

April 20, 2021

Imputing Missing Race and Ethnicity Data in COVID-19 Cases

Health equity is a cornerstone of public health. This is especially true during the COVID-19 pandemic when many populations are suffering from both the health and economic consequences of the disease. Accurate information on disparities in disease incidence, outcomes, and social and economic impacts is necessary to guide and develop an appropriate response. Efforts to study these disparities have been hampered by missing data. Almost a quarter of confirmed cases are missing race and ethnicity data. Accounting for this missing data is essential to understanding COVID-19 and to facilitate research into health disparities.

The methodology to impute missing race and ethnicity, which relies on surname and neighborhood demographics, was [published](#) along with the results of the first assessment. This data is intended for research purposes and to assist understanding of COVID-19-related health disparities. The COVID-19 Daily Dashboard will continue to report the unimputed data, including the number of cases with race and ethnicity not reported. Virginia Department of Health, Surveillance and Investigations staff continue to [pursue multiple strategies](#) to fill in missing data. Data presented here uses the department's [tiered race and ethnicity](#) framework.

Results

As of April 1st, 2021, when data was pulled for this analysis, 16.8% of cases were missing race or ethnicity data. The greater majority of missing cases were imputed as White. In 94,159 records with imputed data, 74.6% were White followed by Hispanic which was 15.5%. Notably, 3.6% of Asian Americans and Pacific Islander cases were imputed in the missing dataset. In the final dataset, White was the only race or ethnicity that saw the share of cases attributed to it, from 54.6% to 57.9%.

Virginia April 1st, 2021						
Race	Reported Race		Imputed Race		Final Data	
	Reported Race (Count)	Reported Race Share	Imputed Race (Count)	Imputed Race Share	Final Data (Count)	Final Data Share
American Indian or Alaska Native	646	0.1%	4	0%	650	0.1%
Asian or Pacific Islander	22,641	5%	3,398	4%	26,039	5%
Black	107,660	23%	5,968	6%	113,628	20%
Latino	77,702	17%	14,574	16%	92,276	16%
Two or More Races	3,517	1%	8	0%	3,525	1%
White	255,099	55%	70,207	75%	325,306	58%
Total	467,265	100%	94,159	100%	561,424	100%

Regional Results

Northern Virginia had the highest share of cases imputed (26.8%) followed by Eastern (22.7%), Southwest (18.3%), Central (17.4%) and Northwest (14.8%). In Central Virginia, the share of cases attributed to White residents increased from 48.5% to 53.2% in terms of race completeness, while the share for Hispanic residents decreased slightly from 10.5 % to 10.2 %. In Northern Virginia, the share of cases attributed to White residents increased, from 35.2% to 37.8% completeness.

Central Health Region

Central April 1st, 2021		Reported Race		Imputed Race		Final Data	
Race	Reported Race (Count)	Reported Race Share	Imputed Race (Count)	Imputed Race Share	Final Data (Count)	Final Data Share	
American Indian or Alaska Native	135	0.2%	2	0%	137	0.1%	
Asian or Pacific Islander	2,139	3%	342	2%	2,481	3%	
Black	30,807	37%	2,008	12%	32,815	33%	
Latino	8,681	10%	1,416	9%	10,097	10%	
Two or More Races	653	1%	2	0%	655	1%	
White	39,905	48%	12,656	77%	52,561	53%	
Total	82,320	100%	16,426	100%	98,746	100%	

Eastern Health Region

Eastern April 1st, 2021		Reported Race		Imputed Race		Final Data	
Race	Reported Race (Count)	Reported Race Share	Imputed Race (Count)	Imputed Race Share	Final Data (Count)	Final Data Share	
American Indian or Alaska Native	152	0.1%	1	0%	153	0.1%	
Asian or Pacific Islander	2,800	3%	375	2%	3,175	3%	
Black	40,992	40%	2,532	12%	43,524	35%	
Latino	6,278	6%	1,648	8%	7,926	6%	
Two or More Races	794	1%	5	0%	799	1%	
White	51,619	50%	16,782	79%	68,401	55%	
Total	102,635	100%	21,343	100%	123,978	100%	

Northern Health Region

Northern Region February 28 th , 2021						
Race	Reported Race		Imputed Race		Final Race	
	Reported Race (Count)	Reported Race Share	Imputed Race (Count)	Imputed Race Share	Final Data (Count)	Final Data Share
American Indian or Alaska Native	167	0.2%	12	0%	179	0.1%
Asian or Pacific Islander	11,408	11%	3,536	11%	14,944	11%
Black	12,878	12%	1,264	4%	14,142	10%
Latino	43,871	42%	12,748	40%	56,619	42%
Two or More Races	803	1%	13	0%	816	1%
White	34,719	33%	14,012	44%	48,731	36%
Total	103,846	100%	31,585	100%	135,431	100.0

Northwest Health Region

Northwest April 1st, 2021						
Race	Reported Race		Imputed Race		Final Data	
	Reported Race (Count)	Reported Race Share	Imputed Race (Count)	Imputed Race Share	Final Data (Count)	Final Data Share
American Indian or Alaska Native	68	0.1%	0	0%	68	0.1%
Asian or Pacific Islander	1,389	2%	137	1%	1,526	2%
Black	8,785	12%	149	1%	8,934	10%
Latino	10,380	14%	1,836	13%	12,216	14%
Two or More Races	506	1%	0	0%	506	1%
White	52,330	71%	11,812	85%	64,142	73%
Total	73,458	100%	13,934	100%	87,392	100%

Southwest Health Region

Southwest April 1st, 2021						
Race	Reported Race		Imputed Race		Final Data	
	Reported Race (Count)	Reported Race Share	Imputed Race (Count)	Imputed Race Share	Final Data (Count)	Final Data Share
American Indian or Alaska Native	90	0.1%	0	0%	90	0.1%
Asian or Pacific Islander	980	1%	140	1%	1,120	1%
Black	10,962	13%	496	3%	11,458	11%
Latino	3,529	4%	307	2%	3,836	4%
Two or More Races	471	1%	1	0%	472	1%
White	66,684	81%	16,287	94%	82,971	83%
Total	82,716	100%	17,231	100%	99,947	100%