

III. THE COVID RACIAL DATA TRACKER EXPERIENCE



This section lays out the deficiencies of existing national COVID-19 datasets that necessitated the CRDT team’s work, describes the CRDT methodology, details the specific challenges that the CRDT team faced, and summarizes the CRDT experience through an antiracist lens.

A. The Need for the CRDT

Federal data repositories with COVID-19 race and ethnicity information on cases, deaths, and hospitalizations have been insufficient for evidence-based policy making. To begin with, no high quality national dataset exists for COVID-19 cases by race and ethnicity. For much of 2020, even the federal government was relying on the CRDT for COVID-19 race and ethnicity case data.³⁵ National case surveillance datasets from the U.S. Centers for Disease Control and Prevention³⁶ (CDC) have been documented as highly incomplete throughout the pandemic.³⁷ For example, the Satcher Institute found that in May 2020, the CDC’s COVID-19 Case Surveillance Restricted Access dataset had race and ethnicity information for only 43% of cases, and that by April 2021, that proportion had increased to just 65%.³⁸ Put another way, as of April 2021, 8.6 million out of 24.4 million cases were missing race and ethnicity information.³⁹ The percentage of COVID-19 cases with known race and ethnicity did not improve through October 2021.⁴⁰

This lack of comprehensive racial and ethnic data for COVID-19 case reports is largely due to the fact that data reporting to the CDC’s National Notifiable Disease Surveillance System (NNDSS) is voluntary.⁴¹ Many local and state jurisdictions fail to provide all COVID-19 case reports to the CDC,⁴² and even for COVID-19 case reports that *are* provided to NNDSS, race and ethnicity are not mandatory fields in all jurisdictions.⁴³ Whether race and ethnicity are mandatory data fields

and how highly they are prioritized varies by jurisdiction, and this is not a problem unique to COVID-19.⁴⁴ As a result, case reports to NNDSS can be, and are, filed with missing race and ethnicity information. Compounding this problem is that individual jurisdictions are not consistent in their reporting to the NNDSS. Voluntary case reporting varies by state and over time, and some states have reported less and less data as the pandemic continues.

Racial and ethnic data concerning COVID-19 deaths, which are collected by the National Center for Health Statistics (NCHS) through the National Vital Statistics System (NVSS),⁴⁵ are superior to the data currently collected through the NNDSS regarding cases. In 2020, for instance, the NCHS had race and ethnicity data for over 99% of all deaths nationwide.⁴⁶ “NCHS has legislative authority and is mandated under 42 U.S.C. § 242k, Section 306(h) of the Public Health Service Act to collect vital statistics,” which includes births, deaths, marriages, and divorces.⁴⁷ The completeness of this dataset is due to a “cooperative relationship between the states and the federal government,” which is facilitated by the National Association for Public Health Statistics and Information Systems (NAPHSIS).⁴⁸ The NVSS is not without its own flaws, however. The data are limited in usefulness because they only concern deaths, and not cases or hospitalizations. Additionally, while the dataset is close to complete across states, it has gaps at the county and finer geographic levels. For example, due to the suppression of small counts for confidentiality reasons, some counties with small numbers of deaths attributed to certain conditions or racial groups are excluded. These data are also collected at a much slower pace, making it less than ideal for disease surveillance. In the CRDT team’s experience, NCHS data were backlogged by approximately six weeks during the data collection period—a delay too lengthy to allow the team to draw useful conclusions in real time.

High quality COVID-19 hospitalization data by race and ethnicity are also unavailable from any national data source. The CDC’s COVID-NET system collects data from hospitals in select counties in just fourteen states.⁴⁹ A newer surveillance system established by HHS in response to the pandemic requires facility-level daily reports of COVID-19-related metrics from all U.S. hospitals to a national tracking system directly maintained by HHS instead of the CDC.⁵⁰ While this reporting system is mandatory, it does not ask hospitals for any demographics aside from age brackets.⁵¹ It notably does not include any racial or ethnic demographics in the facility-level information requested.⁵²

Federal agencies have publicly acknowledged the need for a more robust nationwide system of reporting and collecting COVID-19 race and ethnicity data.⁵³ The U.S. Government Accountability Office (GAO) issued reports in September 2020 and March 2021 commenting on the federal response to the COVID-19 pandemic. In its September report, the GAO called on the CDC to “determine whether having the authority to require states and jurisdictions to report race and ethnicity

information for COVID-19 cases, hospitalizations, and deaths is necessary for ensuring more complete data, and if so, seek such authority from Congress.”⁵⁴ The CDC responded that “it was conducting an analysis to determine whether additional authorities given to the agency to mandate the collection of race and ethnicity information could enhance the robustness and completeness of data shared with the agency.”⁵⁵ In March 2021, the GAO reiterated that federal race and ethnicity data continued to be limited, and added a recommendation for the CDC to collect race and ethnicity data on COVID-19 vaccinations.⁵⁶ The CDC agreed with this recommendation.⁵⁷ However, the CDC reported race and ethnicity data for vaccinations at the national level only, with no geographic breakdown by state or county, and race and ethnicity remained unknown for more than 25% of vaccine recipients as of February 2022.⁵⁸

B. CRDT Methodology

The CRDT team collected all publicly available racial and ethnic demographic data on COVID-19 outcomes from U.S. states and territories. To do this, the CRDT enlisted hundreds of volunteers who collected data twice weekly from April 12, 2020 to March 7, 2021 for COVID-19 cases, deaths, and tests; and from June 17, 2020 to March 7, 2021 for hospitalizations (April 2020 to March 2021 is hereinafter referred to as the “data collection period”). The CRDT data were reported cumulatively, and have been publicly available since the start of the project.⁵⁹

The CRDT data came from governmental websites, dashboards, reports, press releases, and other online sources. The team did not use any public information requests or similar tools to collect data that were not otherwise publicly available online. This Report focuses on CRDT data from the fifty states and District of Columbia (hereinafter referred to as “states” or “jurisdictions”), as the CRDT was not able to obtain meaningful data from the other U.S. territories.⁶⁰

The CRDT classified race and ethnicity according to the OMB categories, and included separate racial categories for individuals with “Multiple Races” or “Other Race.” The most recent (1997) racial and ethnic demographic data categories used by the OMB are “American Indian or Alaska Native (AIAN),” “Asian,” “Black or African American,” “Native Hawaiian or Other Pacific Islander (NHPI),” or “white,” and the OMB ethnic categories are “Hispanic” and “Not Hispanic.” We discuss the OMB further in Section VI.C.

C. The Challenges of Obtaining COVID-19 Data by Race and Ethnicity

The major challenges that the CRDT faced included missing and incomplete data, inconsistencies and deficiencies in reporting practices, inconsistencies and

deficiencies in the treatment of the Hispanic/Latino/a/e/x and multiracial groups, non-standard racial and ethnic categories, failure to include additional information beyond OMB categories, and infrequent data updates. Each of these challenges is described in detail below.

1. Data Were Missing and Incomplete across Jurisdictions and over Time

In the first year of the pandemic, the CRDT was the most comprehensive and up-to-date source of COVID-19 race and ethnicity data available, but its data completeness depended on the completeness of the data it collected from the states. In order to assess the quality of its inputs, the CRDT team separately kept track of each state’s data completeness by determining whether, at a minimum, each state reported some data on race (for this purpose only, the team did not track ethnicity).⁶¹

During the CRDT collection period, all states technically reported some COVID-19 case data by race except New York, but states varied greatly in the completeness of the data reported. For example, Texas technically reported race data, but race was only known for 3% of its cases statewide. COVID-19 death data by race were reported by all states (with varying degrees of completeness) except for North Dakota. Hospitalization data by race were less widely reported—more than half the states failed to report any racial data for COVID-19 hospitalizations.⁶² The proportions of COVID-19 outcomes with known race across states at the end of the CRDT data collection period are shown in Figures 1, 2, and 3. COVID-19 outcomes that were reported without race information included instances of “missing”⁶³ data and “reported unknowns.”⁶⁴ Some states also reported a number of cases as “pending,” based on positive antigen testing, with racial data unavailable for such cases.

