

Boston University School of Law

## Scholarly Commons at Boston University School of Law

---

Faculty Scholarship

---

5-2022

### Toward Evidence-Based Antiracist Policymaking: Problems and Proposals for Better Racial Data Collection and Reporting

Neda Khoshkhoo  
*Boston University*

Aviva Geiger Schwarz

Luisa Godinez Puig

Caitlin Glass

Geoffrey S. Holtzman

*See next page for additional authors*

Follow this and additional works at: [https://scholarship.law.bu.edu/faculty\\_scholarship](https://scholarship.law.bu.edu/faculty_scholarship)



Part of the [Law and Race Commons](#), and the [Legal Writing and Research Commons](#)

---

#### Recommended Citation

Neda Khoshkhoo, Aviva Geiger Schwarz, Luisa Godinez Puig, Caitlin Glass, Geoffrey S. Holtzman, Elaine O. Nsoesie & Jasmine Gonzales Rose, *Toward Evidence-Based Antiracist Policymaking: Problems and Proposals for Better Racial Data Collection and Reporting* (2022).

Available at: [https://scholarship.law.bu.edu/faculty\\_scholarship/3433](https://scholarship.law.bu.edu/faculty_scholarship/3433)

This Report is brought to you for free and open access by Scholarly Commons at Boston University School of Law. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarly Commons at Boston University School of Law. For more information, please contact [lawlessa@bu.edu](mailto:lawlessa@bu.edu).



---

**Authors**

Neda Khoshkhoo, Aviva Geiger Schwarz, Luisa Godinez Puig, Caitlin Glass, Geoffrey S. Holtzman, Elaine O. Nsoesie, and Jasmine Gonzales Rose

# TOWARD EVIDENCE-BASED ANTIRACIST POLICYMAKING:

**Problems and Proposals  
for Better Racial Data  
Collection and Reporting**

May 2022



# ACKNOWLEDGEMENTS



---

This report was researched and written by Neda A. Khoshkhoo, Aviva Geiger Schwarz, Luisa Godinez Puig, Caitlin Glass, Geoffrey S. Holtzman, Professor Elaine O. Nsoesie, and Professor Jasmine B. Gonzales Rose. It was produced by the Boston University (BU) Center for Antiracist Research (the “Center”), with generous funding from the Rockefeller Foundation and the Commonwealth Fund. The authors are grateful to the Center’s faculty and staff, particularly Professor Ibram X. Kendi, Founding Director; Professor Katherine Levine Einstein, Assistant Director of Research; Professor Spencer Piston, Assistant Director of Policy; Dawn Graham, Data Scientist for the Antiracist Tech Initiative; and Nadia Frye Leinhos, Assistant to the Deputy Directors, for their guidance, feedback, and revisions. We thank our external reviewers, Professor Nancy Lopez, Professor Shawnita Sealy-Jefferson, Professor Hannah L. Walker, and Professor Charley E. Willison, for their comments and feedback. We thank BU Statistical Practice graduate student Anna Cook for creating the data visualizations and tables. We thank BU Law students Ashley Korkeakoski-Sears, Julian Burlando-Salazar, Gabriela Rosario, Taylor Mckinnon, and Fatima Elmansy, and BU Public Health graduate student Chloe Miller, for their research and editing assistance. We also thank Fenway Group for their design and layout work in the creation of this Report. Finally, we thank the hard-working team and volunteers from the COVID Tracking Project at *The Atlantic*, as well as the Center’s Racial Data Tracker team, without whom this project would not have been possible.

Cover image by elenabs.

© Copyright 2022, Boston University Center for Antiracist Research.  
All Rights Reserved.

# ABOUT THE BOSTON UNIVERSITY CENTER FOR ANTIRACIST RESEARCH



---

The Boston University Center for Antiracist Research was founded in 2020 by Professor Ibram X. Kendi. The vision of the Center is to create novel and practical ways to understand, explain, and solve seemingly intractable problems of racial inequity and injustice. We convene varied researchers and practitioners to foster exhaustive racial data collection, research-based policy innovation, solutions-focused narrative change initiatives, and practical advocacy campaigns. Together, these efforts power real-world antiracist change and impact.

Learn more at [bu.edu/antiracism-center](https://bu.edu/antiracism-center).

# TABLE OF CONTENTS



<b>I. EXECUTIVE SUMMARY.....</b>	<b>1</b>
A. The CRDT Team’s Findings Regarding Racial and Ethnic Data Collection and Reporting	
B. The RDT Team’s Findings Regarding Racial and Ethnic Data Collection and Reporting	
C. Policy Recommendations	
<b>II. BACKGROUND.....</b>	<b>8</b>
<b>III. THE COVID RACIAL DATA TRACKER EXPERIENCE.....</b>	<b>11</b>
A. The Need for the CRDT	
B. CRDT Methodology	
C. The Challenges of Obtaining COVID-19 Data by Race and Ethnicity	
D. Summary	
<b>IV. THE RACIAL DATA TRACKER EXPERIENCE.....</b>	<b>36</b>
A. RDT Methodology	
B. The Challenges of Obtaining Houselessness Data by Race and Ethnicity	
C. The Challenges of Obtaining Arrest and Police Violence Data by Race and Ethnicity	
D. Summary	
<b>V. POLICY RECOMMENDATIONS.....</b>	<b>49</b>
<b>VI. POLICY AND LEGAL CONSIDERATIONS REGARDING RACIAL AND ETHNIC DATA COLLECTION.....</b>	<b>54</b>
A. Potential Mischaracterization and Misuses of Racial and Ethnic Data	
B. Legal Considerations	
C. Limitations of the Federal Office of Management and Budget Categories	
<b>VII. CONCLUSION.....</b>	<b>58</b>
<b>VIII. APPENDICES.....</b>	<b>59</b>
<b>IX. ENDNOTES .....</b>	<b>61</b>

# I. EXECUTIVE SUMMARY



---

Race is a “power construct of collected or merged differences that lives socially.”<sup>1</sup> Race is neither a biological fact<sup>2</sup> nor a “fixed trait,”<sup>3</sup> and conceptions of race change over time.<sup>4</sup> In this sense, race is not “real.” However, racism is real. The collection and study of data concerning racial and ethnic<sup>5</sup> inequities and disparities allows us to better understand experiences of racism. Racial and ethnic data collection is thus “a crucial step in fighting racism and structural racial inequality.”<sup>6</sup>

Comprehensive demographic data on racial and ethnic disparities and inequities across critical policy areas allow us to see more clearly how and where racism manifests, including in access to resources and exposure to harms.<sup>7</sup> Studying the effects of racism, in turn, allows us to identify the policies that create and perpetuate inequities and to craft antiracist interventions and alternatives. Alongside racial and ethnic data, we need data concerning many other characteristics and variables that are connected to experiences of oppression and subordination, such as gender identity, disability, and socioeconomic status. Data collection across variables such as these provides a better understanding of the nuances of existing inequities and enables robust intersectional research.

Unfortunately, the current state of racial and ethnic data collection and reporting in the United States reflects many gaps and deficiencies that hinder antiracist policymaking. The efforts of the Boston University Center for Antiracist Research (the “Center”) to collect race and ethnicity data across key policy areas confirms that existing data sources are inadequate. State and local sources of racial and ethnic data often vary in standards for how to report, “what to report, when to report it, and even whether to report it at all.”<sup>8</sup> Moreover, lack of coordination regarding data collection at the federal level “can result in contradictory guidance to local and state agencies,”<sup>9</sup> further compounding the challenges of inconsistent, incomplete,

and inaccurate race and ethnicity data collection and reporting practices at the local and state levels. Additionally, in the rare instances when states are required to report racial and ethnic data to the federal government—for instance, as part of their involvement in Medicaid—such requirements have historically not been well enforced.<sup>10</sup>

The lack of uniform and robust standards for racial and ethnic data collection and reporting has meant that existing data repositories are incomplete, contain errors, are usually incompatible with each other, and are often internally inconsistent across years, jurisdictions, subjects, and levels of analysis.<sup>11</sup> Moreover, a great deal of race and ethnicity data are not publicly available or easily accessible. Some entities are reluctant to make any changes to their own practices until a more centralized and standardized system is in place.<sup>12</sup>

In early Spring of 2020, Center staff<sup>13</sup> and *The Atlantic's* COVID Tracking Project team saw that the poor state of race and ethnicity data collection and reporting in the United States was hindering meaningful evidence-based policy responses to the COVID-19 pandemic. Among other things, the lack of robust data on COVID-19 cases, hospitalizations, and deaths by race and ethnicity was preventing policymakers from identifying and responding to resource inequities for their jurisdictions. The two entities quickly collaborated to fill this gap through the COVID Racial Data Tracker (CRDT), which collected race and ethnicity data on COVID-19 cases, hospitalizations, deaths, and to a lesser extent, testing (hereinafter collectively referred to as “COVID-19 outcomes”) in the United States from April 2020 to March 2021.<sup>14</sup> While the CRDT was active, it was the most comprehensive and timely source of COVID-19 racial and ethnic data in the country, and was frequently cited by news outlets, government representatives, advocates, and scholars.

The Center then expanded its race and ethnicity data collection efforts through its Racial Data Tracker (RDT), which collects data on other key issue areas such as houselessness, criminal arrests, and police violence. This data collection work builds upon decades of effort by scholars and advocates from a variety of disciplines to obtain accurate and complete racial and ethnic data to inform evidence-based policy making. Public health scholars and advocates have been at the forefront of the call for better racial and ethnic data,<sup>15</sup> but the need extends to all issue areas in which people of color have been pushed to the margins. Indeed, race and ethnicity data are often strategically and deliberately omitted in order to preserve existing inequities, or neglected out of convenience or indifference—all of which works to perpetuate racism.

The CRDT and RDT teams' experiences provide great insight into the deficiencies of the racial and ethnic data collection and reporting methods used by local, state, and federal agencies, as well as the variety of methods they employ. These teams' work illustrates the need for a single, standardized, nationwide system of



data collection and reporting by race and ethnicity.<sup>16</sup> Below, we summarize the challenges the teams faced and offer policy recommendations, which are discussed in depth in the following sections of this Report.

## A. The CRDT Team's Findings Regarding Racial and Ethnic Data Collection and Reporting

The CRDT team's experience manually collecting state-reported racial and ethnic data on COVID-19 outcomes demonstrates that such state-reported data suffer from deficiencies that can cause errors and underestimations of racial and ethnic inequities. The CRDT team encountered the following challenges:

**1. Data were incomplete.** Many states failed to report any racial and ethnic data on COVID-19 outcomes for several months after the outbreak, and some states *never* reported such information for the duration of the CRDT collection period (April 2020 to March 2021). States that did report racial and ethnic data were not consistent about whether and how they did so from one month to the next. When states did report, race and ethnicity data were often only available for some of the COVID-19 outcomes (cases, hospitalizations, tests, or deaths). Additionally, for each of these outcomes, the data were often incomplete. This means, for example, that the race and ethnicity of every known person who was hospitalized for COVID-19 were not necessarily recorded. Likewise, different states failed to collect data concerning various racial or ethnic groups. For example, some states did not report any data concerning Hispanic/Latino/a/e/x<sup>17</sup> people. Such incomplete race and ethnicity data within and across states impede thorough analysis of national trends and preclude comparisons across jurisdictions.

**2. States' data reporting methods varied.** States varied not only in terms of what race and ethnicity data they reported and whether they reported, but also in *how* they presented their data to the public. This lack of uniformity across jurisdictions presents a barrier for those who are studying national trends. Moreover, some of the states' reporting conventions did not provide sufficient information. For example, some states presented the data *only* as percentages (rounded to certain decimal points or to the nearest whole number) without making raw numbers available. Reporting methods that do not make raw numbers available to the public can cause estimation and calculation errors that make it impossible for researchers, policymakers, and advocates to accurately interpret the full extent of racial and ethnic inequities.

**3. States did not adequately account for the ways that race and ethnicity can intersect.** Many states' reporting methods did not account for the fact that some people fall into both a racial *and* an ethnic group, or into multiple racial or ethnic groups. Those that did had a variety of approaches in how they handled this complexity, with varying degrees of effectiveness. Failure to carefully present race and ethnicity as intersecting measures when appropriate can lead to

errors such as double counting individuals or undercounting group membership. This was particularly the case for the Hispanic/Latino/a/e/x and multiracial groups.

**4. States used non-standard racial and ethnic categories.** States were not consistent in how they categorized race and ethnicity. Some failed to count certain racial and ethnic categories outright, and others lumped categories together in nonstandard and overly broad ways, both of which obscure experiences of racism and subordination.

**5. States reported the bare minimum, or less.** Most states, at best, provided information about racial and ethnic categories that are included in the federal Office of Management and Budget (OMB) racial and ethnic categories (“OMB categories”),<sup>18</sup> which are used for certain federal data collection purposes such as the Decennial Census. Only a few jurisdictions departed from the OMB to provide more detailed race and ethnicity breakdowns, while many jurisdictions provided even fewer race and ethnicity categories than those outlined by the OMB.

**6. States infrequently updated their data.** States varied in terms of when and how often they updated their race and ethnicity data, making it difficult to compare trends across states. Many states updated data infrequently, preventing real-time assessments of health inequities.

## **B. The RDT Team’s Findings Regarding Racial and Ethnic Data Collection and Reporting**

Complementing the work of the CRDT, the RDT’s examination of datasets regarding houselessness, criminal arrests, and police violence<sup>19</sup> confirms that many existing public datasets that report information by race and ethnicity are insufficient to inform policy choices. Indeed, despite ostensibly “national” data collection, these data—like COVID-19 information—are collected in a highly decentralized and uncoordinated way by state and local entities. The datasets examined for this study contained the following problems:

**1. Data are incomplete.** National datasets that provide race and ethnicity information frequently rely on local, regional, and state reporting. Since participation in such reporting schemes is voluntary, many entities do not contribute data to these sources, or fail to include racial and ethnic information when they do report data. As a result, existing datasets with race and ethnicity information are often incomplete, especially when assessed longitudinally (across time). Furthermore, when data are missing (or inaccurate) at the state, regional, or local level, their aggregation results in data that are inaccurate and potentially at odds with data from higher-level sources.

**2. It is often not possible to obtain data that are disaggregated at the local or state levels.** National datasets often fail to disaggregate data at local or state levels. Moreover, some race and ethnicity data, such as data concerning houselessness, are gathered and organized by non-governmental organizations or federally-designated entities with arbitrary geographical boundaries that do not correspond to the boundaries of cities, towns, and the like. For example, such an entity's jurisdiction might cover multiple cities, or even parts of cities. These entities often cannot or do not disaggregate data at the city level, where important policy decisions are often made. As a result, policymakers cannot access data for their specific jurisdiction. The lack of disaggregated data can obscure important disparities that may exist at state or local levels, or between states and locales.

**3. Longitudinal data are largely unavailable.** Datasets with racial and ethnic information have varied and arbitrary time periods. This temporal patchwork prevents analysis of trends over time or the impacts of policy changes.

**4. Methodologies are varied and incomplete.** Existing datasets vary too much in their methodologies to be used side-by-side in an attempt to overcome the above-mentioned challenges.

### C. Policy Recommendations

Based on the data deficiencies and challenges described above, we make the following policy recommendations,<sup>20</sup> which are discussed in more detail in Section V of this Report:

**1. Centralize and standardize racial and ethnic demographic data collection and reporting across critical issue areas.** Federal leadership is needed to create a single standardized, nationwide system of data collection and reporting by race and ethnicity in key policy areas including, but not limited to, health, housing, employment, education, the criminal legal system, and the environment.

**2. Use existing federal race and ethnicity standards as a starting point, and regularly reevaluate and amend such standards.** Existing federal race and ethnicity data standards should be the starting point for a centralized and standardized race and ethnicity data collection system. The federal government should review and amend these standards immediately, and continue to do so regularly with input from scholars, community members, and advocates. This process should include a close examination and reevaluation of the OMB racial and ethnic categories.

