2021. [a]

[a] See attached spreadsheet in Supplement 3. Study regions are geographic regions uniquely defined for this study based on state and metropolitan borders that blood donors consistently reside in. Infection-induced seroprevalence is the seroprevalence of the population with both spike and nucleocapsid antibodies. Sample totals do not include specimens that should have been tested for nucleocapsid antibodies per the study algorithm but had missing results. Results for these specimens were imputed.

CI: Confidence interval

eTable 8. Weighted combined infection- and vaccine-induced SARS-CoV-2 seroprevalence by study region, and within study region, by age group, by gender, and by race and ethnicity, United States, July 2020–May 2021. [a]

[a] See attached spreadsheet in Supplement 4. Study regions are geographic regions uniquely defined for this study based on state and metropolitan borders that blood donors consistently reside in. Combined infection- and vaccine-induced SARS-CoV-2 seroprevalence is the seroprevalence of the population with spike antibodies.

CI: Confidence interval