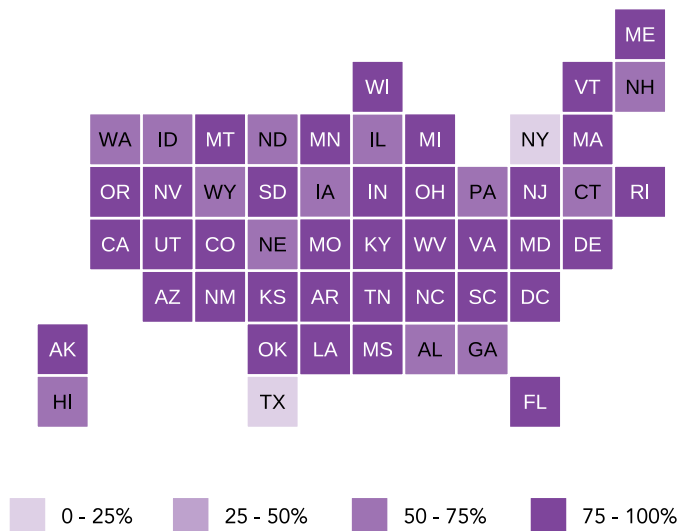


Figure 1: Percent of Cases with Known Race

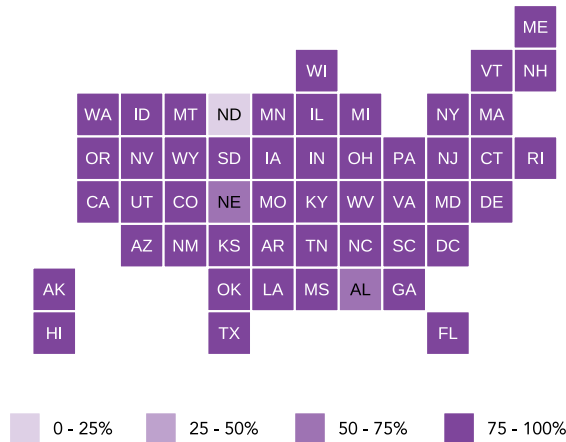


Figure 2: Percent of Deaths with Known Race

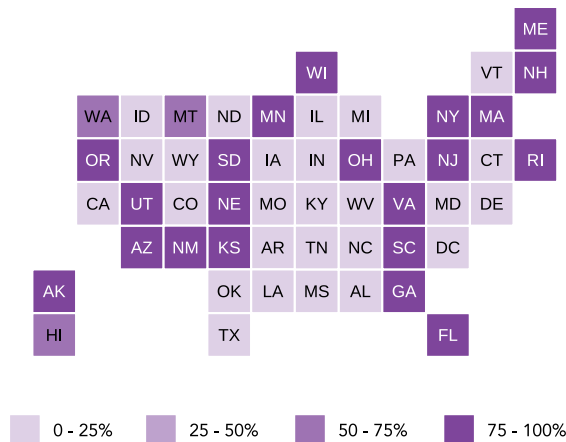


Figure 3: Percent of Hospitalizations with Known Race

States also varied greatly in terms of when they *first* began reporting COVID-19 racial data.⁶⁵ A plurality of states began reporting some racial data for COVID-19 outcomes in April 2020. Forty-three states and the District of Columbia reported racial data for cases by the end of that month; thirty-seven states and the District of Columbia reported racial data for deaths by the end of that month.⁶⁶ However, data completeness was insufficient to draw nationwide conclusions at that time, because race remained unknown for 63.5% of cumulative cases and 32.5% of cumulative deaths through the end of April 2020. The CRDT team began collecting COVID-19 hospitalization data in June 2020, two months after it began collecting case and death data. Over the course of that month, sixteen states reported racial data for hospitalizations.⁶⁷ By this time, data completeness had improved for deaths, as race was unknown for just 7% of cumulative deaths nationwide. But data completeness remained problematic for tracking nationwide trends in cases (as race remained unknown for 44.4% of cases) and hospitalizations (as race remained unknown for over 50%⁶⁸ of hospitalizations).

The amount of missing data showed gradual improvement over time for all three outcomes (cases, hospitalizations, and deaths), but remained far from complete by the end of the collection period. By August 2020, all states except New York were reporting racial data for cases, and by September 2020, all states except North Dakota were reporting racial data for deaths. Hospitalization data remained the most incomplete, with a total of twenty-three states reporting racial data for hospitalizations by December 2020.⁶⁹ Nationwide, through the end of February 2021, race was unknown for 5% of cumulative deaths and 33.2% of cumulative cases. A percentage of nationwide hospitalizations with unknown race cannot be calculated because approximately one-third of the states did not report cumulative total counts of patients ever hospitalized. Restricting to the states that reported cumulative hospitalizations, race was unknown for over 26%⁷⁰ of cumulative hospitalizations through the end of the collection period.

Finally, data on testing remained woefully underreported during the collection period. As shown in Figure 4, COVID-19 testing data were reported by race and ethnicity in only nine states,⁷¹ and definitions of testing were not consistent across these states. Testing was defined by three states (California, Illinois, and Missouri) in terms of the number of specimens tested, including repeated tests on the same individual. Six other states (Nevada, Utah, Indiana, Kansas, Delaware, and Rhode Island), by contrast, reported this information in terms of the number of individual people tested. The CRDT could not determine how the inclusion of repeated tests on the same individuals in California, Illinois, and Missouri might impact the racial and ethnic distribution of COVID-19 testing data. This made data comparisons between states challenging and interfered with the team’s ability to draw regional or national conclusions regarding testing.

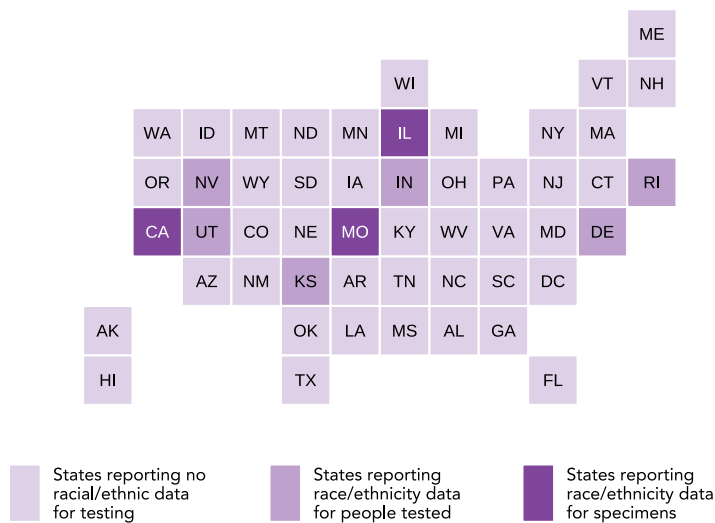


Figure 4: Type of Covid-19 Testing Data Reported

2. States’ Data Reporting Conventions Were Deficient and Inconsistent

The CRDT team also encountered inconsistencies in the ways states reported their race and ethnicity data that led to erasure of important information. The preferred reporting practice, employed by some states, is to make disaggregated exact counts (the raw numbers) publicly available for each racial and ethnic category, so that researchers and policymakers can interpret and analyze the data easily. However, some states only reported race and ethnicity information as a percentage of total COVID-19 cases.