**3. Create an oversight board to monitor how the government uses race and ethnicity data.** An external oversight board, which should include community partners and advocates, should monitor the creation of a centralized data collection system to ensure that the government's use of the data complies with existing laws and is used responsibly, ethically, and equitably.

**4. Incentivize consistent and timely state and local participation through adequate funding.** The federal government should create funding incentives that encourage states and localities to collect and report racial and ethnic data to the centralized system in a consistent and timely manner. This funding should include conditional funds to incentivize participation, and additional funds to support data administration infrastructure.

**5. Monitor noncompliance.** States and other entities that choose not to participate in a centralized and standardized system of racial and ethnic data reporting should be monitored by the aforementioned oversight board. Data reporting dashboards, websites, reports, and similar platforms should explicitly note which entities have chosen not to participate in order to increase public visibility and accountability.

**6. Make data available at the national level, and also disaggregate at the state and local levels.** Race and ethnicity data should be collected and reported such that researchers and policymakers have the option of obtaining data aggregated at the national level and disaggregated at state and local levels.

**7. Incentivize more granular race and ethnicity reporting.** State and local entities should be incentivized to collect data at more granular levels than the OMB racial and ethnic categories to reflect the ethnic and racial makeup of their particular jurisdictions, and to reveal inequities between subpopulations. These granular data should be collected and disaggregated such that they can be collapsed into future permutations of the OMB categories as needed.

**8. Report race and ethnicity data as intersecting measures.** Detailed race and ethnicity data should be collected and reported as intersecting characteristics in order to provide the most clarity and flexibility regarding the breakdown of the data. For example, reported data should make clear how many people experiencing houselessness who fell within the Hispanic/Latino/a/e/x group also fell within the Black group, and how many did not. This system should not limit the number of ethnicities or races that may apply to each person.

**9. Collect and report data across additional characteristics and variables.** Experiences of racism are multidimensional. In order to better understand the nuances of racism and enable robust intersectional research, race and ethnicity data should be disaggregated by additional characteristics and variables, including, but not limited to, sex, gender identity, sexual orientation, age, disability, religion, income, linguistic usage, educational attainment, socioeconomic status, “street race,”<sup>21</sup> and national origin.

**10. Tailor privacy-related data suppression practices to the realities of each dataset.** Small numbers may need to be suppressed for privacy reasons in local or state databases, but may not need to be suppressed when they are aggregated at the national level. State and local entities that collect race and ethnicity data should remove any identifying information, such as names and birth dates, before reporting the data to a national database. Those entities should then report *all* disaggregated data to a national database through an encrypted platform, and the national database should decide whether suppression of small numbers is necessary to protect privacy at the national level. State and local entities should, of course, continue to suppress small numbers for their own, publicly-facing reporting when necessary for privacy reasons.

**11. Make data publicly available and accessible.** Apart from data that must be suppressed or protected due to privacy or serious confidentiality concerns, racial and ethnic data should be made freely, publicly, and easily accessible for use by advocates, scholars, policymakers, and others. All such data should contain explicit disclaimers and guidance on best practices concerning any gaps and limitations associated with them.

## II. BACKGROUND



In the earliest months of the COVID-19 pandemic, the Center Director, Ibram X. Kendi, feared that the virus was disproportionately harming and killing people of color.<sup>22</sup> He could not get data to confirm or deny this fear, however, as nationwide racial and ethnic data on COVID-19 outcomes were practically nonexistent.<sup>23</sup> At the time, some public officials referred to COVID-19 as the “great equalizer.”<sup>24</sup> In reality, public health emergencies disproportionately devastate under-resourced communities of color, and COVID-19 was no exception.<sup>25</sup> But the lack of reliable data regarding COVID-19 outcomes by race and ethnicity was a major barrier to creating equitable policy interventions.<sup>26</sup> Dr. Kendi began publishing pieces in *The Atlantic* urgently calling for COVID-19 race and ethnicity data.<sup>27</sup> As the months wore on and people continued to die and suffer, various states slowly began to release some racial and ethnic data for COVID-19 outcomes in an uncoordinated, piecemeal manner, but nationwide trends remained unknown, and there were many gaps in the data.

With limited information forthcoming from any national agency, Dr. Kendi partnered with *The Atlantic*’s COVID Tracking Project in the early Spring of 2020 to quickly assemble a team of journalists, researchers, scholars, and volunteers to fill this gap, and the COVID Racial Data Tracker (CRDT), the Center’s first research project, was born. From April 2020 to March 2021, the CRDT team painstakingly worked to manually collect the available racial and ethnic COVID-19 data reported by state and U.S. territory.<sup>28</sup> The CRDT, which constituted the most complete and timely source of COVID-19 race and ethnicity data at the time, confirmed that COVID-19 reflected the long-standing racial and ethnic health inequities that plague the U.S. population: Black, Brown, and Indigenous people were disproportionately suffering and dying from the virus.

Nevertheless, the CRDT data could only be as complete and accurate as the data the team was collecting from each state. The CRDT team encountered a number of data quality challenges due to the varied ways states reported (or failed to report) data, and responded to those challenges as best it could. While the CRDT team's work successfully demonstrated the existence of racial and ethnic inequities in COVID-19 outcomes, state data quality issues prevented the team from revealing the full extent of the inequities.<sup>29</sup> The CRDT is therefore a useful case study of how state data sources are an insufficient substitute for single, standardized, nationwide data, and how difficulties in combining data reports from the different states and territories, each with their own reporting practices, can introduce errors and obscure evidence of racism. This case study allows us to identify better practices for collecting and reporting race and ethnicity data and the measures needed to fill existing gaps in the data.

Aware that the data quality issues the CRDT team faced were not unique to COVID-19 or the public health context, Center staff then decided to expand the CRDT's work by establishing the Racial Data Tracker (RDT) to continue this investigation. The Racial Data Tracker team is advancing Racial Data Science, a multidisciplinary field marrying antiracist research and data science that involves the application of mathematics, statistics, computer science, visualizations, storytelling, and social science methods to large volumes of data. To do this, the RDT team is working to amass the largest online, publicly accessible collection of racial and ethnic data at the local, regional, and national levels, spanning key issue areas including health, education, employment, politics, housing, and the criminal legal system. The RDT team began by collecting racial and ethnic data on houselessness, criminal arrests, and police violence.

The RDT team's experience sheds more light on many of the data quality issues identified by the CRDT team. The RDT team tried to gather data from numerous existing national, state, and local datasets in order to obtain the clearest picture of racial and ethnic inequities across important issue areas. The RDT team's experience confirms that racial and ethnic data collection and reporting in the United States are frequently left to state, local, and nonprofit entities, whose participation is often voluntary, inconsistent, incomplete, and poorly enforced.

The incompleteness of racial and ethnic data is a major barrier to crafting equitable policies and meaningfully measuring progress toward racial equity.<sup>30</sup> "If we can't see racial disparities, then we can't see the racist policies behind any disparities and deaths."<sup>31</sup> And if we can't identify racist policies—that is, policies that produce or sustain racial inequity—we cannot dismantle them.<sup>32</sup> The CRDT and RDT teams' experiences underscore the urgent need for a single, standardized, nationwide system of collecting and reporting data by race and ethnicity. Such a system would

allow policymakers to see whether and how particular policies and practices are racist, and respond with informed solutions, paving the way for eradicating “the original American virus: racism.”<sup>33</sup>

The following sections of this Report examine the data collection processes of the CRDT and RDT teams from the Spring of 2020 to the Summer of 2021. Their experiences illustrate the deficiencies of existing publicly available racial and ethnic data<sup>34</sup> for COVID-19, houselessness, criminal arrests, and police violence. We describe the current challenges of collecting robust and accurate racial and ethnic data, demonstrate ways in which incomplete data obscure racial inequities and erase evidence of racism, and offer antiracist policy recommendations.



# 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.<sup>35</sup> National case surveillance datasets from the U.S. Centers for Disease Control and Prevention<sup>36</sup> (CDC) have been documented as highly incomplete throughout the pandemic.<sup>37</sup> 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%.<sup>38</sup> Put another way, as of April 2021, 8.6 million out of 24.4 million cases were missing race and ethnicity information.<sup>39</sup> The percentage of COVID-19 cases with known race and ethnicity did not improve through October 2021.<sup>40</sup>

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.<sup>41</sup> Many local and state jurisdictions fail to provide all COVID-19 case reports to the CDC,<sup>42</sup> and even for COVID-19 case reports that *are* provided to NNDSS, race and ethnicity are not mandatory fields in all jurisdictions.<sup>43</sup> 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.<sup>44</sup> 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),<sup>45</sup> 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.<sup>46</sup> “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.<sup>47</sup> 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).<sup>48</sup> 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.<sup>49</sup> 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.<sup>50</sup> While this reporting system is mandatory, it does not ask hospitals for any demographics aside from age brackets.<sup>51</sup> It notably does not include any racial or ethnic demographics in the facility-level information requested.<sup>52</sup>

Federal agencies have publicly acknowledged the need for a more robust nationwide system of reporting and collecting COVID-19 race and ethnicity data.<sup>53</sup> 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.”<sup>54</sup> 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.”<sup>55</sup> 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.<sup>56</sup> The CDC agreed with this recommendation.<sup>57</sup> 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.<sup>58</sup>

## **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.<sup>59</sup>

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.<sup>60</sup>

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**

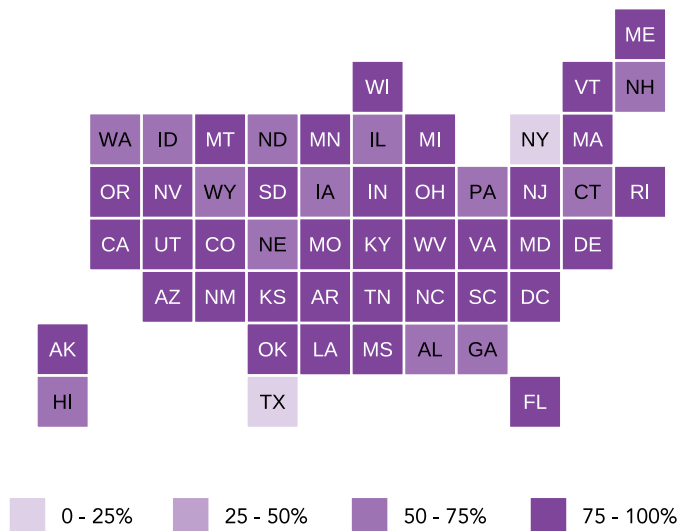
The major challenges that the CRDT faced included missing and incomplete data, inconsistencies and deficiencies in reporting practices, inconsistencies and

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).<sup>61</sup>

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.<sup>62</sup> 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”<sup>63</sup> data and “reported unknowns.”<sup>64</sup> 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**

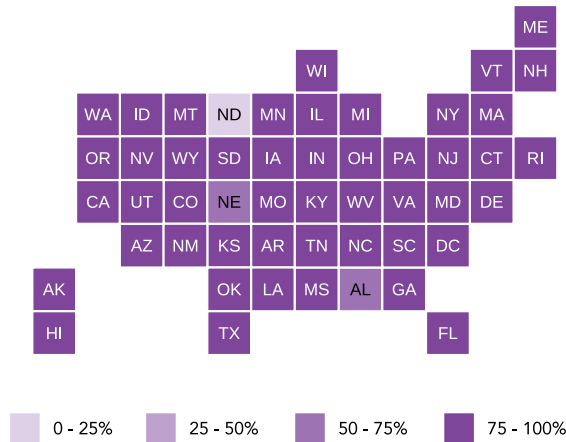


Figure 2: Percent of Deaths with Known Race

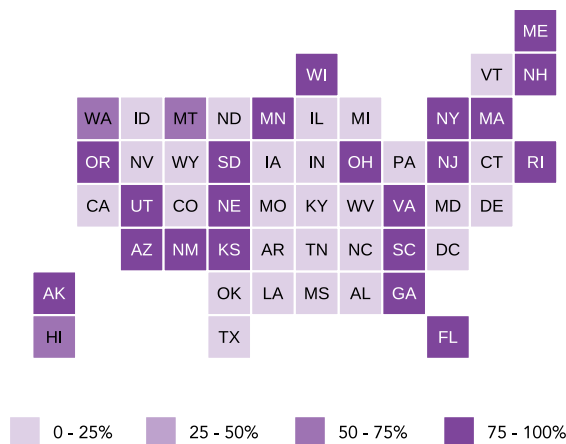
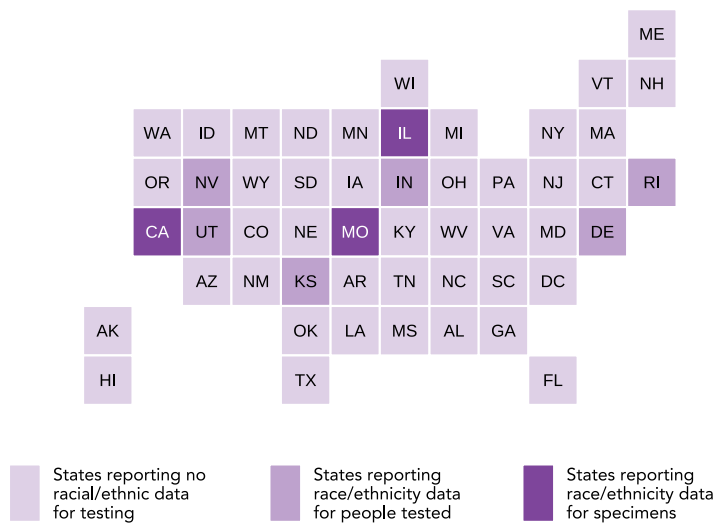


Figure 3: Percent of Hospitalizations with Known Race

States also varied greatly in terms of when they *first* began reporting COVID-19 racial data.<sup>65</sup> 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.<sup>66</sup> 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.<sup>67</sup> 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%<sup>68</sup> 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.<sup>69</sup> 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%<sup>70</sup> 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,<sup>71</sup> 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**

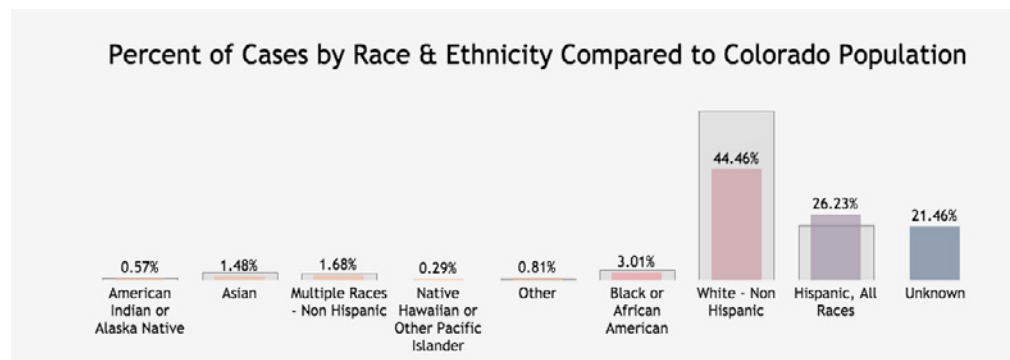
**2. States’ Data Reporting Conventions Were Deficient and Inconsistent**

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).<sup>72</sup> 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.<sup>73</sup>

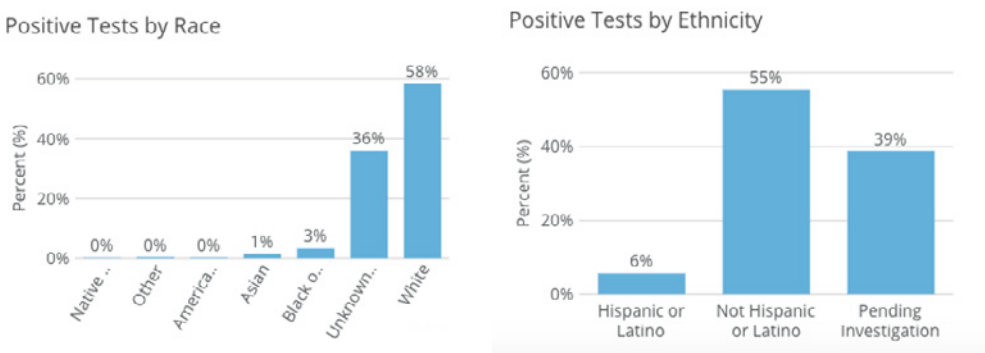


**Image 1: Colorado’s Rounding Convention for COVID-19 Race and Ethnicity Data, 2021**

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.”<sup>74</sup> 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.<sup>75</sup> 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.<sup>76</sup> 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
<b>Total</b>	<b>2,286,332</b>	<b>95,607</b>	<b>36,973</b>

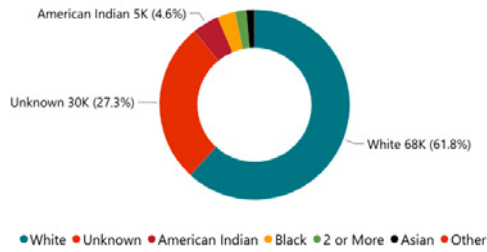
**Image 5: Florida’s Display of Intersecting COVID-19 Race and Ethnicity Data**

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/](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.

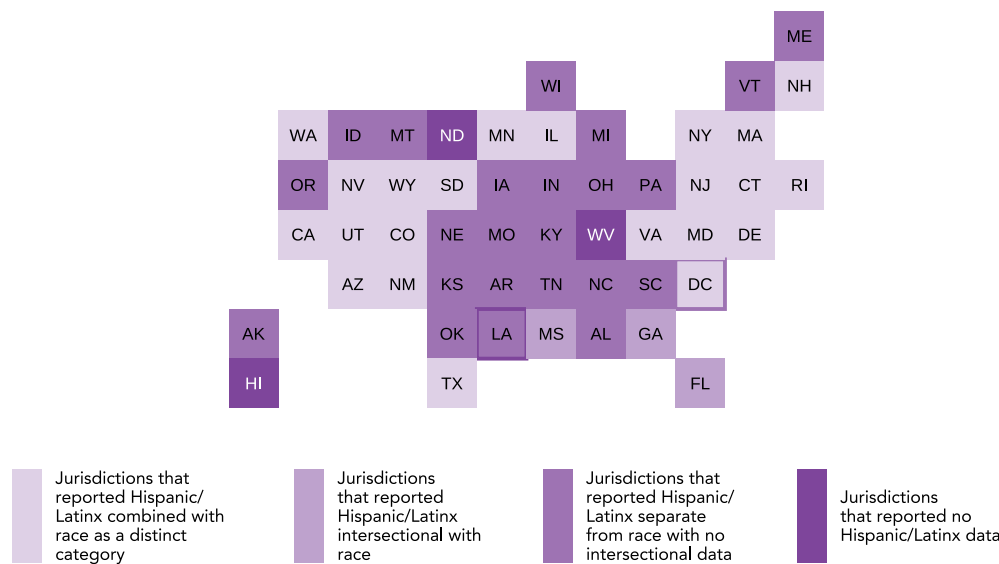
Actual Race	Selected Calculation Demographic
White	68,358
Unknown	30,243
American Indian	5,068
Black	3,387
2 or More	2,118
Asian	1,462
Other	44



**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.<sup>77</sup>



**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.<sup>78</sup> 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.<sup>79</sup>

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

		Cases	Population*	Cases per 1000 pop.	Case rate ratio (vs. White)
<b>Race</b>	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
<b>Ethnicity</b>	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:

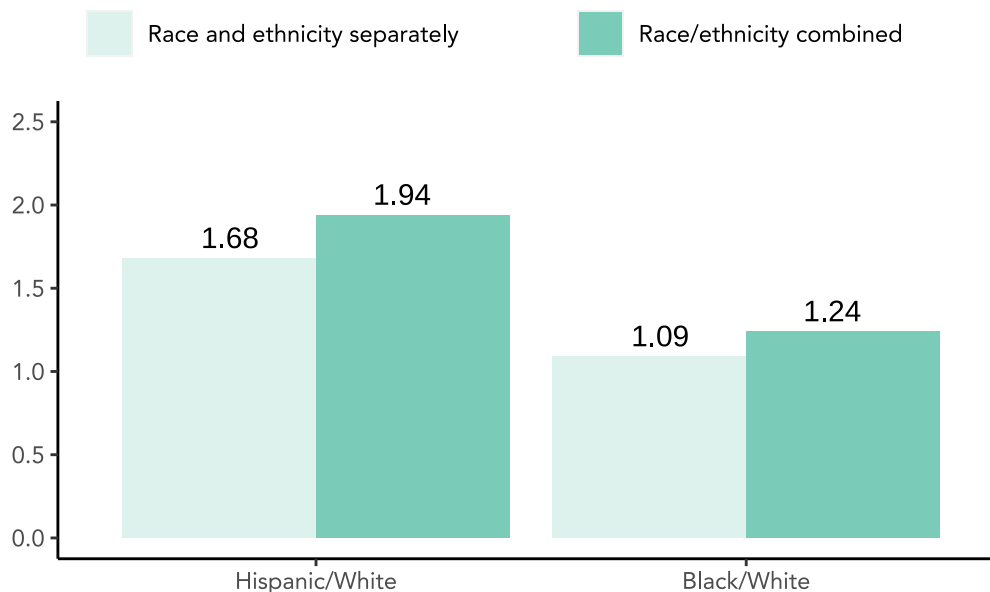
\* Population data from US Census American Community Survey 5-year estimates.

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).<sup>80</sup> 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).<sup>81</sup> 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.<sup>82</sup> 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	
	<b>108.53</b>	

**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,<sup>83</sup> 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.

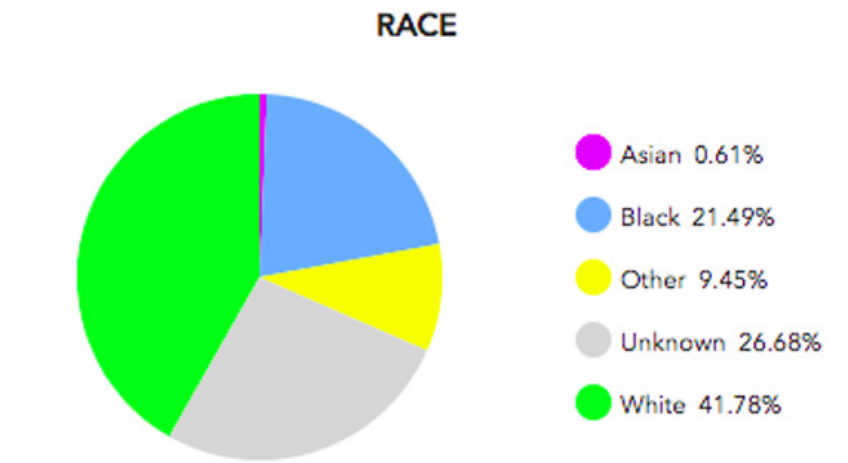
Below, we describe the major ways that the states departed from the OMB and varied from each other in their treatment of racial categories.

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.<sup>84</sup>

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.

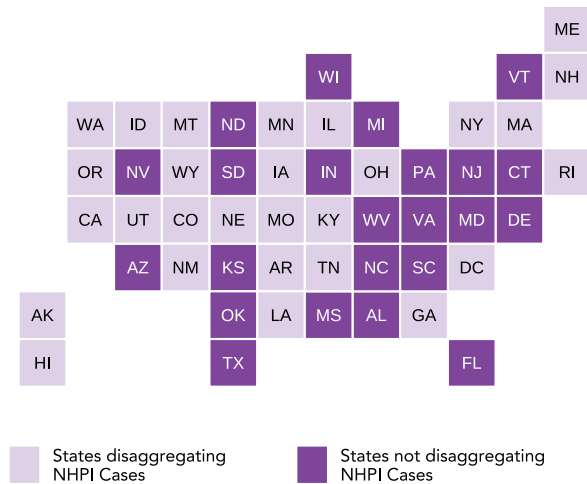
Montana grouped “American Indian or Alaska Native” and “Native Hawaiian or Other Pacific Islander” into a combined category labeled “American Indian, Alaska Native, Native Hawaiian, Pacific Islander.” Montana was the only state to aggregate these groups in COVID-19 reporting, though it later began disaggregating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

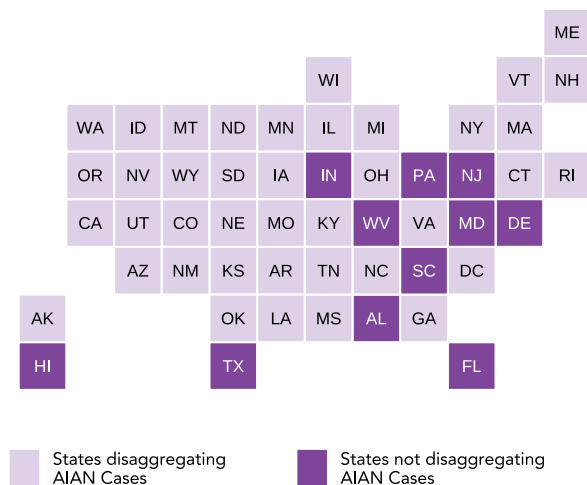
Finally, a number of states put one or more of the OMB categories into a combined “Other” category. Indiana, for example, reported a category labeled “Other Race” with a note specifying that “[o]ther races included American Indian, Alaska Native, Native Hawaiian, Other Pacific Islander, two or more races,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

ISDH as other race.”<sup>85</sup> 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.

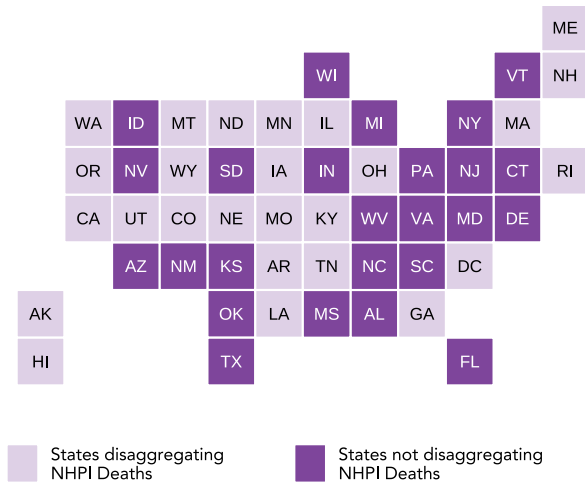
Taken as a whole, the result of the practices described in Sections (a) and (b) above was an undercount of national totals for some racial categories. The two race categories most affected by this issue were the two smallest by population: AIAN and NHPI. Figures 7 through 10 show the states that did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



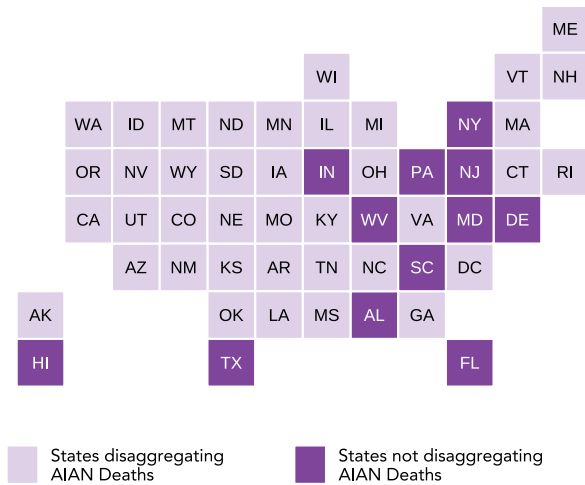
**Figure 7: States Reporting with Non-Standard Race Categories for NHPI Cases**



**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.”<sup>86</sup> 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.<sup>87</sup> “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.’”<sup>88</sup> 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.<sup>89</sup> 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.<sup>90</sup> 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.”<sup>91</sup> 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.

# IV. THE RACIAL DATA TRACKER EXPERIENCE



---

In this Section, we describe the experience of the RDT team collecting data in the summer of 2021. We describe the RDT methodology, detail the challenges and data deficiencies the RDT team encountered for each issue area it studied, and summarize the implications for antiracist policymaking.

## A. RDT Methodology

In the summer of 2021, the RDT team, which consisted of Center faculty, staff, postdoctoral fellows, and student interns, examined racial and ethnic data in the areas of houselessness, criminal arrests, and police violence, and identified many deficiencies in the available information. The team began by collecting data at the national level to see what information was available and disaggregated for each issue area. It then tried to fill data gaps by obtaining information directly from states and the fifty largest cities, but often could not do so because data were unavailable.<sup>92</sup> The RDT data collection efforts are ongoing, but for purposes of this report, the data collection period began in May 2021 and ended in August 2021.

The RDT team gathered data from federal, state, and city websites and data dashboards, as well as reports and data published by research institutions. In some instances, the team informally contacted cities' statistics departments, health departments, and social services departments to supplement missing data. Data ranged widely in its accessibility, with some data easily downloadable from publicly available websites and others only obtainable upon request to public officials. The team focused its efforts on collecting data in the form of cumulative counts.

## **B. The Challenges of Obtaining Houselessness Data by Race and Ethnicity**

The RDT team encountered several deficiencies in houselessness data concerning race and ethnicity. First, the data sources employ different methodologies. The datasets vary in terms of whether they reflect houselessness data collected on a given night, during a three-day period, or during a one-year period. Second, houselessness data are aggregated in a manner that does not aid in the creation of evidence-based policies that address particular geographic or jurisdictional needs: generally such data are publicly reported as aggregated national statistics. Additionally, when the data are disaggregated beyond the national level, they are often disaggregated at units of geography that do not correspond to the jurisdictions of governments that make policy decisions (such as cities). Compounding this issue is the fact that the jurisdictions that do report disaggregated data are not fixed, and their boundaries often change. Third, longitudinal data are severely lacking, which prevents both an analysis of trends over time and an evaluation of the efficacy of policy interventions. These challenges demonstrate the inadequate and piecemeal nature of houselessness race and ethnicity data.

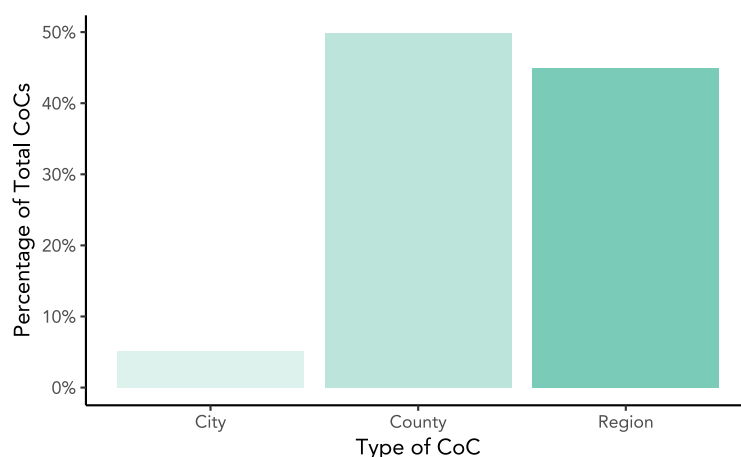
### **1. Datasets with Race and Ethnicity Information Are Deficient and Vary in Methodology**

The RDT team examined four major national sources of houselessness data by race and ethnicity: Point-in-Time (PIT) counts, the Homeless Management Information System (HMIS), the federal Census, and the American Community Survey (ACS).<sup>93</sup> These sources are generally compliant with the OMB racial and ethnic categories except for the ACS, as explained below. These datasets each contain limitations and posed challenges for the RDT team.