Colorado and Iowa are two states that reported information about race and ethnicity as percentages, and their methods are useful as case studies. These examples demonstrate how providing percentages instead of exact counts can either introduce minor rounding errors or seriously obscure racial disparities, depending on the amount of additional information provided alongside those percentages.

a) Example: Rounding Errors in Colorado

Colorado’s convention for reporting, while not ideal, introduced only minor rounding errors. Colorado reported percentages to two decimal places on its state COVID-19 dashboard, as shown in Image 1, a screenshot taken from the state’s website. The Colorado dashboard’s “Tooltips” section specified that these percentages were calculated with a denominator of “All Cases.” The state provides this denominator as an exact count (559,704) in a different section of the dashboard. This additional information allowed users of the data to calculate counts within each racial or ethnic group. For example, a user could multiply .57% by 559,704 to obtain, within a small margin of error, the number of COVID-19 cases among the AIAN population in Colorado (approximately 3,190 ±27 cases).⁷² So, while it is not ideal that Colorado did not report exact counts of COVID-19 outcomes within each group, counts could at least be estimated within a narrow range using the percentages reported to two decimal places alongside the additional raw numbers provided.⁷³

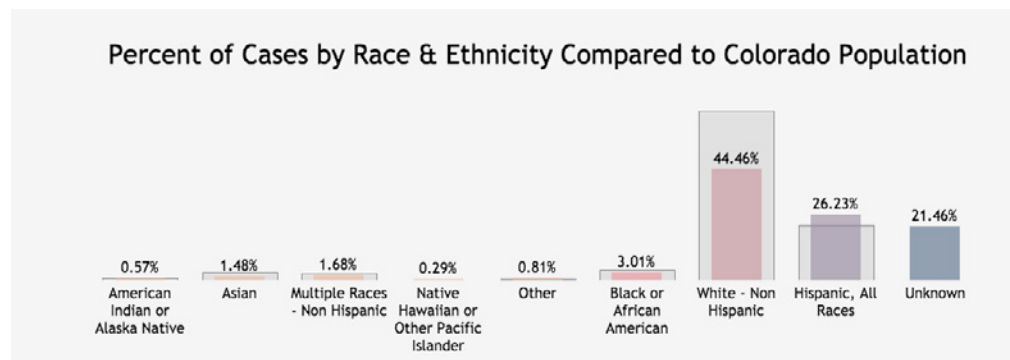


Image 1: Colorado’s Rounding Convention for COVID-19 Race and Ethnicity Data, 2021

This screenshot was taken from the Colorado state dashboard (<https://covid19.colorado.gov/data>) on July 6, 2021.

b) Example: Rounding Errors in Iowa

Iowa’s convention for reporting race and ethnicity information obscured the extent of racial and ethnic inequities due to two specific practices. Unlike Colorado, Iowa reported case percentages rounded to the nearest whole number (which is far from precise), as shown in Image 2, a screenshot taken from the state’s website. Iowa also used positive tests, as opposed to positive cases, as the denominator for these percentages, forcing the CRDT team to make several inferences about the data that potentially introduced additional errors.

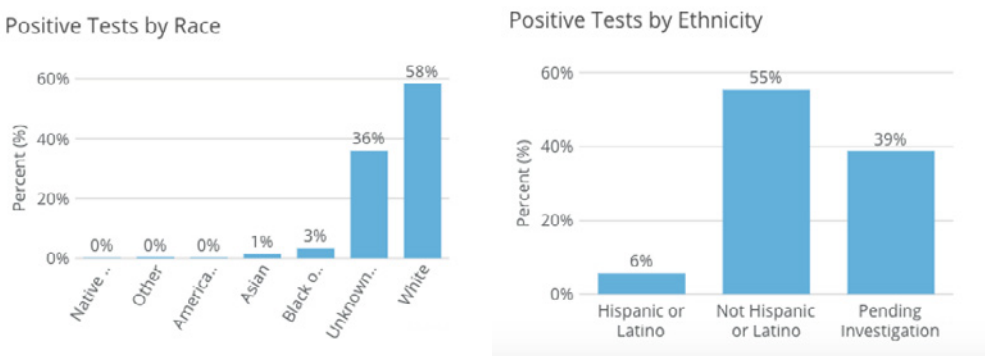


Image 2: Iowa’s Rounding Convention for COVID-19 Race and Ethnicity Data, 2021

These screenshots were taken from the Iowa state dashboard (<https://coronavirus.iowa.gov/pages/case-counts>) on July 6, 2021.

First, Iowa’s choice of denominator was problematic. Iowa’s website dashboard (as shown in Image 2) presented race and ethnicity data as a percentage of “Positive Tests” rather than as a percentage of “cases.”⁷⁴ The problem with this practice is that the measure of “Positive Tests” includes repeated tests for the same individual, obscuring the number of unique individuals who contracted COVID-19. Most states, by contrast, reported racial data in terms of cases (what Iowa termed “Individuals Positive”). So, for consistency with other states, the CRDT team applied the percentages shown in the race- and ethnicity-specific bar charts in Image 2 to the total PCR and Antigen Individuals Positive information provided elsewhere on the Iowa website dashboard (308,623 + 65,632 = 374,255 cases), instead of the total Positive Tests. In doing so, the CRDT team had to assume that the percentage distribution by race and ethnicity is the same for Positive Tests and Individuals Positive. This assumption may not be accurate, introducing an unknown amount of error to the CRDT dataset.

Iowa’s use of percentages rounded to the nearest whole number was also problematic. The CRDT team attempted to estimate the number of cases in each race and ethnicity category by applying the percentages shown in Image 2 to the total count of “Individuals Positive.” Because the percentages were rounded to the nearest whole number, however, a wide range of other case counts could also have been true and have resulted in the same reported percentages. For example, as illustrated in Table 1, 0% of cases occurring in a group could refer to 0 actual cases or as many as 1,871 cases.

Table 1. Case-Study of Impact from Reporting Whole Number Percentages (Iowa)

Race		State-reported percent of cases*	Estimated # of cases (calculated based on N=374,255* statewide cases)	Lowest possible # of cases (calculated**)	Highest possible # of cases (calculated**)	Population***	Estimated case rate per 1,000	Range of possible case rates per 1,000	Estimated case rate ratio (vs. White reference group)	Range of possible case rate ratios (vs. White reference group)
Race	Native Hawaiian or Other Pacific Islander (NHPI)	0.00	0	0	1871	3729	0	0 to 501.7	0	0.0 to 6.6
	Other	0.00	0	0	1871	105633	0	0 to 17.7	0	0.0 to 0.2
	American Indian or Alaska Native (AIAN)	0.00	0	0	1871	11976	0	0 to 156.2	0	0.0 to 2.1
	Asian	0.01	3743	1872	5613	75741	49.4	24.7 to 74.1	0.6	0.3 to 1.0
	Black or African American	0.03	11228	9357	13098	116359	96.5	80.4 to 112.6	1.3	1.0 to 1.5
	Unknown	0.36	134732	132861	136602	N/A	N/A	N/A	N/A	N/A
	White	0.58	217068	215197	218938	2826070	76.8	76.1 to 77.5	1	1
Ethnicity	Hispanic or Latino	0.06	22455	20584	24325	188311	119.2	109.3 to 129.2	1.6	1.4 to 1.7
	Not Hispanic or Latino	0.55	205840	203969	207710	2951197	69.7	69.1 to 70.4	0.9	0.89 to 0.93
	Pending Investigation	0.39	145959	144088	147829	N/A	N/A	N/A	N/A	N/A

Note:

* Data from Iowa state COVID-19 dashboard 7/6/2021 <https://coronavirus.iowa.gov/pages/case-counts>

** Lowest and highest whole numbers that round to the state-reported percent of cases when divided by total statewide cases (N=374,255). For example: $0/374,255 * 100 = 0.0000\%$ and $1,871/374,255 * 100 = 0.4999\%$, making 1,871 the highest possible whole number of cases that would round to the state-reported value of 0%.