#### **a) Deficiencies of Datasets That Rely on Continuums of Care for Reporting**

A major obstacle to robust race and ethnicity data analysis in the houselessness context is the wide reliance on Continuums of Care (CoCs) for reporting. CoCs are federally designated planning bodies responsible for coordinating the funding and delivery of services for people experiencing houselessness.<sup>94</sup> Two of the major data repositories studied by the RDT—PIT counts and the HMIS—are collected by the U.S. Department of Housing and Urban Development (HUD), which in turn obtains data from CoCs.<sup>95</sup> HUD depends on CoCs to report point-in-time and yearly counts of people experiencing houselessness.<sup>96</sup> CoCs are incentivized to provide data, including racial and ethnic data, to HUD as a condition of receiving federal funding.<sup>97</sup>

Reliance on CoCs for data reporting is problematic for many reasons. CoCs operate at various jurisdictional levels that rarely match the geographic boundaries of local government or other Census geographies that engage in policy making, as demonstrated in Figure 11. Specifically, CoCs often encompass multiple city or



**Figure 11: Level of Jurisdiction of Continuums of Care Reporting to HUD**

county jurisdictions, or portions of such jurisdictions. They also do not necessarily operate as part of a local government. “In 2020, nearly 60% of Continuums of Care surveyed were organized as a non-governmental entity.”<sup>98</sup> Consequently, CoC data are usually not disaggregated at the city or town level where houselessness policies are often made, as discussed in more detail below. The arbitrary CoC boundaries also make it challenging to link Census data—including racial and ethnic data—to CoCs in any kind of analysis. Additionally, CoCs use varying collection and data storage methodologies which can bias final counts because they are not comparable (i.e. one CoC may collect data for unique users and another may collect data for every user per day).<sup>99</sup> And when CoCs merge, as they often do, they change the boundaries under which they operate, precluding longitudinal analysis.

The amount of race and ethnicity data HUD has managed to obtain from CoCs has also decreased over time, likely because CoCs have decreased in number. The RDT’s case study of CoC participation in data reporting for three separate years (2007, 2013, and 2016) vis-a-vis 2020 demonstrates that even the same CoCs do not always participate in data reporting, and that overall participation has decreased since 2007.<sup>100</sup> The RDT team investigated the reasons why some CoCs did not participate in data reporting in 2020, but this information was not widely available. Those that did provide a reason indicated that the CoC had merged with another CoC, had ceased to exist, or had not applied for HUD funding and thus was not required to report.

Separate from their reliance on CoCs, PIT counts and HMIS have additional limitations regarding the quality of their race and ethnicity data. The most complete publicly available, disaggregated data (at the CoC level) on race and ethnicity comes from the PIT counts, which capture only a point-in-time snapshot of houselessness on a given night each year.<sup>101</sup> PIT counts add up the number of shelter users and unsheltered individuals during a given night in January.<sup>102</sup> PIT counts have limited value because they cannot be easily compared to other sources of houselessness data, which reflect data collected over the course of a year.

By contrast, data from HMIS estimate the number of unique shelter users in a fiscal year as recorded by shelters' administrative records, but HMIS only publicly reports information on race and ethnicity as aggregated, national counts. Instead of disaggregating data by geographic jurisdiction, these publicly available datasets provide national counts that are disaggregated by type of shelter option, such as family or emergency shelters. Additionally, HMIS estimates rely on shelter reports, which omit information about unhoused people who are not in shelter settings and may introduce other inconsistencies in the data.<sup>103</sup> HMIS collects race and ethnicity data at the CoC level as well, but these data are not publicly available.<sup>104</sup>

#### **b) Deficiencies of Datasets That Do Not Rely on Continuums of Care for Reporting**

Houselessness datasets that do not rely on CoCs for data, namely the Census and the ACS, have their own limitations and variations regarding race and ethnicity information. The 2010 Decennial Census included racial and ethnic data on people experiencing houselessness during a three-day period.<sup>105</sup> Advantages of this database include that it employs collection efforts that follow the same methodology throughout the country, unlike those done by individual CoCs, and that it includes some information on unsheltered individuals.<sup>106</sup> However, like the one-day focus of the PIT counts, the three-day unit of measurement is likely not representative of the overall yearly population of people experiencing houselessness, and is difficult to analyze in relation to other standard measures of houselessness that are collected over the course of a year. Additionally, the Census aggregates race and ethnicity data at the state and national level, precluding a better understanding of where and how racial inequities arise at the local level.

Finally, the ACS has collected micro-level data about people in emergency and transitional shelters with sleeping facilities, but it omits data on unsheltered individuals, and the information it provides is not publicly available. One advantage of the ACS is that it collects data in some major cities, which is missing from other sources of collection. While the database includes categories for race and ethnicity, it is unclear what those categories are, as the RDT team could not access the repository.<sup>107</sup>

**c) Differences in Methodologies Cannot Be Reconciled**

The RDT team examined the differences between the houselessness data repositories to investigate whether they could be combined to overcome some of the aforementioned data gaps. The RDT team compiled Table 9, which reveals key differences and deficiencies in the datasets that prevent them from being used in a complementary way.

**Table 9. Comparison of Data Sources**

	HUD Point-in-Time	HUD Homeless Management Information System	ACS	2010 Decennial
<b>Race data</b>	yes	yes	yes	yes
<b>Ethnicity data</b>	yes	yes	yes	yes
<b>Years Available</b>	2015-2020 (with variables on race, otherwise since 2007)	2007-2017	2006-2020	2010
<b>Jurisdiction available</b>	CoC level, state level, US level	US level	State level, US level, some major cities	State level, US level
<b>Public Availability</b>	yes	yes	no	yes
<b>Period of Collection</b>	One night in January	Annual number of shelter users	Yearly	Conducted march 29-31, 2010
<b>Data on Sheltered Individuals</b>	yes	yes	yes	yes
<b>Data on Unsheltered Individuals</b>	yes	no	no	some
<b>Method of Collection</b>	Individual Reports from CoCs	Individual Reports from CoCs	Survey of People residing in shelter facilities + imputation techniques	Survey of People residing in shelter facilities

**2. Race and Ethnicity Data Are Not Sufficiently Disaggregated at Local Levels**

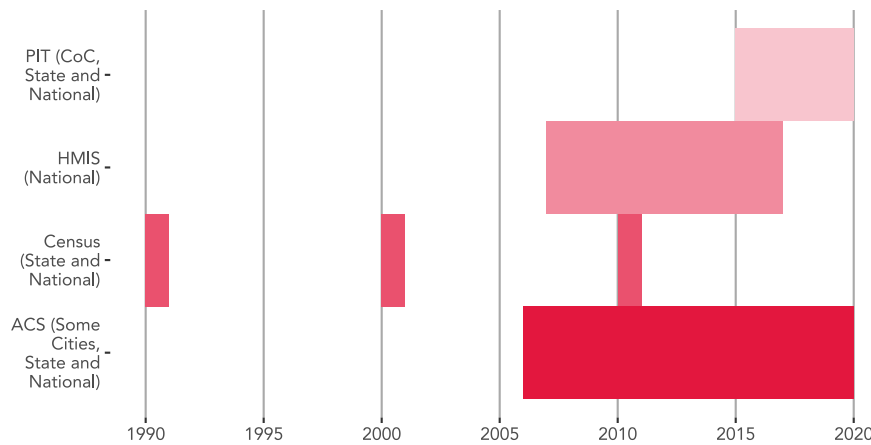
Racial and ethnic data on houselessness are generally not disaggregated by municipality or locality, which makes it difficult to design and evaluate houselessness policies that originate at the local level. While many important policy decisions are made by city officials responding to city problems, city-level racial and ethnic demographic data are lacking in the datasets described above. At best, houselessness data are disaggregated at the CoC level, but as explained above, CoC boundaries often do not correspond with local governments. PIT data are available at the CoC, state, and national levels, HMIS data are only available at the national level, Census data are collected for states and the national level, and ACS data are collected for some cities, all states, and at the national level.

When the RDT team tried to supplement these datasets by contacting the relevant departments of the fifty largest cities to request yearly counts on houselessness,

only about 54% responded and, of those that did, none were able to provide the information requested—indeed, most cities referred the RDT team back to the PIT counts collected by their corresponding CoCs. Given the highly localized nature of houselessness policies, the absence of publicly available, geographically disaggregated racial and ethnic data makes it difficult to craft and evaluate the rules, laws, and ordinances that may have the greatest impact on racial and ethnic inequities in the houselessness context.

### 3. Race and Ethnicity Data Are Not Consistently Available across Time

None of the repositories discussed above consistently or comprehensively collected racial and ethnic demographic data over time. While PIT data have generally been available since 2007, data on race and ethnicity were not included in PIT counts until 2015. HMIS data, by contrast, included race and ethnicity information only between 2007 and 2017. Census data are only collected once every ten years, and only included comprehensive racial and ethnic data in 2010, following partial availability of racial and ethnic data in 2000 and 1990. Finally, racial and ethnic data from the ACS have been available on a yearly basis since 2006, but are not publicly accessible. Figure 12 demonstrates the scarcity of longitudinal data on houselessness and the inconsistency across data sources of the time periods for which race and ethnicity data are available. The absence of such information precludes a comprehensive picture of racial and ethnic disparities over time, which in turn prevents researchers from evaluating the effectiveness of policies intended to promote racial equity.



**Figure 12: Timeline of Data Availability by Race/Ethnicity of Four main Repositories of Data on Houselessness**

## C. The Challenges of Obtaining Arrest and Police Violence Data by Race and Ethnicity

The RDT team also identified several critical problems with major national repositories of racial and ethnic data regarding criminal arrests and police violence. For this study, the RDT tried to collect data on (1) overall arrests, (2) arrests for murder and non-negligent manslaughter, and (3) police use of violence. As with the houselessness databases, these sources varied significantly in their methodologies, had many gaps, lacked longitudinal data, and were not sufficiently disaggregated.

### 1. Datasets Vary in Collection Methodologies

#### a) Arrest Data by Race and Ethnicity

The RDT team examined three national datasets with racial and ethnic data on arrests: the Uniform Crime Reporting (UCR) Program from the FBI, the Bureau of Justice Statistics, and the National Crime Victimization Survey (NCVS), each of which has its own methodologies and particular limitations regarding race and ethnicity data.<sup>108</sup>

The most comprehensive source of racial and ethnic data for arrests in the United States is the UCR. Agencies voluntarily report and submit their data to the federal UCR through a state Uniform Crime Reporting System or directly to the federal UCR via the National Incident Based Reporting System (NIBRS). UCR data are reported at the national, state, and agency level. Information for both instances of arrest (murder and non-negligent manslaughter) was available between 1985 and 2019 through the Crime Data Explorer (CDE).<sup>109</sup> The UCR does not employ the OMB categories; data are disaggregated for race but not ethnicity,<sup>110</sup> and the category of “Native Hawaiian” does not include “other Pacific Islanders.”<sup>111</sup>

A second source of race and ethnicity data on arrests is the Bureau of Justice Statistics, which has data on murder, non-negligent manslaughter, and thirty-one other offenses by race between 1980 and 2014. Like UCR data, information from the Bureau of Justice Statistics is available at the national, state, and agency levels. These data are also non-compliant with the OMB standards, as the only racial categories available are white, Black, AIAN and Asian Pacific Islander (“API”), and no data on ethnicity is included.

A third source of data on arrests is the NCVS, a study that has been administered yearly in the United States since 1973<sup>112</sup> by the Census Bureau on behalf of the Bureau of Justice Statistics. This survey is presented to a nationally representative sample of approximately 169,000 people ages 12 or older in the United States. The survey asks about reported and unreported incidents of crime, why some incidents were not reported, the contexts of these incidents, experiences with the criminal legal system, self-protective measures used, and substance use in the past six months. For this study, the race and ethnicity of victims per type of crime

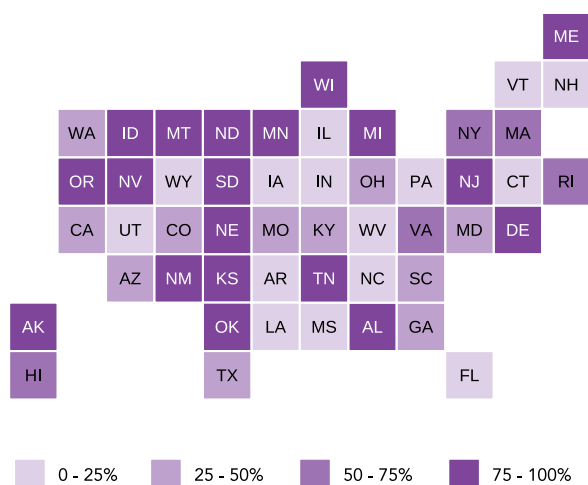


was collected from 2005 through 2019, but only at the national level. The only categories used for the data were white, Black, Hispanic, and other.<sup>113</sup>

The RDT’s examination of these datasets revealed particularly significant gaps in the availability of data regarding the ethnicity of people who are arrested, so the RDT team attempted to manually collect ethnicity data by searching law enforcement agencies’ websites and states’ websites. Only 32% of states reported data on ethnicity locally for all arrests and for non-negligent manslaughter arrests. The absence of ethnicity data was worse in some places than others: such data were available in 29% of states in the South, 33% of states in the Midwest, 38% of states in the West, and 56% of states in the Northeast. The lack of available ethnicity data regarding arrests precludes tailored policy responses to ethnic inequities in the criminal legal system.

**b) Police Violence Data by Race and Ethnicity**

Racial and ethnic data on police violence are scarce, and there is no nationwide repository of such information. While the Deaths in Custody Reporting Act of 2014 imposes financial penalties on states that do not comply with certain police violence data reporting requirements, there has been no thorough enforcement of the Act to date.<sup>114</sup> The FBI began collecting data on reporting compliance in 2019. As Figure 13 illustrates, the percentage of law enforcement agencies that reported police violence data per state in 2021 was very low: 6,543 out of 18,514 federal, state, local, and tribal law enforcement agencies submitted data (approximately 35%). These data included information on the race of the person against whom the police used violence, officer information (including race), and incident information. However, the only data publicly available to date are the number of agencies reporting and number of incidents reported for each; data on race and ethnicity remain unavailable.<sup>115</sup>



**Figure 13: Percentage of Law Enforcement Agencies That Reported Police Violence Per State in 2021**

Some states have their own requirements for reporting racial and ethnic data on police violence, but such piecemeal efforts are insufficient to provide comprehensive information about racism in policing. Data from the National Conference of State Legislatures shows that, at the state level, at least twenty-one states require some sort of data collection on law enforcement use-of-force incidents<sup>116</sup> but only eight of these states require that the data be publicly reported, only fifteen states specify the need to collect data on race, and only eleven require data on the ethnicity of victims.<sup>117</sup> Moreover, definitions of police use of violence and type of data collected vary greatly by state, making the data difficult to compare or aggregate. For example, only sixteen states collect race and ethnicity data on officer-involved deaths.<sup>118</sup>

In the absence of a reliable nationwide data repository, the Mapping Police Violence Organization is the most comprehensive dataset that collects data on police violence, including racial and ethnic data, but it, too, comes with important limitations. This database includes over 9,000 killings by police nationwide between 2013 and 2020, based on data compiled from a variety of sources, including (1) police use-of-force data collection programs in the small set of states that report publicly; (2) nationwide data from the Fatal Encounters database, a crowdsourced database on police killings; and (3) searches in social media, obituaries, criminal record databases, police records, and other sources. This approach allows the Mapping Police Violence Organization to identify and report the race and ethnicity of 90% of the victims of police violence.<sup>119</sup> While deeply impressive in its scope, this database relies on incomplete and often unofficial sources of information. It also generally follows the OMB categories, except that the category of “Pacific Islanders” omits “Hawaiian,” and the “Hispanic” category does not include “Latino” in its name.