*** Iowa state population by race and ethnicity from US Census, 2019 American Community Survey, 5-year estimates, Table B02001 for race and Table B03002 for ethnicity. 'Other' race category includes Census categories 'Some other race alone' and 'Two or more races.'

The lack of precision in Iowa’s methods is particularly problematic for understanding the impact of COVID-19 in groups with small populations. For example, Iowa’s entire statewide population of NHPI people is estimated at 3,729 people, according to U.S. Census data.⁷⁵ Based on the range of possible case counts shown in Table 1 (0 to 1,871), the case rate among NHPI people in Iowa ranged from 0 to 501.7 cases per 1,000 people. This means that somewhere from 0% to 50.17% of NHPI people living in Iowa contracted COVID-19—a range so broad it provides no meaningful information.

The comparison of Iowa's data on NHPI people to white people further shows how the simple practice of reporting percentages rounded to the nearest whole number results in erasure of potentially large racial disparities. To measure the impact of racial disparities, the CRDT team used case rate ratios, with the white population as the reference group. Table 1 shows that white people in Iowa, with a state population of 2,826,070, had a possible case rate ranging from 76.1 to 77.5 cases per 1,000 people—a narrow range that gives a clear sense of magnitude of the true case rate. The NHPI/white case rate ratio would be estimated at 0 based on the data provided by the state, but with possible values ranging from 0 to 6.6. In other words, people in the NHPI group may have been more than six times as likely as white people to have contracted COVID-19, a very large disparity. However, because this range (0–6.6) spans over the value 1.0 (a ratio of 1.0 indicating no disparity is present), it is also within the range of possibility that NHPI people experienced no disparity in case rates, or that white people are in fact more likely than NHPI people to have contracted COVID-19. Based on the way Iowa reports its case information, it is impossible to know which scenario is accurate.⁷⁶ Thus the practice by Iowa of reporting percentages rounded to the nearest whole number obscures the existence and extent of racial disparities.

3. States' Conventions Concerning the Hispanic/Latino/a/e/x Category Were Deficient and Inconsistent

The CRDT encountered additional challenges regarding states' conventions concerning the Hispanic/Latino/a/e/x category, which is often treated as an ethnicity for purposes of data collection. Modeled after the OMB categories, state forms and records frequently collect race and ethnicity information in two separate questions. The first question asks if the individual is of Hispanic or Latino/a/e/x ethnicity, and second asks the individual to select one or more races (Black, white, etc.). As a result, a person may, for example, check boxes for both Hispanic/Latino/a/e/x (ethnicity) and Black (race), and would be included in both counts. States varied in how they treated the nuances of race and ethnicity, making the reporting of COVID outcomes by the Hispanic/Latino/a/e/x category inconsistent across jurisdictions. The four main ways the states treated the Hispanic/Latino/a/e/x category are described below.

First, some states, such as Kansas, reported COVID-19 data by race and ethnicity as two separate measures, as shown in Image 3, a screenshot taken from the state's website. In other words, they reported separate numbers for each race and separate numbers for those who fell within the Hispanic/Latino/a/e/x ethnicity, with no information about how those groups were connected. This reporting structure does not provide or allow disaggregation of the number of cases among white, Black, AIAN, NHPI, or Asian people who are also of Hispanic/Latino/a/e/x ethnicity versus those of non-Hispanic/Latino/a/e/x ethnicity. This convention thus makes comparisons of COVID-19 outcomes across racial and ethnic categories difficult.

Race Case Rates per 100,000

Race	Number of Cases	Rate per 100,000
White	228,561	8,981.77
Black or African American	16,334	7,492.63
American Indian or Alaska Nat..	2,256	3,481.43
Asian	6,131	5,712.98
Other Race	13,846	15,039.70
Not Reported/Missing	50,523	

Image 3: Kansas’s Display of Separate COVID-19 Race and Ethnicity Data

These screenshots were taken from the Kansas state dashboard (<https://www.coronavirus.kdheks.gov/160/COVID-19-in-Kansas>) on June 29, 2021. Data regarding deaths and hospitalizations (not shown) follow the same reporting structure.

Ethnicity Case Rates per 1,000

Ethnicity	Number of Cases	Rate per 1,000
Hispanic/Latino	42,642	122.38
Not Hispanic/Latino	202,195	78.89
Unknown or Missing	72,814	

Second, other states, such as Connecticut, reclassified race and ethnicity into a combined measure, as shown in Image 4, a table downloaded from the state’s website. This means those states counted individuals who responded affirmatively to the ethnicity question (Hispanic/Latino/a/e/x) as one separate group (hence, a combined racial and ethnic group), and counted only those who were part of the non-Hispanic/Latino/a/e/x group in each of the other race categories (Black, white, etc.). This approach allows for some comparisons of COVID-19 outcomes between racial and ethnic categories.

Race/ethnicity	Total cases	Total deaths
Hispanic	50902	879
NH American Indian or Alaskan Native	407	3
NH Asian or Pacific Islander	4076	92
NH Black	25122	967
NH Multiracial	25842	433
NH Other	18471	46
NH White	109807	5722
Unknown	114674	134

Image 4: Connecticut’s Display of Combined COVID-19 Race and Ethnicity Data

Note that “NH” stands for “Not Hispanic.” These data were downloaded from the Connecticut state dashboard (<https://data.ct.gov/Health-and-Human-Services/COVID-19-Cases-and-Deaths-by-Race-Ethnicity/7rne-efic/data>) on June 29, 2021.

Third, a handful of states reported race and ethnicity as intersecting measures. Florida, for example, provided information about only three race categories (white, Black, and other), but did effectively specify the number of people who were Hispanic, Non-Hispanic, and of Unknown ethnicity *within* each of those race categories, as shown in Image 5, a table downloaded from the state’s website. While Florida’s decision to omit several racial categories is ill-advised, its treatment of the intersection of race and ethnicity for those categories that it *did* report provides the most information to researchers, allowing easier comparisons across racial and ethnic categories.

Race and ethnicity	Cases	Hospitalizations	Deaths
White	1,311,318	62,796	27,029
Hispanic	497,861	17,336	6,950
Non-Hispanic	722,095	43,691	18,934
Unknown	91,362	1,769	1,145
Black	306,204	19,903	5,978
Hispanic	18,334	858	242
Non-Hispanic	265,368	18,665	5,555
Unknown	22,502	380	181
Other race	360,091	10,541	2,964
Hispanic	181,982	5,561	1,339
Non-Hispanic	124,780	4,499	1,397
Unknown	53,329	481	228
Unknown race	308,719	2,367	1,002
Hispanic	50,226	517	192
Non-Hispanic	19,145	263	104
Unknown	239,348	1,587	706
Total	2,286,332	95,607	36,973

Image 5: Florida’s Display of Intersecting COVID-19 Race and Ethnicity Data

This table was downloaded from the Florida state dashboard on June 29, 2021 with data updated through June 2, 2021, available at http://ww11.doh.state.fl.us/comm/_partners/covid19_report_archive/cases-monitoring-and-pui-information/state-report/

Hospitalization counts include anyone who was hospitalized at some point during their illness. It does not reflect the number of people currently hospitalized. **Other race** includes any person with a race of American Indian/Alaskan native, Asian, native Hawaiian/Pacific Islander, or other.

Fourth, a few states, such as North Dakota, omitted Hispanic/Latino/a/e/x ethnicity information from their reporting altogether and only reported race data, as shown in Image 6, a screenshot taken from the state’s website. This method is clearly inadequate, as it completely omits any information about those who fall in the Hispanic/Latino/a/e/x category.

Actual Race	Selected Calculation Demographic
White	68,358
Unknown	30,243
American Indian	5,068
Black	3,387
2 or More	2,118
Asian	1,462
Other	44

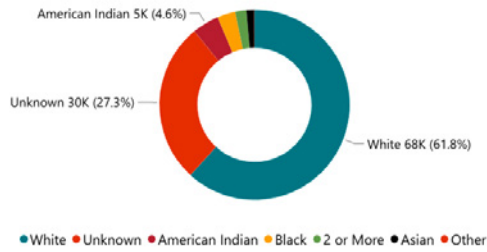


Image 6: North Dakota's Display of COVID-19 Race Data Only

These screenshots were taken from the North Dakota state dashboard (<https://www.health.nd.gov/diseases-conditions/coronavirus/north-dakota-coronavirus-cases>) on June 29, 2021.

Figure 5 displays which states employed each of the four methods described above.⁷⁷

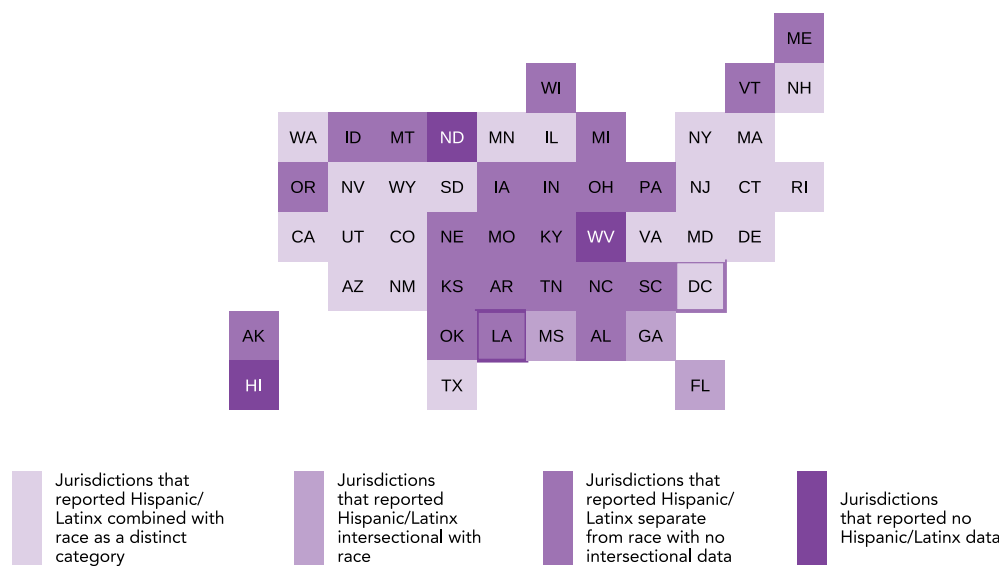


Figure 5: Method of Race and Hispanic/Latino/a/e/x Ethnicity Reporting

The inconsistency in methods of reporting data about the Hispanic/Latino/a/e/x category across the states made racial and ethnic disparities difficult to measure for the United States as a whole and led to the potential underestimation of disparities in some states. Florida's data from Image 5 above can be used as a case study to show how different classification methods may result in underestimation of racial and ethnic disparities.⁷⁸ In Tables 2 and 3 below, we reconfigured Florida's case data from Image 5, presented it in the two other ways states used to report information about the Hispanic/Latino/a/e/x group (race and ethnicity as two separate measures, and race/ethnicity combined), and calculated the resulting racial/ethnic disparities, to demonstrate the impact of each type of reporting.⁷⁹

Table 2. Florida Case Data by Race and Ethnicity, Reported Separately

		Cases	Population*	Cases per 1000 pop.	Case rate ratio (vs. White)
Race	White	1311318	15702256	83.5	1
	Black	306204	3359031	91.2	1.09
	Other Race	360091	1840349	195.7	2.34
	Unknown Race	308719	N/A	N/A	N/A
Ethnicity	Hispanic	748403	5346684	140	1.68
	Non-Hispanic	1311388	15554952	72.7	0.87
	Unknown Ethnicity	406541	N/A	N/A	N/A

Note:

* Population data from US Census American Community Survey 5-year estimates.

Table 3. Florida Case Data by Race and Ethnicity, Combined

Race/Ethnicity combined	Cases	Population*	Cases per 1000 pop.	Case rate ratio (vs. White)
Hispanic (any race)	748403	5346684	140	1.94
White (not Hispanic)	813457	11266347	72.2	1
Black (not Hispanic)	287870	3202687	89.9	1.24
Other race/ethnicity (not Hispanic)	178109	2085918	164	2.27
Unknown race/ethnicity (not Hispanic)	258493	N/A	N/A	N/A

Note:

* Population data from US Census American Community Survey 5-year estimates.

Tables 2 and 3 reveal several differences in the magnitude of case rates and case rate ratios when race and ethnicity are reported separately versus combined. First, the case rate among white people appears *higher* when race and ethnicity are reported separately instead of combined (83.5 cases per 1,000 versus 72.2 cases per 1,000). This occurs because, when race and ethnicity are reported separately, the white race category includes some people of Hispanic/Latino/a/e/x ethnicity. But when race and ethnicity are reported as a combined measure, people of Hispanic/Latino/a/e/x ethnicity are excluded from the white race category, resulting in a lower case rate for white people.

Second, the case rate of white people as compared to other racial and ethnic groups changes depending on whether race and ethnicity are reported separately or combined. The Hispanic/Latino/a/e/x/white case rate ratio appears lower when race and ethnicity are reported separately instead of combined (1.68 times as likely as white people to have contracted COVID-19 versus 1.94 times as likely as white people to have contracted COVID-19).⁸⁰ Because the measure of white cases per

1,000 is higher when race and ethnicity are reported separately, the Hispanic/Latino/a/e/x case rate ratio appears lower. The Black/white case rate ratio also appears lower when race and ethnicity are reported separately instead of combined (1.09 times as likely as white people to have contracted COVID-19 versus 1.24 times as likely as white people to have contracted COVID-19).⁸¹ Once again, because white cases per 1,000 appear higher when race and ethnicity are reported separately, the Black case rate ratio appears lower.

Case rate ratios are a key measure of racial and ethnic disparities. As demonstrated in Figure 6, the magnitude of racial and ethnic disparities may vary based on state reporting practices. The case study of Florida’s data shows that the magnitude of racial and ethnic disparities appears lower for both Hispanic/Latino/a/e/x and Black people when insufficient information is provided about how race and ethnicity intersect. This erasure is likely an issue in other states’ data as well, and may become even more pronounced when calculating disparities for multiple states or for the United States as a whole.