Other non-governmental agency groups assemble and analyze race and ethnicity data on policing, but their repositories are also incomplete. The Police Scorecard, for example, collects data on police violence, accountability, racial bias, and policing outcomes for over 16,000 municipal and county law enforcement agencies in the United States. The data are collected from police arrests, personnel, funding, incarceration rates, and homicide clearance rates from official federal and state databases such as the UCR, the Bureau of Justice Statistics’ Annual Survey of Jails, the U.S. Census Bureau’s Survey of State and Local Government Finances, and the California Department of Justice’s Open Justice database. This information is complemented with agency publications and media reports, including the data from the Mapping Police Violence dataset. Each agency is assigned a score by focusing on a number of criteria. Some of their measures take into account race, specifically the variables: “racial disparities in deadly force,” “racial disparities in drug arrests,” “police violence by race,” and “percent of homicides unsolved by race.” While these scorecards are an important indication of agency performance, the data on race are not disaggregated beyond the state level and are not provided in cumulative counts. Moreover, as with other private organizations, the Police

Scorecard relies on sources of information that are incomplete, and it can only create a database that is as complete as its sources.

Finally, the *Washington Post* has a database of fatal shootings by a police officer in the line of duty since January 1, 2015, which contains some race and ethnicity information. The data are collected by looking at local news reports, law enforcement websites and social media, and by monitoring independent databases such as “Killed by Police” and “Fatal Encounters.”<sup>120</sup> The categories of race and ethnicity are mostly consistent with the OMB regulations, but the category of “Hawaiian and other Pacific Islander” is missing. While the *Washington Post’s* database is useful, it is also incomplete, as it does not include any police shootings or incidents of police violence that do not attract media attention.

In sum, racial and ethnic data concerning police violence are woefully incomplete and missing, hindering the creation of policies that effectively address racialized policing.

## **2. Longitudinal Data Are Scarce**

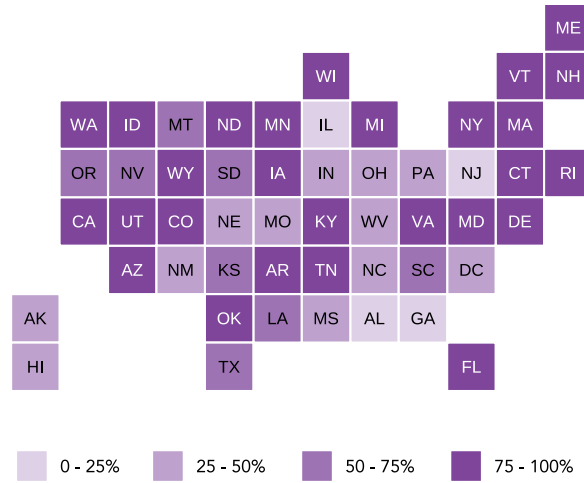
Longitudinal data by race are generally available for the past several decades for criminal arrests, but are essentially nonexistent for police violence.

The longitudinal arrest data come with several caveats. The UCR has longitudinal data on arrests with racial data (but no information on ethnicity) that spans from 1985 to 2019. The Bureau of Justice Statistics provides racial (but, again, not ethnic) data from 1980 to 2014, but with many gaps in reporting from law enforcement agencies. The NCVS has longitudinal data on race and ethnicity from 1973 to 2019, but this is also of limited value because the racial and ethnic categories it utilizes have changed over time. Moreover, because the NCVS focuses on self-reported victimization data, it uses a very different collection methodology than the prevalence counts<sup>121</sup> from UCR and the Bureau of Justice Statistics. Finally, all these sources of information use different race and ethnicity categories. As a result, it is harder to compare, complement, or impute data between data sources because they do not represent the same groups of people.

Efforts to collect race and ethnicity data on police violence are too recent to provide meaningful longitudinal data. Mapping Police Violence has tracked police violence data since 2013, and the *Washington Post* has tracked such data since 2015. Nationwide data before those dates is unavailable.<sup>122</sup>

## **3. The Level of Race and Ethnicity Data Reporting Varies Greatly from State to State**

Crime-related data submission to the federal government is voluntary, so many agencies simply do not submit data, including race and ethnicity data, to the UCR. As a consequence, race and ethnicity data submission levels vary greatly from region to region and from year to year. Figure 14 shows the percentage of agencies reporting

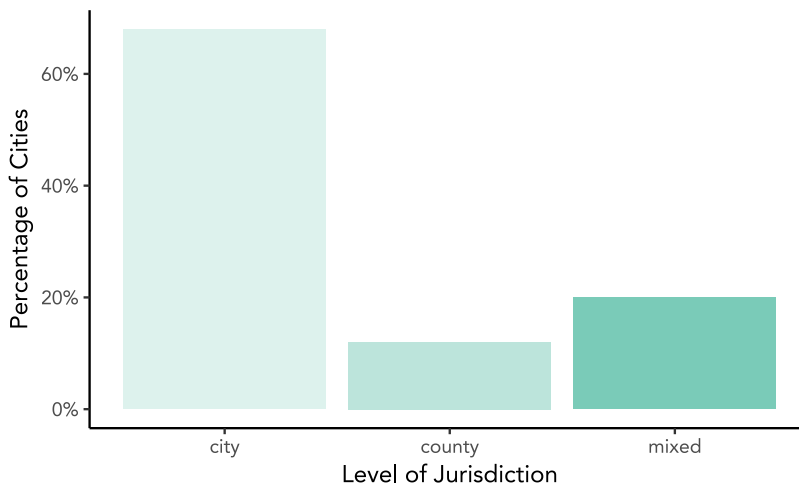


**Figure 14: Percentage of Law Enforcement Agencies That Reported Per State in 2019**

per state in 2019. The differences are stark: while 100% of agencies reported race and ethnicity data in Connecticut, less than 0.1% of agencies provided that data in Illinois. Year-to-year collection also varies greatly; for instance, the city of Boston did not submit race and ethnicity data in 2019 (for the most recent dataset) but did submit such data for 2018.

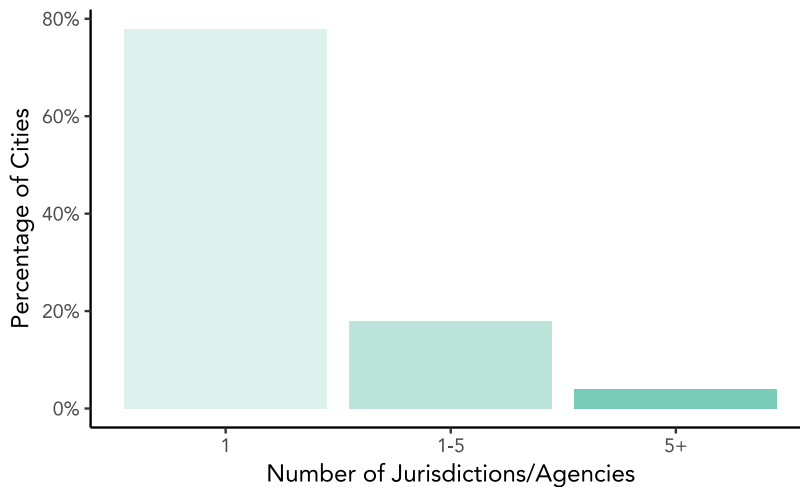
**4. Geographic Boundaries for Data Collection Often Overlap**

There are 18,000 different law enforcement agencies nationwide, which greatly complicates efforts to systematically collect racial and ethnic data on arrests and police violence. These agencies operate at different units of geography, including city, county, and regional levels. Figure 15 demonstrates how law enforcement jurisdictions correspond with various units of geography, including city, county, and regional jurisdictions among the fifty largest cities in the United States.



**Figure 15: Type of Jurisdiction(s) Operating in the 50 Largest Cities in the U.S.**

While a majority of cities have one law enforcement agency operating at the city level, at least 20% of them have more than two agencies that have the authority to arrest individuals in the city. This creates agency overlap, where multiple agencies engage in policing within one jurisdiction. For example, in New York City, the New York Police Department (NYPD) and New York City Transit can both make arrests, which they separately report to the federal government. Similar overlap exists with housing authorities and university-based police departments. Some cities also rely on county sheriff's offices for some or all of their policing. In Los Angeles, the Los Angeles Police Department (LAPD) and Los Angeles County Sheriff Department (LASD) are separate entities; the former is for the city of Los Angeles, but the LASD policed public hospitals, nine community colleges, and public transit (until policing of public transit transitioned to LAPD in 2017). Figure 16 shows the percentage of overlapping law enforcement agencies that operate in the fifty largest cities of the United States.



**Figure 16: Number of Law Enforcement Agencies Overlapping in the 50 Largest Cities in the U.S.**

Agency overlap complicates racial and ethnic data collection. In attempting to collect such data from police jurisdictions for the fifty largest cities in the United States, the RDT team identified a series of problems that have been corroborated in other studies.<sup>123</sup> When two or more overlapping law enforcement entities have separate collecting repositories, separate jurisdictions, and potentially different collection methodologies, data on race and ethnicity are likely to be incomplete. Just obtaining race and ethnicity data from the city police force, for example, might miss data on arrests made by transit, county, and university police forces. This is compounded by the fact that policing jurisdictions change over time. Moreover, no geocoded national data exist that would allow researchers to identify the number of police forces operating within a jurisdiction—and to assemble their race and ethnicity data in a way that aligns with geographic boundaries. Nor is it always clear which law enforcement agencies have the power to arrest within city limits, even if they do patrol a city. Additionally, while city level authorities will enact policy on

crime for their jurisdiction, they will not be able to evaluate the exact prevalence of the issue by race and ethnicity if the data are collected at another level. This array of challenges precludes the accurate measurement of racialized law enforcement at the local level, as well as targeted public policies that might redress it.

#### **D. Summary**

The RDT's first wave of data collection efforts confirms that existing datasets on houselessness, arrests, and police violence that report counts by race and ethnicity do not provide sufficient information to policymakers and advocates interested in combating racial inequities, particularly at the local level. Like the state-reported data that informed the CRDT, the local, regional, and state data on houselessness, arrests, and police violence are incomplete, uncoordinated, and unreliable. Existing datasets cannot be used side by side to try to fill these gaps because their methodologies are too varied. Due to the organization and structure of existing data collection entities, data disaggregated at the local level (where many policy choices are made) are unavailable.

The work of the CRDT and RDT, analyzed together, affirms that state and local data collection infrastructure must be financed and strengthened, and that such systems should report to a single, standardized, nationwide system of data collection and reporting by race and ethnicity.

# V. POLICY RECOMMENDATIONS



---

Based on the challenges the CRDT and RDT teams faced in collecting race and ethnicity data on COVID-19 outcomes, houselessness, criminal arrests, and police violence, we make the following policy recommendations concerning racial and ethnic data collection and reporting in the United States:

## **1. Centralize and standardize racial and ethnic demographic data collection and reporting across critical issue areas.**

Many of the challenges the CRDT and RDT faced in trying to collect the most accurate and complete race and ethnicity data possible could have been largely avoided if states and local entities reported their data to a national source that could present the data in a standardized way. Such a system would ensure consistency across local and state entities in terms of how data are collected and reported,<sup>124</sup> what data are collected and reported, and how often the data are updated, which would allow robust data analysis and comparisons across geography and time. Federal leadership is best suited to create such a system. In order to understand the full ecosystem of racial inequity and subordination, this system should include race and ethnicity data in key policy areas such as health, housing, employment, education, the criminal legal system, and the environment.

## **2. Use existing federal race and ethnicity standards as a starting point, and regularly reevaluate and amend such standards.**

The CRDT and RDT teams found that, across different jurisdictions and issue areas, the OMB categories are ubiquitous and are often treated as default standards even when not required. To maximize efficiency and increase the likelihood of compliance, existing federal standards, including the OMB racial categories, should

be the starting point for a centralized system of race and ethnicity data collection. The federal government should review and amend these standards immediately, and continue to do so regularly with input from scholars, community members, and advocates. This process should include a reevaluation of the OMB racial and ethnic categories.

### **3. Create an oversight board to monitor how the government uses race and ethnicity data.**

Race and ethnicity data, like all information, are not always used appropriately. To mitigate that risk, an external oversight board should monitor the creation and use of a centralized and standardized data collection system. This board would ensure that the government's data practices comply with existing laws and are responsible, ethical, and equitable. The board's duties should include making sure that data concerning racial inequities is presented with appropriate context about structural racism as a root cause,<sup>125</sup> so that such disparities are less likely to be used to discriminate or promote racist ideas and stereotypes. This board should include community partners and advocates who are most likely to be impacted by or familiar with counteracting racist abuses and biases associated with data.<sup>126</sup>

### **4. Incentivize consistent and timely state and local participation through adequate funding.**

The CRDT and RDT teams' efforts were hampered by the voluntary and haphazard nature of local and state data reporting. The RDT team also found that race and ethnicity data quality was often subject to the arbitrary geographic boundaries and practices of service providers, indicating a need for stronger data infrastructure at the local and state government levels. The federal government should provide financial incentives to states and localities to collect and report racial demographic data in a consistent and timely manner to a centralized source. Conditional funding should be tied to critical state needs, such as infrastructure development, to adequately incentivize participation. Additional, separate funding should also be provided to all entities that participate to support data administration infrastructure. As part of this system, states and local entities should report data that they are already collecting *and* be incentivized with more funds to increase their data collection efforts to fill existing data gaps. Additional incentives to collect and report racial and ethnic data should be provided by way of research grants to entities that are committed to studying and counteracting racial and ethnic inequities that these data collection efforts reveal.

### **5. Monitor noncompliance.**

Public visibility and accountability can complement financial incentives to motivate participation in a centralized and standardized race and ethnicity data



collection system. The aforementioned oversight board should monitor the participation of entities that collect (or should be collecting) data. Data reporting dashboards, websites, reports, and similar platforms associated with this centralized data collection system should explicitly note which entities have chosen not to participate.

## **6. Make data available at the national level, and also disaggregate at the state and local levels.**

Race and ethnicity data should be aggregated at the national level, but not at the expense of information about local and state-specific trends, because it is also necessary to understand local and state differences in how policies are implemented and the impacts they have. Data should be collected and reported in a manner that provides researchers and policymakers with the option of seeing aggregated data at the national level and disaggregated data at the local and state level.

## **7. Encourage and facilitate more granular race and ethnicity reporting.**

The CRDT and RDT teams' efforts to analyze racial and ethnic inequities were limited by the amount of granularity reflected in the available data. Entities should be encouraged to collect data at more granular levels than the OMB racial and ethnic categories to reflect the ethnic and racial makeup of their particular jurisdictions, and to reveal inequities between subpopulations.<sup>127</sup> Specifically, states should collect data about additional categories that are not included in the OMB (as relevant to their populations), but which can be collapsed into the OMB categories (if necessary at the national level).<sup>128</sup> This approach "can provide the level of detail needed to improve quality at the local level while providing standardized data to assess national progress."<sup>129</sup> More granular reporting also ensures that, if larger racial categories are changed over time, the original data are reported with sufficient detail such that they can be reconfigured to fit within a new reporting scheme, preserving longitudinal data.