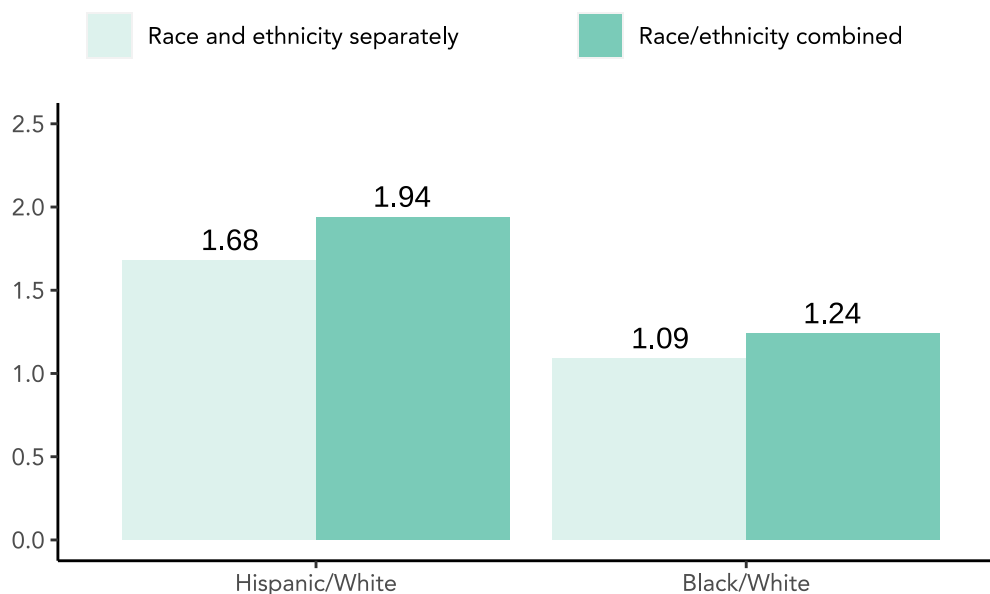


Figure 6: Florida Hispanic/White and Black/White Case Rate Ratios by Reporting Structure

These case studies demonstrate how a comprehensively detailed approach for reporting the intersections of race and ethnicity data, as used by states like Florida, is the method least likely to result in errors. Disaggregating race by ethnicity and presenting race and ethnicity as intersecting measures allows for the most clarity and flexibility (short of providing a detailed case-level dataset). Researchers and analysts can further aggregate measures as desired under this method. By contrast, they cannot disaggregate race and ethnicity or make comparisons to states using a different aggregation method when race and ethnicity are reported using the other three methods described above.

4. States’ Conventions Concerning the Multiracial Category Were Deficient and Inconsistent

The CRDT team also observed a variety of methods for reporting information about multiracial people, some of which introduced errors by counting multiracial people as two or more separate people. Various institutions’ forms often allow multiracial people to select more than one race category. Most states reported COVID-19 data by reclassifying those who selected two or more races into a separate multiracial category, or in a combined category with “other race.” However, two states, Utah and Wyoming, reported outcomes for multiracial individuals in each of the race categories selected, meaning that such individuals were double counted. For example, if an individual selected both “Black” and “white” for race, they were included in both counts. While it is important to obtain data concerning exactly which racial and ethnic categories multiracial people fall into, treating multiracial people as two (or more) separate people misrepresents such information.

Because data from Utah and Wyoming were reported in aggregate, it was not possible to discern exactly how many people in each category represented double-counted multiracial people and reclassify the information described above into the “Multi Race” or “Other Race” categories employed by most states. Utah acknowledged this methodological choice, but did not address the problem that this method poses for measuring racial disparities.⁸² Wyoming, by contrast, did not state its practice explicitly, but the double counting of multiracial individuals can be inferred because percentages in its reported data totaled more than 100%. For example, Image 7, a table downloaded from the state’s website, showed percentages of COVID-19 cases by race and ethnicity summing to over 108%. Below, we explore Utah and Wyoming’s approaches more closely.

NH White	62.78	
Hispanic	9.29	
American Indian	4.76	
Black	1.11	
Asian	0.62	
Native Hawaiian/	0.29	
Other	4.29	
Unknown	25.40	
	108.53	

Image 7: Wyoming’s Display of Percentage of COVID-19 Cases by Race and Ethnicity

This table was downloaded from the Wyoming state dashboard (<https://sites.google.com/wyo.gov/covid-19/home>) on July 7, 2021.

a) Case Study: Double Counting Can Lead to Underestimation of Racial Inequities

To demonstrate how Utah and Wyoming’s classification schemes may result in underestimation of racial disparities, we use a hypothetical example. For simplicity, the only racial categories in this example are Black, white, or multiracial (Black and white). The total number of COVID-19 cases statewide is 1,090. The practice by most states of reporting separate race categories would present the sample data shown in Table 4.

Race	Cases
White alone	1000
Black alone	70
Multiracial	20
Total	1090

Table 4: Case Counts by Race (Hypothetical Data) - Version 1

Reclassifying the same data, but with multiracial people counted in *both* the Black and white categories, would result in this alternate version of the same table, as shown in Table 5.

Race	Cases
White, alone or in combination	1020
Black, alone or in combination	90
Total	1110

Table 5: Case Counts by Race (Hypothetical Data) - Version 2

The number of cases appears higher in each race category in the second version, as compared to the first. Additionally, in the second version, the total value of 1,110 exceeds the actual statewide total number cases, which is 1,090.

Combining these tables with sample population data,⁸³ Tables 6 and 7 demonstrate how the choice of classification scheme impacts the measurement of racial disparities.

Table 6. Case Rates and Case Rate Ratio by Race (Hypothetical Data) – Version 1

Race	Cases	Population	Cases per 1000 pop.	Case rate ratio (vs. White)
White alone	1000	1000000	1.0	1.0
Black alone	70	28000	2.5	2.5
Multiracial	20	10000	2.0	2.0
Total	1090			

Table 7. Case Rates and Case Rate Ratio by Race (Hypothetical Data) – Version 2

Race	Cases	Population	Cases per 1000 pop.	Case rate ratio (vs. White)
White, alone or in combination	1020	1010000	1.01	1.0
Black, alone or in combination	90	38000	2.37	2.3
Total	1110			

In this hypothetical, the Black/white disparity in case rates appears smaller in magnitude (2.5 in Table 6/Version 1 versus 2.3 in Table 7/Version 2) when the same data are reclassified using categories that double count multiracial people. As a result, double counting individuals can result in underestimation of racial disparities. The actual extent to which this issue may have obscured racial or ethnic disparities in Utah and Wyoming cannot be determined from the limited information reported by those states and captured by CRDT.

5. States Omitted or Lumped Together Racial and Ethnic Categories

The CRDT team also encountered high variation in the racial and ethnic categories that were used by states, which often precluded state-by-state comparisons of disparities and made it more difficult to understand disparities at the national level. The OMB categories were not required for the state-reported COVID-19 outcome data that the CRDT collected, but they were treated as default categories by most states. By the end of the CRDT data collection period, several states reported COVID-19 outcomes disaggregated for all the OMB racial categories, as listed in Table 8. However, many states did not even follow the OMB categories, either by failing to include some of the categories or reporting categories in aggregate with other groups. These practices contributed to undercounting certain racial groups.

Below, we describe the major ways that the states departed from the OMB and varied from each other in their treatment of racial categories.

Cases	Deaths	Hospitalizations	Tests
Alaska	Alaska	Alaska	California
Arkansas	Arkansas	Maine	Missouri
California	California	Minnesota	Rhode Island
Colorado	Colorado	Ohio	
Kentucky	Kentucky	Oregon	
Maine	Maine	Rhode Island	
Minnesota	Minnesota	Washington	
Missouri	Missouri		
New Hampshire	New Hampshire		
Ohio	Ohio		
Oregon	Oregon		
Rhode Island	Rhode Island		
Washington	Washington		

Table 8: States Reporting Disaggregated Data for All Federal Standard Race Categories, March 7, 2021

a) States Failed to Include Some OMB Categories in Reporting

Some states, such as Alabama, completely omitted some of the OMB racial categories in their reporting. Alabama only reported COVID-19 cases and deaths

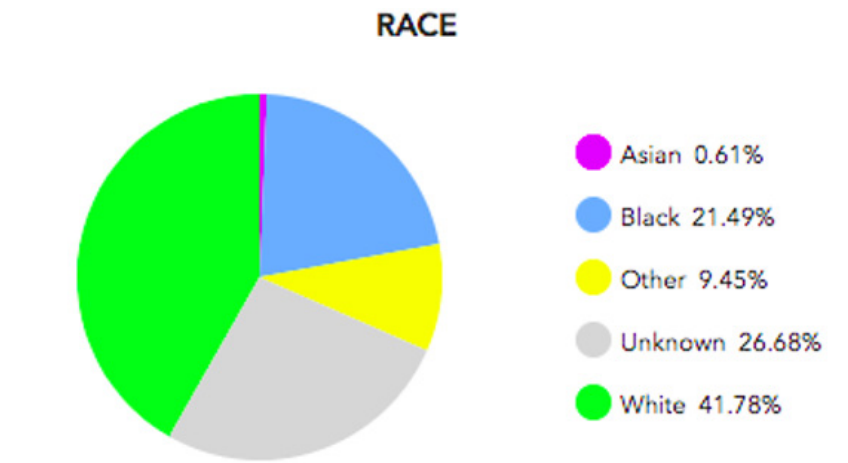


Image 8: Alabama’s Display of COVID-19 Race Information

This screenshot was taken from the Alabama state dashboard (<https://alpublichealth.maps.arcgis.com/apps/dashboards/6d2771faa9da4a2786a509d82c8cf0f7>).

for the Asian, Black, Other, Unknown, and white races, as shown in Image 8, a screenshot taken from the state’s website.⁸⁴

The state did not make any mention of the OMB categories for AIAN or NHPI. No data notes are provided to specify whether these categories are included under “Other” or if there were zero cases among members of these groups.

b) States Reported Standard Categories in Aggregate with Other Groups

Some states combined the existing OMB categories into larger groups. This practice took several forms, the most common of which was to use a combined “Asian or Pacific Islander” category instead of the disaggregated “Asian” and “Native Hawaiian or Other Pacific Islander” categories.

Montana grouped “American Indian or Alaska Native” and “Native Hawaiian or Other Pacific Islander” into a combined category labeled “American Indian, Alaska Native, Native Hawaiian, Pacific Islander.” Montana was the only state to aggregate these groups in COVID-19 reporting, though it later began disaggregating AIAN and NHPI on February 5, 2021.

South Carolina adopted another nonstandard practice, reporting an aggregated category labeled “Asian, Alaskan, Hawaiian.” No disaggregated data were reported by this state for the Asian, AIAN, or NHPI categories.

Finally, a number of states put one or more of the OMB categories into a combined “Other” category. Indiana, for example, reported a category labeled “Other Race” with a note specifying that “[o]ther races included American Indian, Alaska Native, Native Hawaiian, Other Pacific Islander, two or more races, and those reported to

ISDH as other race.”⁸⁵ A number of other states reported a category labeled “Other” without providing any notes to specify which race categories “Other” included. In the example of Alabama (shown above in Image 8), several OMB categories that were not reported were most likely included in the “Other” category, but the state websites did not say this explicitly.

Taken as a whole, the result of the practices described in Sections (a) and (b) above was an undercount of national totals for some racial categories. The two race categories most affected by this issue were the two smallest by population: AIAN and NHPI. Figures 7 through 10 show the states that did not report disaggregated data for these groups in state-level reporting for COVID-19 cases or deaths.

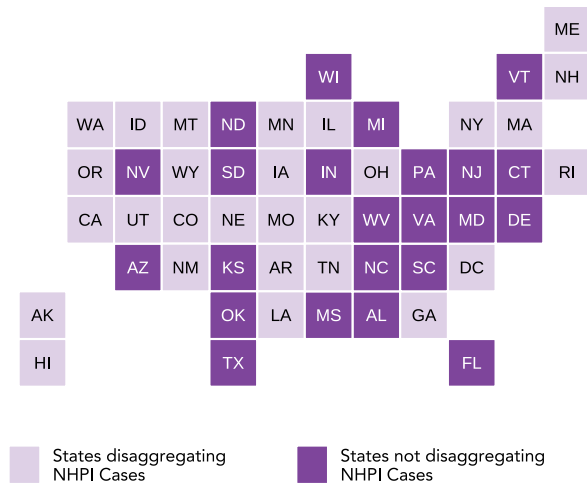


Figure 7: States Reporting with Non-Standard Race Categories for NHPI Cases

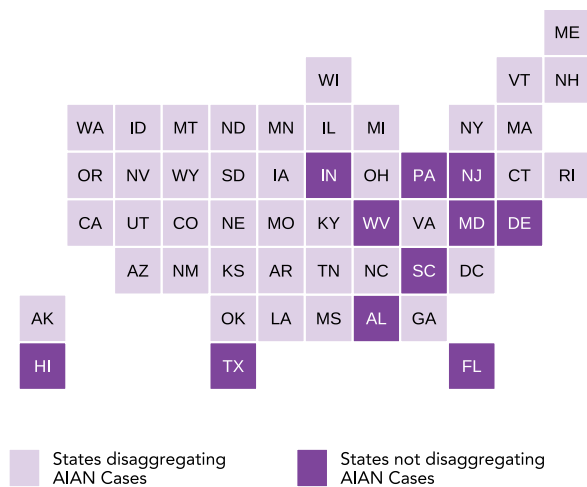


Figure 8: States Reporting with Non-Standard Race Categories for AIAN Cases

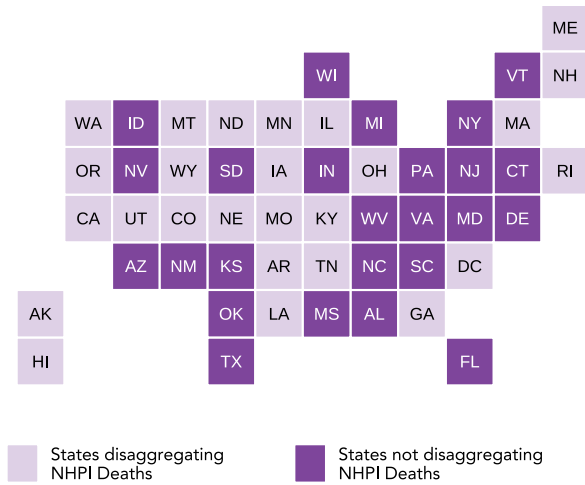


Figure 9: States Reporting with Non-Standard Race Categories for NHPI Deaths

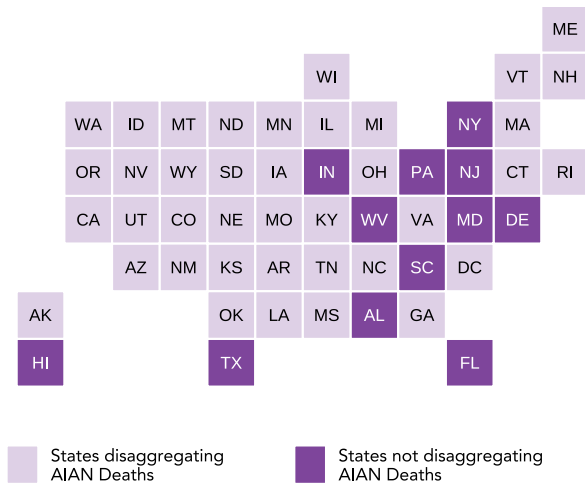


Figure 10: States Reporting with Non-Standard Race Categories for AIAN Deaths

c) States Omitted Certain Groups and Potentially Overcorrected for Privacy Concerns

One challenge of racial and ethnic data reporting is that, where group populations are small, it is potentially possible to deduce a specific individual’s identity from the reported data, jeopardizing patient confidentiality. Some states may have responded to these concerns by aggregating COVID-19 data from categories with small populations with COVID-19 data for other categories, or omitting data on categories with small populations altogether. The process of not reporting small numbers out of a concern for privacy is sometimes referred to as “data suppression.”⁸⁶ For example, if a jurisdiction had a very small number of COVID-19 cases among the AIAN population, it would conceivably be possible for a member of the public to narrow down the identity of an individual who had COVID-19 based on the reported data. Some states may have omitted or aggregated the AIAN racial category to prevent this disclosure, though the CRDT could not confirm this.