## **8. Report race and ethnicity data as intersecting measures.**

Although "dominant conceptions of discrimination condition us to think about subordination as disadvantage occurring along a single categorical axis,"<sup>130</sup> demographic characteristics, such as race and gender, are not "mutually exclusive categories of experience and analysis."<sup>131</sup> An antiracist approach to data collection must account for the complexity and nuance of race, ethnicity, and (as discussed in the following recommendation below) additional characteristics and experiences. The CRDT and RDT teams' data collection efforts were often hampered by reporting styles that failed to account for the fact that people can be part of multiple racial or ethnic groups. Race and ethnicity data should be collected and reported

as intersecting characteristics in order to provide the most clarity and flexibility regarding the breakdown of ethnicity and race. For example, it is not enough to know how many people who experience homelessness separately fall into the Black and Hispanic/Latino/a/e/x categories. Reporting entities should specify how many of those who fall into the Black category *also* fall into the Hispanic/Latino/a/e/x category, and how many do not. Reporting entities must provide more detail about the breakdowns across all the race and ethnicity categories, including those people who are part of multiple racial or ethnic groups. This system should not limit the number of ethnicities or races that may apply to each person.

### **9. Collect and report data across additional characteristics and variables.**

Race and ethnicity data should be disaggregated by additional characteristics and variables in order to better understand and address the experiences of people who are subjected to multiple simultaneous forms of oppression based on their racialized, gendered, and otherwise minoritized identities. This additional information, including but not limited to sex, gender identity, sexual orientation, age, disability, religion, income, linguistic usage, educational attainment, socioeconomic status, “street race,”<sup>132</sup> and national origin, will enable more robust intersectional research.

### **10. Tailor privacy-related data suppression practices to the realities of each dataset.**

Data suppression is sometimes a useful method to protect individual privacy, such as when a number is so small that it risks disclosing the identity of those it pertains to. However, suppression should only be implemented when there is a real privacy concern. Small numbers may need to be suppressed in local or state databases, but may not need to be suppressed when they are aggregated at the national level. Accordingly, state and local entities that collect race and ethnicity data should remove any identifying information, such as names and birth dates, before reporting the data to a national database. Those entities should then report all disaggregated data to a national database through an encrypted platform, and the national database should decide whether suppression of small numbers is necessary at the national level. State and local entities should, of course, continue to suppress problematically small numbers for their own, publicly-facing reporting. Reporting entities should not be permitted to omit or merge racial and ethnic categories in reports to the national platform as an alternative way to address privacy concerns, and should be discouraged from omitting or merging categories in their own data reports.

**11. Make data publicly available and accessible.**

Public information requests and similar processes are energy-intensive, slow, and cumbersome. These procedures are simply not feasible when researchers and policy makers need to obtain data quickly to respond to public inequities, as was the case for the CRDT team. For this reason, apart from data that must be suppressed or protected due to privacy or serious confidentiality concerns, racial and ethnic data should be made freely, publicly, and easily accessible for use by advocates, scholars, policymakers, and others. All such data should contain explicit disclaimers and guidance on best practices concerning any gaps and limitations associated with them.

# VI. POLICY AND LEGAL CONSIDERATIONS REGARDING RACIAL AND ETHNIC DATA COLLECTION



---

In this Section, we address several policy and legal considerations regarding the collection and reporting of racial and ethnic data that helped shape our policy recommendations. These considerations include the risk of data misuse, legal limitations on federal mandates, and the shortcomings of the OMB categories.

## A. Potential Mischaracterizations and Misuses of Racial and Ethnic Data

Concerns about potential mischaracterizations and misuses of racial and ethnic data should not halt the collection and reporting of such data. “Some scholarly and civic leaders believe that measuring [racial inequities] promotes social divisions and fuels a mistaken perception that race is a biological concept.”<sup>133</sup> Of course, there is always a risk that information will be used incorrectly. Data about racial and ethnic inequities can be misinterpreted, misused, or politicized<sup>134</sup> to fuel racist stereotypes<sup>135</sup> and falsely justify racist policies.<sup>136</sup> In a “vacuum,” such data can reinforce socially constructed racial categories and mischaracterize the causes of racial inequities.<sup>137</sup> Racial and ethnic data at local levels can also at times be misused to enforce “‘territorial stigmatization,’ whereby resource-deprived neighborhoods suffer from ‘blemish of place’ and are thought to be ‘composed essentially of poor people, minorities and foreigners,’ many of whom have already been marginalized by the broader society.”<sup>138</sup> People of color may also, understandably, be wary of government efforts to track their race and identity for fear of misuse of such data, based on prior abuse and oppression.<sup>139</sup>

We need racial and ethnic data, however, to ensure that the “groups suffering the worst receive the most attention, treatment, and resources.”<sup>140</sup> Refusing to examine and measure racial and ethnic inequities will “[a]t best . . . preserve the status quo.”<sup>141</sup>

Moreover, racial and ethnic data are not needed to engage in racist discrimination—predictive policing algorithms, for example, use factors like zip code that act as proxies for race.<sup>142</sup> Other examples include prison-based gerrymandering<sup>143</sup> and the racist effects of the use of “big data” and artificial intelligence in credit scores.<sup>144</sup> The antiracist solution to racial inequity is not to avoid tracking racial inequities for fear of their misuse, but to actively educate the public about inequities and the policies that contribute to them.

When racial data are used and contextualized appropriately, they provide critical information about experiences of racism that can inform advocacy and policymaking. We need race and ethnicity data, not concerning any one societal problem in isolation, but across major policy areas, in order to understand the full ecosystem of racial and ethnic subordination and oppression. Racial inequities must be studied alongside resource disparities and the histories and policies that contributed to them, so that the data are effectively wielded toward antiracist policy change. For example, when racial disparities in the criminal legal system are understood in the context of racially targeted and disproportionate policing, prosecution, and sentencing, racial data can help inform the corresponding policy interventions. Ultimately, data collection will allow us to gain a better understanding, not of race, but of *racism*.

## B. Legal Considerations

Data collection is subject to several federal laws that have constructed procedural safeguards to protect individuals’ privacy, prevent discrimination, and minimize burdensome requirements, among other protections.<sup>145</sup> Some states also “impose restraints on when and how such data may be collected.”<sup>146</sup> These laws are not barriers to racial and ethnic data collection; rather, they help ensure that data are collected, stored, and shared in ways that protect peoples’ rights.

The most important legal consideration in designing a single, standardized, nationwide system of data collection and reporting by race and ethnicity is the constitutional limitation on the federal government’s ability to impose mandates on states. The Tenth Amendment to the U.S. Constitution provides that “the powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.” As a result, the “Federal Government may not compel the States to implement, by legislation or executive action, federal regulatory programs.”<sup>147</sup>

Federal mandates do not have a settled definition<sup>148</sup> and take a variety of forms—they may be funded or unfunded, and may consist of direct orders, generally applicable regulations, and conditions of assistance, among other formats.<sup>149</sup> The constitutionality of a federal mandate depends largely on the source of power that is used to justify the mandate (such as Congress’s commerce power<sup>150</sup> and spending power<sup>151</sup>), but on the whole, this remains a murky area of law.<sup>152</sup> Additionally,

“[d]uring the last two decades, the Supreme Court has reopened the debate on whether the Tenth Amendment imposes any limits on the authority of the federal government to subject states and their subdivisions to federal regulations.”<sup>153</sup> In short, while the federal government often employs mandates,<sup>154</sup> it is difficult to predict how each mandate will fare in court.<sup>155</sup>

The clearest course, then, is to use conditions of federal financial assistance that are rooted in Congress’s spending power to incentivize, rather than require, participation. When invoking its spending power, “Congress may attach conditions on the receipt of federal funds,” so long as this power is used “in pursuit of the general welfare” and the conditions are unambiguous and “reasonably related to the federal interest in particular national projects or programs.”<sup>156</sup> In practice, courts are heavily deferential to Congress in determining whether these requirements are met.<sup>157</sup> Thus, for example, in *South Dakota v. Dole*, the National Minimum Drinking Age Act, “which provided that federal highway funds otherwise payable to a state would be withheld if that state did not raise the minimum drinking age to twenty-one,” was not unconstitutional.<sup>158</sup> For these reasons, the federal government should incentivize participation of state and local entities in a standardized and centralized racial and ethnic data collection and reporting system through the use of conditional, highly desirable funding.

### **C. Limitations of the Federal Office of Management and Budget Categories**

This Report’s recommendation to use the OMB categories as a starting point for standardization of racial and ethnic data collection and reporting is not to suggest that these categories are ideal or comprehensive. The OMB categories, which “were honed for bureaucratic and political purposes,”<sup>159</sup> are rightfully critiqued for many reasons, including that they do not reflect the diversity of the U.S. population,<sup>160</sup> and improperly lump together many distinct racial and ethnic experiences.<sup>161</sup> The current OMB racial and ethnic categories are clearly lacking, as more and more people are choosing to check boxes for “Other Race” in data collection efforts.<sup>162</sup> In the 2020 Census, “[t]he Some Other Race population was the second-largest alone or in combination race group, comprising 15.1% of the total population.”<sup>163</sup> Heavy reliance on the “other race” category can lead to data confusion and obscures the true extent of inequities.<sup>164</sup>

Despite the OMB categories’ deficiencies, they have become default categories for many federal, state, and local data collection efforts. So much of existing racial and ethnic data are organized according to the OMB categories that their wholesale replacement could make vast amounts of longitudinal data unusable.<sup>165</sup> The OMB itself “encourages,” but does not require, “additional granularity where it is supported by sample size and as long as the additional detail can be aggregated back to the minimum standard set of race and ethnicity categories.”<sup>166</sup> While the

**VI. POLICY AND LEGAL CONSIDERATIONS REGARDING  
RACIAL AND ETHNIC DATA COLLECTION**

---

OMB categories “are political constructs,” they “nonetheless help to code past and present forms of inequality and discrimination,” which are also significantly shaped by politics.<sup>167</sup> As a result, the conversation among scholars and advocates regarding nationwide data collection generally calls for amending and supplementing, rather than replacing, the OMB categories.<sup>168</sup>

We encourage the continued critique of the OMB categories and echo scholars and advocates that have called for their supplementation and amendment.<sup>169</sup> This issue will be the subject of a future Center project.

# VII. CONCLUSION



---

The CRDT and RDT teams' experiences collecting racial and ethnic data across several key issue areas reveal major deficiencies regarding the state of racial and ethnic demographic data collection and reporting in the United States. Existing data collection efforts are riddled with gaps and errors, including missing and incomplete data, insufficiently disaggregated data, lack of meaningful longitudinal data, infrequently updated data, non-standardized methodologies, and other problems. These data quality issues lead to underestimations of racial inequities, obscure evidence of racism, prevent cross-jurisdictional analysis, and, ultimately, hinder evidence-based antiracist policymaking. The experiences of the CRDT and RDT teams underscore the need for a single standardized and nationwide system of data collection and reporting by race and ethnicity across important issue areas. We must standardize, centralize, and bolster our race and ethnicity data infrastructure and practices in order to strengthen the process of analyzing, contextualizing, and dismantling racism.



# VIII. APPENDICES



**Appendix 1. Summary of Jurisdictions Reporting Race Data Over Time, April 2020 – February 2021**

Month-yr	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21
<b>Exact data collection period (start and end date)*</b>	4/12/2020 to 4/29/2020	4/29/2020 to 5/31/2020	5/31/2020 to 6/28/2020 for cases and deaths; 6/17/2020 to 6/28/2020 for hospitalizations	6/28/2020 to 7/29/2020	7/29/2020 to 8/30/2020	8/30/2020 to 9/30/2020	9/30/2020 to 10/28/2020	10/28/2020 to 11/29/2020	11/29/2020 to 12/30/2020	12/30/2020 to 1/31/2021	1/31/2021 to 2/28/2021
<b># Jurisdictions reporting new race data for cases (out of N=56)</b>	44	48	49	50	51	51	51	51	51	51	51
<b># Jurisdictions reporting new race data for deaths (out of N=56)</b>	37	43	47	45	48	49***	49***	50	50	51	51
<b># Jurisdictions reporting new race data for hospitalizations (out of N=56)</b>	N/A	N/A	16	16	17	19	20	21	23	23	23
<b>Jurisdictions reporting no new race data for cases</b>	AS, GU, LA, NE, NV, NJ, NY, ND, MP, PR, SD, VI	AS, LA, NE, NY, ND, MP, PR, VI	AS, LA, NY, ND, MP, PR, VI	AS, NY, ND, MP, PR, VI	AS, NY, MP, PR, VI	AS, NY, MP, PR, VI	AS, NY, MP, PR, VI	AS, NY, MP, PR, VI	AS, NY, MP, PR, VI	AS, NY, MP, PR, VI	AS, NY, MP, PR, VI
<b>Jurisdictions reporting no new race data for deaths</b>	AS, FL, GU, HI, IA, ME, MT, NE, NV, NM, ND, MP, OR, PR, SD, UT, VI, WV, WY	AS, GG, HI, ME, NE, NM, ND, MP, PR, SD, UT, VI, WY	AS, GU, HI, ND, MP, PR, SD, VI, WV	AS, GU, HI, NM, ND, MP, PR, SD, VI, WV	AS, GU, HI, ND, MP, PR, VI, WV	AS, GU, ND, MP, PR, SD, VI, WV	AS, GU, ND, MP, PR, VT***, VI	AS, GU, ND, MP, PR, VI	AS, GU, ND, MP, PR, VI	AS, ND, MP, PR, VI	AS, ND, MP, PR, VI
<b>Jurisdictions reporting no new race data for hospitalizations</b>	N/A	N/A	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, HI, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, MT, NV, NM, NY, NC, PA, PR, SC, SD, TN, TX, VT, VI, WV, WI, WY	AL, AR, AS, CA, CO, CT, DE, DC, GE, GU, HI, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, MT, NV, NM, NY, NC, ND, MP, OK, PA, PR, SC, SD, TN, TX, VT, VI, WV, WI, WY	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, HI, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, MT, NV, NM, NY, NC, ND, MP, OK, PA, PR, SC, TN, TX, VT, VI, WV, WI, WY	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, HI, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, NV, NC, ND, MP, OK, PA, PR, SC, TN, TX, VT, VI, WV, WI, WY	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, HI, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, NV, NC, ND, MP, OK, PA, PR, SC, TN, TX, VT, VI, WV, WI, WY	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, HI, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, NV, NC, ND, MP, OK, PA, PR, TN, TX, VT, VI, WV, WY	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, NV, NC, ND, MP, OK, PA, PR, TN, TX, VT, VI, WV, WY	AL, AR, AS, CA, CO, CT, DE, DC, GA, GU, ID, IL, IN, IA, KY, LA, MD, MI, MS, MO, NV, NC, ND, MP, OK, PA, PR, TN, TX, VT, VI, WV, WY	AL, AR, AS, CA, CO, CT, DE, DC, DE, GU, IA, ID, IL, IN, KY, LA, MD, MI, MS, MO, NV, NC, ND, MP, OK, PA, PR, TN, TX, VT, VI, WV, WY

Appendix 1 continued

Proportion of cumulative cases with no race information out of total cases nationwide**	0.635	0.473	0.444	0.421	0.387	0.362	0.345	0.348	0.345	0.337	0.332
Proportion of cumulative deaths nationwide with no race information out total deaths nationwide**	0.325	0.116	0.070	0.049	0.038	0.038	0.021	0.031	0.048	0.052	0.050
Proportion of cumulative hospitalizations with no race information out of cumulative/ever hospitalized nationwide**	N/A	N/A	0.506	0.518	0.508	0.505	0.351	0.364	0.342	0.340	0.262

Note:

- \* The COVID Racial Data Tracker collected data from states and territories on a twice weekly schedule beginning 4/12/2020 for cases and deaths and 6/17/2020 for hospitalizations. Data are not available from exact first and last day of each month.
- \*\* Proportions are calculated from total nationwide cases (confirmed and probable), deaths (confirmed and probable), and hospitalizations (cumulative hospitalized/ever hospitalized) reported in the COVID Tracking Project core data set. For hospitalizations, the COVID Tracking Project noted that only about two-thirds of states and territories reported data for the cumulative hospitalized/ever hospitalized metric. Proportions in this column exclude states and territories not reporting this metric and are not representative of the US as a whole.
- \*\*\* Vermont reported no new race data for deaths because there were 0 reported deaths statewide in this month.
- \*\*\*\* **Bold jurisdictions had reported racial data for this metric in previous months but no new race data in the current month.**

Appendix 2. CoC Participation in HUD’s Funding Program in 2016, 2013, and 2007 Compared to 2020

Year of comparison with 2020	CoCs missing
2016	<ul style="list-style-type: none"> <li>AR-508-Fort Smith CoC: No information available on this CoC.</li> <li>NY-504-Cattaraugus County CoC: CoC in operation.</li> <li>NY-516-Clinton County CoC: CoC absorbed by the Balance of State CoC.</li> <li>NY-607-Catholic Charities Community Services of Orange County: CoC in operation.</li> </ul>
2013	<ul style="list-style-type: none"> <li>NY-504-Cattaraugus County CoC: CoC in operation.</li> <li>NY-506-Fulton, Montgomery CoC: absorbed by the Balance of State CoC.</li> <li>NY-516-Clinton County CoC: absorbed by the Balance of State CoC.</li> <li>NY-607-Catholic Charities Community Services of Orange County CoC: CoC in operation.</li> </ul>
2007	<ul style="list-style-type: none"> <li>AR-502: No information available on this CoC.</li> <li>AR-504: Delta Hills CoC: No information available on this CoC after 2019.</li> <li>AR-506: No information available on this CoC.</li> <li>AR-508-Fort Smith CoC: No information available on this CoC.</li> <li>AR-509: No information available on this CoC.</li> <li>AR-510: No information available on this CoC.</li> <li>MA-501: No information available on this CoC.</li> <li>MA-514: No information available on this CoC.</li> <li>NY-504-Cattaraugus County CoC: CoC in operation.</li> <li>NY-516-Clinton County CoC: absorbed by the Balance of State CoC.</li> <li>NY-607-Catholic Charities Community Services of Orange County: CoC in operation.</li> <li>NY-609- Putnam County: absorbed by the Balance of State CoC.</li> </ul>

# IX. ENDNOTES



---

<sup>1</sup>Ibram X. Kendi, *How to Be an Antiracist* (New York: One World, 2019), 35.

<sup>2</sup>Racial and ethnic categories, such as those used for the U.S. Census, “emerge, or are negotiated, in . . . a political fashion—none exists in nature.” G. Cristina Mora and Michael Rodríguez-Muñiz, “Latinos, Race, and the American Future: A Response to Richard Alba’s ‘The Likely Persistence of a White Majority,’” *New Labor Forum* 26, no. 2 (Spring 2017): 41, <https://www.jstor.org/stable/26420068>.

<sup>3</sup>Trans-disciplinary “Race” Working Group, “Trans-Disciplinary Guidelines For Researching “Race,”” (University of New Mexico: Spring 2010), <https://race.unm.edu/about/race-research.html>.

<sup>4</sup>American Sociological Association, “The Importance of Collecting Data and Doing Social Scientific Research on Race,” (Washington, D.C.: 2003), [https://www.asanet.org/sites/default/files/savvy/images/press/docs/pdf/asa\\_race\\_statement.pdf](https://www.asanet.org/sites/default/files/savvy/images/press/docs/pdf/asa_race_statement.pdf).

<sup>5</sup>For the purpose of this report, we discuss “race” and “ethnicity” data as defined and categorized by the federal Office of Management and Budget (OMB), since its standards are adopted by many federal agencies, states, and localities. As discussed below in Section VI.C, those categories—including distinctions made between race and ethnicity—warrant deeper examination. While outside the scope of this report, potential reforms to the OMB categories will be the subject of a future Center project.

<sup>6</sup>Trans-disciplinary “Race” Working Group, “Trans-Disciplinary Guidelines.”

<sup>7</sup>American Sociological Association, “The Importance of Collecting Data.”

<sup>8</sup>Megan D. Douglas et al., “Variation in Reporting of the Race and Ethnicity of COVID-19 Cases and Deaths across US States: April 12, 2020, and November 9, 2020,” *American Journal of Public Health* 111, no. 6 (May 5, 2021): e2, <https://satcherinstitute.org/wp-content/uploads/2021/04/ePub-ajph.2021.306167.pdf>.

<sup>9</sup>Grant Makers in Health and National Committee for Quality Assurance, “Federal Action Is Needed to Improve Race and Ethnicity Data in Health Programs,” (New York: The Commonwealth Fund, 2021), 9–10, <https://www.gih.org/wp-content/uploads/2021/10/GIH-Commonwealth-Fund-federal-data-report-part-1.pdf>; National Research Council, “Introduction,” in *Improving Racial and Ethnic Data on Health: Report of a Workshop*, eds. Daniel Melnick and Edward Perrin (Washington, D.C.: National Academies Press, 2003), 4, <https://www.nap.edu/read/10833/chapter/2>.

<sup>10</sup>Ruth Perot and Mara Youdelman, “Racial, Ethnic, and Primary Language Data Collection in the Health Care System: An Assessment of Federal Policies and Practices” (New York: The Commonwealth Fund, September 2001), vii, [https://www.commonwealthfund.org/sites/default/files/documents/\\_\\_\\_media\\_files\\_publications\\_fund\\_report\\_2001\\_sep\\_racial\\_\\_ethnic\\_\\_and\\_primary\\_language\\_data\\_collection\\_in\\_the\\_health\\_care\\_system\\_\\_an\\_assessment\\_of\\_fed\\_perot\\_raciaethnic\\_492\\_pdf.pdf](https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_fund_report_2001_sep_racial__ethnic__and_primary_language_data_collection_in_the_health_care_system__an_assessment_of_fed_perot_raciaethnic_492_pdf.pdf).

<sup>11</sup>Perot and Youdelman, “Racial, Ethnic, and Primary Language Data Collection,” 21.

<sup>12</sup>Cheryl Ulmer et al., “Implementation,” in *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*, Institute of Medicine, (Washington, D.C.: National Academies Press, 2009), 153, <https://www.ncbi.nlm.nih.gov/books/NBK219753/>. As the CDC explains, in the public disease surveillance context, the “scope and nature of reporting requirements vary considerably by state, differing, for example, by the number of conditions required for reporting, time periods within which conditions must be reported, agencies to which reports must be submitted, and persons or sources required to report. Moreover, despite the legal requirements for reporting, adherence to and completeness of reporting also vary substantially by infectious disease agent, ranging from 6% to 90% for different common infectious conditions.” James D. Holt et al., “Legal Considerations,” *The CDC Field Epidemiology Manual*, (Centers for Disease Control and Prevention, December 13, 2018), <https://www.cdc.gov/eis/field-epi-manual/chapters/Legal.html>.

<sup>13</sup>The Boston University Center for Antiracist Research was founded in July of 2020, at which time some of those who were involved in the CRDT were formally hired by the Center.

<sup>14</sup>More information about the CRDT can be found at <https://covidtracking.com/race>.

<sup>15</sup>See, e.g., Grant Makers in Health and National Committee for Quality Assurance, “Federal Action Is Needed”; Perot and Youdelman, “Racial, Ethnic, and Primary Language Data Collection”; Allen Fremont and Nicole Lurie, “The Role of Racial and Ethnic Data Collection in Eliminating Disparities in Health Care,” in *Eliminating Health Disparities: Measurement and Data Needs* (Washington, D.C.: National Academies Press, 2004), <https://www.ncbi.nlm.nih.gov/books/NBK215740/>; David R. Williams and Toni D. Rucker, “Understanding and Addressing Racial Disparities in Health Care,” *Health Care Financial Review* 21, no. 4 (Summer 2000): 75–90, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194634/>; Institute of Medicine, “Data Collection and Monitoring,” in *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, eds. Brian Smedley et al. (Washington, D.C.: National Academies Press, 2003), <https://pubmed.ncbi.nlm.nih.gov/25032386/>.

<sup>16</sup>A foundational report by the Institute of Medicine to Congress in 2003 concluded that “[f]ederal leadership is needed to spearhead [racial and ethnic] data collection efforts,” which “should be accomplished using a standard racial/ethnic classification scheme.” Institute of Medicine, “Data Collection and Monitoring.” Subsequent IOM reports have made similar recommendations. Ulmer et al., “Implementation.”

<sup>17</sup>We use the term “Hispanic/Latino/a/e/x” to be as inclusive as possible and to take note of the various ways this group has been referred to by data collection entities. Federal entities often use the term “Hispanic.” “Latino” and “Latina” are broader terms, but are gender binary, so some use “Latinx” to avoid gendering the term. Recognizing that the “x” sound, as used in “Latinx,” is not part of Spanish language pronunciation, a newer trend is emerging to use “Latine.” For a detailed analysis of the categorizations of people in this group, see generally, Laura Gomez, *Inventing Latinos*, (New York: The New Press, 2020); G. Cristina Mora, Reuben Perez, and Nicholas Vargas, “Who Identifies as ‘Latinx’? The Generational Politics of Ethnoracial Labels,” *Social Forces* 100, no. 3 (March 2022): 7–8, <https://academic.oup.com/sf/advance-article-abstract/doi/10.1093/sf/soab011/6148737>; Salvador Vidal-Ortiz and Juliana Martínez, “Latinx Thoughts: Latinidad with an X,” *Latino Studies* 16 (2018): 384–95 (2018), <https://link.springer.com/article/10.1057/s41276-018-0137-8>; Rogelio Sáenz and Maria Cristina Morales, *Latinos in the United States: Diversity and Change* (Cambridge: Polity Press, 2015).

<sup>18</sup>U.S. Office of Management and Budget, “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity,” (1997), [https://obamawhitehouse.archives.gov/omb/fedreg\\_1997standards](https://obamawhitehouse.archives.gov/omb/fedreg_1997standards).

<sup>19</sup>The team selected these issue areas to (1) understand data that was spread across multiple policy areas and federal departments; (2) target areas of general policy import; and (3) target areas where the availability of geographically disaggregated state and local data are especially important, since housing and policing are areas of high local control.

<sup>20</sup>These policy recommendations are designed to create a strong foundation to overcome existing deficiencies in race and ethnicity data. For best practices regarding the *study* of race and ethnicity, we recommend that researchers consult the “Trans-Disciplinary Guidelines for Researching ‘Race,’” compiled by the Trans-disciplinary “Race” Working Group at the University of New Mexico, which can be accessed at <https://race.unm.edu/about/race-research.html>.

<sup>21</sup>Racism is experienced along many axes. For example, a person can personally identify as Latina or Afro-Latina, but be racialized by others as Black. Data collection concerning race and ethnicity often conflates how people identify and how they are racialized. The term “street race” refers to “how you believe other ‘Americans’ perceive your race at the level of the street.” Nancy López et al., “What’s Your ‘Street Race’? Leveraging Multidimensional Measures of Race and Intersectionality for Examining Physical and Mental Health Status among Latinxs,” *Sociology of Race and Ethnicity* 4, no. 1 (June 9, 2017): 50, <https://doi.org/10.1177/2332649217708798>. See Edward D. Vargas et al., “Critical ‘Street Race’ Praxis: Advancing the Measurement of Racial Discrimination among Diverse Latinx Communities in the U.S.,” *Critical Public Health* 31, no. 4 (2021), <https://www.tandfonline.com/doi/abs/10.1080/09581596.2019.1695040>.

<sup>22</sup>For a discussion of the racially essentializing nature of certain language about disparities in COVID-19 outcomes, see Coronasocial, “Death Rates, Deaths Per Capita, and Racial Essentialism” (June 19, 2020), <https://coronasocial.wordpress.com/2020/06/19/data-illiteracy-state-indifference-death/>.

<sup>23</sup>The current presidential administration’s COVID-19 Health Equity Task Force has made data equality a priority. David Raths, “Task Force Chair: Better Data Key to COVID Equity Response,” *Healthcare Innovation* (February 17, 2021), <https://www.hcinnovationgroup.com/policy-value-based-care/public-health/article/21210649/task-force-chair-better-data-key-to-covid-equity-response>; U.S. Department of Health and Human Services, “Readout of the Second COVID-19 Health Equity Task Force Meeting,” (April 13, 2021), <https://www.hhs.gov/about/news/2021/04/13/readout-second-covid-19-health-equity-task-force-meeting.html>; U.S. Department of Health and Human Services, “Readout of the Sixth Covid-19 Health Equity Task Force Meeting” (August 4, 2021), <https://www.hhs.gov/about/news/2021/08/04/readout-sixth-covid-19-health-equity-task-force-meeting.html>.

<sup>24</sup>See, e.g., Bethany L. Jones and Jonathan S. Jones, “Gov. Cuomo Is Wrong, Covid-19 Is Anything but an Equalizer,” *Washington Post*, April 5, 2020, <https://www.washingtonpost.com/outlook/2020/04/05/gov-cuomo-is-wrong-covid-19-is-anything-an-equalizer/>.

<sup>25</sup>Jones, “Gov. Cuomo Is Wrong.”; Yaryna Serkez, “These Charts Show That Your LockDown Experience Wasn’t Just About Luck,” *New York Times* (March 11, 2021), <https://www.nytimes.com/interactive/2021/03/11/opinion/covid-inequality-race-gender.html>.

<sup>26</sup>Jess Bidgood, “How Patchy COVID Data Hampered the Pandemic Response,” *Boston Globe* (February 1, 2022), <https://www.bostonglobe.com/2022/02/01/nation/how-patchy-covid-data-hampered-pandemic-response/>.

<sup>27</sup>Ibram X. Kendi, “Why Don’t We Know Who the Coronavirus Victims Are?,” *Atlantic* (April 1, 2020), <https://www.theatlantic.com/ideas/archive/2020/04/stop-looking-away-race-covid-19-victims/609250/>.

<sup>28</sup>So little COVID-19 race and ethnicity data were available from U.S. territories that, for ease of reference, we at times hereinafter refer to the jurisdictions included in the CRDT as “states.”

<sup>29</sup>Ibram X. Kendi, “We Still Don’t Know Who the Coronavirus’s Victims Are,” *Atlantic*, May 2, 2021, <https://www.theatlantic.com/ideas/archive/2021/05/we-still-dont-know-who-the-coronaviruss-victims-were/618776/>.

<sup>30</sup>Kendi, “We Still Don’t Know.” In the public health context, for example, “[s]urveilliance data, including on race/ethnicity, are used in mathematical modeling to assess the trajectory of illness among populations and to inform subsequent distribution of resources to affected communities.” Douglas et al., “Variation in Reporting of the Race and Ethnicity,” e2.

<sup>31</sup>Kendi, “Why Don’t We Know?”

<sup>32</sup>Kendi, “We Still Don’t Know.”

<sup>33</sup>Kendi, “We Still Don’t Know.”

<sup>34</sup>By “publicly available,” we mean data that is available to any member of the public via a government website, press release, press conference, or report, or by a simple, informal request. This does not include data which may become available through an application or review process. Indeed, for members of the public, government officials, and researchers, such processes are lengthy and cumbersome and unlikely to be used outside of longer, more involved research projects. Consequently, this report focuses on data that support government transparency, easy communication between government officials, and the ability to quickly and seamlessly evaluate the impact of public policies on racial inequities.

<sup>35</sup>“How Complete Are the CDC’s COVID-19 Case Surveillance Datasets for Race/Ethnicity at the State and County Levels?” *Satcher Institute* (May 24, 2021), [https://satcherinstitute.github.io/analysis/cdc\\_case\\_data](https://satcherinstitute.github.io/analysis/cdc_case_data). See “Disparities in Rates of COVID-19 Infection, Hospitalization, and Death by Race and Ethnicity,” *Office of the Assistant Secretary for Planning and Evaluation* (November 9, 2020), 21, <https://aspe.hhs.gov/reports/disparities-rates-covid-19-infection-hospitalization-death-race-ethnicity>.