While patient privacy is very important, overcorrection in the data suppression context creates other harms for groups that have historically been subjected to tactics of erasure, like Indigenous people.⁸⁷ “The National Congress of American Indians has expressed concern that AI/AN people exist as the ‘Asterisk Nation’ in national studies because AI/AN populations are often described as ‘too small to be included,’ ‘too difficult to enumerate,’ or ‘too costly to be sampled appropriately.’”⁸⁸ Lack of data regarding the AIAN population precludes a comprehensive study of the inequities suffered by that group, and hinders the crafting of antiracist policy interventions to address those inequities. Privacy concerns must be balanced against the need for data about racial and ethnic inequities, particularly regarding often-ignored groups. Data suppression should only be implemented when actually necessary for the geographic level of reporting at issue, and the preceding sections illustrate the importance of states providing complete case-level race and ethnicity information to a national reporting system.

Continuing the above example, a state may have omitted or aggregated the number of AIAN cases reported at the state level in order to protect patient privacy, but these case numbers could still have been provided in a *national* count of AIAN cases without additional risk of disclosure, since the number of AIAN cases would not be so small as to raise privacy concerns at the national level. Once included in a national case surveillance dataset, standard suppression rules can be applied to prevent disclosure of an individual’s identity within smaller geographic units, as is common in federal health-related datasets. However, in the absence of robust national case-level reporting, the state practice of omitting or improperly aggregating small counts means that the nationwide impact of COVID-19 on these groups cannot be computed accurately due to incomplete data from a large number of states.

6. Few States Provided Information about Additional Racial and Ethnic Categories beyond the OMB Categories

Only two states, Michigan and Hawaii, reported any COVID-19 data for additional race or ethnicity categories beyond those specified in the OMB standards.

Michigan is the only state to have provided ethnicity data by “Arab ethnicity” (in addition to Hispanic/Latino/a/e/x ethnicity) in COVID-19 case and death data.⁸⁹ As a result, it is not possible to measure the impact of COVID-19 on people of Arab descent in any state other than Michigan.

Hawaii is the only state that provided detailed racial data for Asian, Native Hawaiian, and Pacific Islander subgroups in COVID-19 case, death, and hospitalization data. Hawaii broke down the Asian race category into Filipino, Japanese, Chinese, and Other Asian subcategories. Hawaii also reported two separate racial subcategories for Native Hawaiian and Pacific Islander. Hawaii’s reporting of subcategories indicated important differences in the impact of COVID-19 among these racial subcategories. For example, in data reported through July 2021, Native Hawaiians

represented 21% of the state population and 21% of COVID-19 cases; however, Pacific Islanders represented 4% of the state population and 19% of COVID-19 cases. The practice of collapsing these two categories into the standard OMB grouping “Native Hawaiian or Other Pacific Islander” would have obscured a clear disparity in Hawaii, with COVID-19 disproportionately impacting Pacific Islanders but not Native Hawaiians in the state. Similarly, Hawaii’s state population is composed of 15% Japanese and 16% Filipino people; however, Japanese people made up only 7% of COVID-19 cases whereas Filipino individuals comprised 20% of cases. By providing information within subcategories, Hawaii demonstrated how the pan-racial Asian category obscures important differences in COVID-19 risk between specific subpopulations. It is not possible to determine whether similar patterns are present outside of Hawaii due to lack of such detailed reporting from any other state.

The fact that so few states provided information for additional categories beyond the OMB categories demonstrates the central role that the OMB categories have come to play in race and ethnicity data collection and reporting, and the importance of reevaluating and amending those categories.

7. States Infrequently Updated Reported Data

The speed at which race and ethnicity data are made available, and how regularly such data are updated, is critical in understanding how racial and ethnic inequities change over time, particularly when dealing with emergencies that require a quick response. But the states infrequently updated their reported COVID-19 race and ethnicity data. By the *end* of the CRDT data collection period, most states updated the data on a daily basis, but a few states continued to update data weekly or less than weekly throughout the collection period, obscuring the CRDT team’s ability to see shifts in COVID-19 demographics in real time.

Louisiana, for example, reported race data for COVID-19 cases and deaths on a weekly basis.⁹⁰ The state reported Hispanic/Latino/a/e/x ethnicity information even less frequently. In a table entitled “Additional Data on COVID-19 Deaths in Louisiana” on the state’s website, a note stated: “Information in these tables is based on deaths where there is complete epidemiological data, and will be updated every two weeks.”⁹¹ Data in these additional tables were reported in percentage form only, with no stated denominator, making it unclear if those tables reflected long lag times due to delays in the collection of complete epidemiological data. No Hispanic/Latino/a/e/x ethnicity information was provided by Louisiana for COVID-19 cases.

In the event of newly arising COVID-19 cases disproportionately impacting the Hispanic/Latino/a/e/x community in Louisiana, the effect would not be immediately apparent through state-reported case data. The impact to this population would only become evident after cases had progressed to cause new

COVID-19 deaths, deaths had undergone complete epidemiological investigation, and the biweekly update to the state dashboard had occurred. Thus, infrequent updates of state reporting can result in the erasure of important health disparities until several weeks after the time when a targeted public health response may have intervened to save lives within the impacted community.

8. Summary

The CRDT fulfilled a vital need for race and ethnicity data during a global emergency. The team's work revealed critical racial and ethnic inequities in COVID-19 outcomes as the United States, like most other countries, was struggling to understand and respond to the disease. But the prominence and impact of the CRDT speaks to the insufficiency of existing racial and ethnic data collection structures in the United States.

Moreover, the challenges that the CRDT team experienced demonstrate that it is impractical to rely on state-reported racial and ethnic data in its current form for evidence-based policy making regarding nationwide problems. The CRDT had to contend with data incompleteness, inaccuracies, and outdatedness across multiple variables related to race, ethnicity, and time. States were inconsistent in whether they reported, what they reported, when they reported, how they reported, how much they reported, and how often they reported. The states also did not provide enough information to accurately understand the impact of COVID-19 on people whose racialized experiences are not adequately captured by the OMB categories. These problems often led to an underestimation of racial and ethnic disparities, preventing us from seeing and understanding the full extent of inequities in COVID-19 outcomes, which in turn, precluded the creation of antiracist policy interventions.

To be antiracist is to actively seek the information needed to counteract racism. The challenges and data quality issues the CRDT team faced are not inevitable and could largely be addressed through a single standardized system of nationwide racial and ethnic data reporting. Racial and ethnic data cannot continue to be an afterthought, nor should it be subject to the whims and idiosyncrasies of individual states. As the following section demonstrates, these issues are not unique to COVID-19 or the public health context. There are a variety of additional, pressing issue areas that require better racial and ethnic data.