<sup>36</sup>“COVID-19 Case Surveillance Public Use Data,” Centers for Disease Control and Prevention, December 21, 2021, <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf>; “COVID-19 Case Surveillance Restricted Access Detailed Data,” *Centers for Disease Control and Prevention* (Dec. 21, 2021), <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Restricted-Access-Detai/mbd7-r32t>.

<sup>37</sup>*E.g.*, Jeremy A.W. Gold, Jennifer DeCuir, and Jayme P. Coyle, “COVID-19 Case Surveillance: Trends in Person-Level Case Data Completeness, United States, April 5–September 30, 2020,” *Public Health Reports* 136, no. 4 (July 2021): 466–74, <https://journals.sagepub.com/doi/10.1177/00333549211006973?icid=int.sj-full-text.samecollection-articles.2>; Satcher Institute, “COVID-19 Case Surveillance Datasets.”

<sup>38</sup>Satcher Institute, “COVID-19 Case Surveillance Datasets.”

<sup>39</sup>Satcher Institute, “How Complete Are the CDC’s COVID-19 Case Surveillance Datasets.”

<sup>40</sup>Samantha Artiga, Latoya Hill, and Sweta Haldar, “COVID-19 Cases and Deaths by Race/Ethnicity: Current Data and Changes Over Time,” *KFF* (October 8, 2021), <https://www.kff.org/racial-equity-and-health-policy/issue-brief/covid-19-cases-and-deaths-by-race-ethnicity-current-data-and-changes-over-time/>.

<sup>41</sup>“What is Case Surveillance?” *Centers for Disease Control and Prevention* (June 11, 2021), <https://www.cdc.gov/nndss/about/index.html> (“States voluntarily inform CDC when a person meets certain criteria to become a case.”). NNDSS recommendations for case reporting COVID-19 can be found at <https://www.cdc.gov/coronavirus/2019-ncov/php/reporting-pui.html>.

<sup>42</sup>Gold et al., “Trends in Person-Level Case Data Completeness” (“We also identified important regional differences in the estimated sensitivity of case ascertainment by case notifications sent to CDC, which poses a challenge when assessing the national burden of disease.” (Sensitivity of case ascertainment here refers to the proportion of cases that were reported to NNDSS)).



<sup>43</sup>Nelson Adekoya, Benedict I. Truman, and Uhmed A. Ajani, “Completeness of Reporting of Race and Ethnicity Data in the Nationally Notifiable Diseases Surveillance System, United States, 2006–2010,” *Journal of Public Health Management and Practice* 21, no. 2 (March–April 2015), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4708284/>.

<sup>44</sup>The most recent Summary of Notifiable Diseases describes NNDSS as “neither a single surveillance system nor a method of reporting. Rather, it is a ‘system of systems,’ which is coordinated by CDC at the national level across disease-specific programs to optimize data compilation, analysis, and dissemination of notifiable disease data.” Deborah A. Adams et al., “Summary of Notifiable Infectious Diseases and Conditions—United States, 2015,” *CDC Morbidity and Mortality Weekly Report* 64, no. 53 (August 11, 2017): 1–143, <https://www.cdc.gov/mmwr/volumes/64/wr/mm6453a1.htm>.

<sup>45</sup>“National Vital Statistics System,” *Centers for Disease Control and Prevention*, accessed January 3, 2022, <https://www.cdc.gov/nchs/nvss/index.htm>.

<sup>46</sup>Elizabeth Arias et al., “Provisional Life Expectancy Estimates for January through June, 2020,” NVSS Vital Statistics Rapid Release 6; Report No. 010 (February 2021):1–8, <https://stacks.cdc.gov/view/cdc/100392>; Jeremy A.W. Gold et al., “Race, Ethnicity, and Age Trends in Persons Who Died from COVID-19—United States, May–August 2020,” *CDC Morbidity and Mortality Weekly Report* 69, no. 42 (October 23, 2020): 1517–21, <https://www.cdc.gov/mmwr/volumes/69/wr/mm6942e1.htm> (“Age, race and ethnicity, and place of death were unknown for two (<0.01%), 465 (0.4%), and 46 (0.04%) deaths, respectively.”).

<sup>47</sup>National Research Council (US) Committee on National Statistics, “Vital Statistics: Summary of a Workshop,” (New York: National Academic Press, 2009), <https://www.ncbi.nlm.nih.gov/books/NBK219884/>.

<sup>48</sup>National Research Council (US) Committee on National Statistics, “Vital Statistics.”

<sup>49</sup>Government Accountability Office, “COVID-19: Federal Efforts Could Be Strengthened by Timely and Concerted Actions,” (Washington, D.C.: September 21, 2020), <https://files.gao.gov/reports/GAO-20-701/index.html#fnref104>.

<sup>50</sup>“COVID-19 Reported Patient Impact and Hospital Capacity by Facility,” Healthdata.gov, last modified January 3, 2022, <https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u>.

<sup>51</sup>“COVID-19 Guidance for Hospital Reporting and FAQs for Hospitals, Hospital Laboratory, and Acute Care Facility Data Reporting,” Healthdata.gov, last modified January 12, 2021, <https://healthdata.gov/stories/s/COVID-19-Reporting-and-FAQS/kjst-g9cm/>.

<sup>52</sup>Healthdata.gov, “COVID-19 FAQs for Hospitals, Hospital Laboratory, and Acute Care Facility Data Reporting.”; Government Accountability Office, “COVID-19: Federal Efforts.” The HHS National Diagnostic Testing dataset also does not include any demographics in the data available for public use. HHS Protect Public Data Hub, “COVID-19 Diagnostic Laboratory Testing,” <https://protect-public.hhs.gov/pages/national-testing>.

<sup>53</sup>Government Accountability Office, “COVID-19:Federal Efforts.”

<sup>54</sup>Government Accountability Office, “COVID-19:Federal Efforts.”

<sup>55</sup>Government Accountability Office, “COVID-19:Federal Efforts.”

<sup>56</sup>Government Accountability Office, “Sustained Federal Action Is Crucial as Pandemic Enters Its Second Year,” (Washington, D.C.: March 31, 2021), <https://www.gao.gov/products/gao-21-387>.

<sup>57</sup>Government Accountability Office, “Sustained Federal Action.”

<sup>58</sup>The CDC historical dataset regarding COVID-19 vaccination is available at <https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Demographics-in-the-United-St/km4m-vcsb>.

<sup>59</sup>All datasets may be downloaded at <https://covidtracking.com/race>.

<sup>60</sup>Of the other U.S. territories, Guam reported race information for cases and deaths, but with a non-standard set of race categories reflecting the specific demographics of their population. Puerto Rico, American Samoa, Northern Mariana Islands, and U.S. Virgin Islands did not report any race or ethnicity information for COVID-19 outcomes.

<sup>61</sup>The CRDT assessed data completeness by state in this way to give “partial credit” to states that reported at least some racial information, even if they failed to report information about all races or about any ethnic group. *See* Appendix 1. As a result, this Section only discusses states’ data completeness by race (except for testing, described at the end of this section). For the remaining sections, we discuss data quality in terms of both race and ethnicity information.

<sup>62</sup>The states that did not report race data for COVID-19 hospitalizations were Alabama, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maryland, Michigan, Mississippi, Missouri, Nevada, North Carolina, North Dakota, Oklahoma, Pennsylvania, Tennessee, Texas, Vermont, West Virginia, and Wyoming.

<sup>63</sup>Racial data are considered “missing” if the state’s total number of COVID-19 events (cases, deaths, hospitalizations, or tests) is greater than the total number of events for which any racial data were provided. For example, if a state dashboard reports 1,000 cases statewide but the racial demographics section of the dashboard only includes race information for 990 cases, then racial data would be considered “missing” for 10 cases. Depending on the state, data may be missing in this manner if there is a time lag between the updating of the statewide totals and the updating of racial demographics section of a state’s dashboard.

<sup>64</sup>Racial data are considered “reported unknown” if the state’s racial reporting includes a category labeled “Unknown.” Continuing the example from the previous footnote, the state may have included 990 cases in the racial demographics section of its dashboard, but 90 of those 990 cases are classified in a category labeled “Unknown Race.” Depending on the state, the reported unknowns may include case reports where a race question was left blank or case investigations where a subject refused to answer the question about their race.

<sup>65</sup>For more details, see Appendix 1, which discusses race and not ethnicity for the reasons stated in footnote 61, *supra*.

<sup>66</sup>See Appendix 1.

<sup>67</sup>See Appendix 1.

<sup>68</sup>Proportions are calculated from total nationwide cases (confirmed and probable), deaths (confirmed and probable), and hospitalizations (cumulative hospitalized/ever hospitalized) reported in the COVID Tracking Project core dataset. For hospitalizations, the COVID Tracking Project noted that only about two-thirds of states and territories reported data for the cumulative hospitalized/ever hospitalized metric. Proportions in this column exclude states and territories not reporting this metric and are not representative of the United States as a whole.

<sup>69</sup>For more details, see Appendix 1.

<sup>70</sup>Proportions are calculated from total nationwide cases (confirmed and probable), deaths (confirmed and probable), and hospitalizations (cumulative hospitalized/ever hospitalized) reported in the COVID Tracking Project core dataset. For hospitalizations, the COVID Tracking Project noted that only about two-thirds of states and territories reported data for the cumulative hospitalized/ever hospitalized metric. Proportions in this column exclude states and territories not reporting this metric and are not representative of the United States as a whole.

<sup>71</sup>All nine states reported race and ethnicity, and there were no states that reported race only for testing.

<sup>72</sup>This count can vary up to twenty-seven in either direction and still round to 0.57% when rounded to the nearest hundredth.

<sup>73</sup>For example, the number of COVID-19 cases among the two smallest groups, American Indian or Alaska Native (AIAN) and Native Hawaiian or Other Pacific Islander (NHPI), can be calculated as follows:

$$\text{AIAN: } 0.57\% * 559,704 = 3,190.3$$

We can estimate approximately 3,190 cases occurred, with a range of possible whole number values from 3,163 to 3,218 cases. ( $3,163/559,704 = 0.5651\%$  to  $3,218/559,704 = 0.5749\%$ )

$$\text{NHPI: } 0.29\% * 559,704 = 1,623.1$$

We can estimate 1,623 cases occurred, with a range of possible values from 1,596 to 1,651.

<sup>74</sup>The statewide total of Positive Tests was available in a different part of the dashboard, broken into polymerase chain reaction (PCR) tests (335,444) and Antigen tests (69,708). We can assume, but the state website does not confirm, that the denominator for the reported percentages is the sum of these two counts (405,152).

<sup>75</sup>Data on Iowa state population by race and ethnicity was taken from U.S. Census, 2019 American Community Survey, 5-year estimates, Table B02001 for race and Table B03002 for ethnicity.

<sup>76</sup>Iowa's use of percentages rounded to whole numbers obscures the true extent of racial disparities for other racial and ethnic groups as well. For the AIAN group in Iowa with a state population of 11,976 people, the case rate ranges in possible values from 0 to 156.2 cases per 1,000 people. Comparing AIAN to white people, the case rate ratio ranges in possible values from 0 to 2.1. This means that it is possible that AIAN people are more than twice as likely as white people to have contracted COVID-19. It is also possible that no disparity exists, or that white people were in fact more likely than AIAN people to have contracted COVID-19. Meanwhile, for Black people, the state population is 116,359. The case rate among Black people ranges from 80.4 to 112.6 cases per 1,000 people. The Black/white case rate ratio is estimated at 1.3, but ranges in possible values from 1.0 to 1.5. The value of 1.3 indicates that Black people are approximately 30% more likely than white people to have contracted COVID-19. However, a ratio of 1.0 is also within the range of possible values, which would mean that no Black/white disparity exists. By contrast, the Hispanic/Latino/a/e/x to white case rate ratio is estimated at 1.6 with a range of possible values from 1.3 to 1.7. We can conclude that a disparity exists in this case, since the range of possible values are all above 1.0. The point estimate of 1.6 means Hispanic/Latino/a/e/x people are approximately 60% more likely than white people to have contracted COVID-19, with a range of possible values from 30% to 70%.

<sup>77</sup>In Figure 5, Louisiana and the District of Columbia are each represented by two different colors because they used different methods for reporting COVID-19 cases and COVID-19 deaths. Louisiana reported race and ethnicity as separate measures for deaths; it reported race only (no ethnicity information) for cases. The District of Columbia used a combined race/ethnicity measure for cases but separated race and ethnicity for measuring deaths.

Two states, Minnesota and Virginia, changed reporting practices over the course of the CRDT data collection period (reporting race and ethnicity separately at first and later switching to a combined race/ethnicity measure). These states are shown in Figure 5 with the reporting practice they used for the majority of the reporting period.

<sup>78</sup>As explained above, Florida did not report any racial data for several OMB categories including Asian, AIAN, and NHPI, and the state did not specify if these categories were included under “Other race.” Due to these limitations, the discussion of Florida’s data in this Section pertains only to case rate disparities among Black and Hispanic/Latino/a/e/x populations relative to the white population.

<sup>79</sup>Population data from U.S. Census American Community Survey 5-year estimates (obtained from [data.census.gov](https://data.census.gov)):

- Table B02001 for population by race in the separate race and ethnicity table
- Table B03002 for ethnicity in the separate race and ethnicity table
- Table B03002 for race/ethnicity in the race/ethnicity combined table

<sup>80</sup>The Hispanic/Latino/a/e/x case rate ratio is calculated as Hispanic/Latino/a/e/x cases per 1,000 population divided by white cases per 1,000 population.

<sup>81</sup>The Black case rate ratio is calculated as Black cases per 1,000 population divided by white cases per 1,000 population.

<sup>82</sup>“COVID-19 Data,” Utah Department of Health, captured July 7, 2021, <https://coronavirus.utah.gov/case-counts/> (“Race and ethnicity groups follow US Census estimates for race alone or in combination in order to provide a broad snapshot of Utah’s growing diversity, including the many multi-racial and multi-ethnic individuals who call Utah home. Groups are not mutually exclusive and will not sum to total.”).

<sup>83</sup>Table B02001 and Table B03002 of the U.S. Census American Community Survey (ACS) provide population data with multiracial individuals in a separate category labeled “Two or more races,” comparable to version 1 of our sample data. Tables B02008 and B02009 of the U.S. Census ACS provide the population for each race alone or in combination, comparable to version 2 of our sample data.

<sup>84</sup>Alabama also reported Hispanic/Latino/a/e/x ethnicity as a separate measure from race (chart not shown).

<sup>85</sup>“Indiana COVID-19 Data Report,” Indiana COVID-19 Dashboard and Map, captured July 8, 2021, <https://www.coronavirus.in.gov/2393.htm>. ISDH stands for the Indiana State Department of Health.

<sup>86</sup>See United States Census Bureau, “Data Suppression,” (October 1, 2016), <https://www.census.gov/programs-surveys/acs/technical-documentation/data-suppression.html>.

<sup>87</sup>See Roxanne Dunbar-Ortiz, *An Indigenous Peoples’ History of the United States* (Boston: Beacon Press, 2014), 8–10.

<sup>88</sup>NCAI Policy Research Center, “Disaggregating American Indian & Alaska Native Data: A Review of Literature,” (Washington DC: National Congress of American Indians, 2016), <https://www.policylink.org/sites/default/files/AIAN-report.pdf>.

<sup>89</sup>“Michigan Data,” Coronavirus, Michigan.gov, last modified January 5, 2022, [https://www.michigan.gov/coronavirus/0,9753,7-406-98163\\_98173---,00.html](https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html).

<sup>90</sup>“Covid-19 Information,” Louisiana Department of Health, accessed July 8, 2021, <https://ldh.la.gov/Coronavirus/>. This was indicated by notes on the state dashboard stating: “Cases by Race by Region (updated weekly on Wednesday)” and “Deaths by Race by Region (updated weekly on Wednesday).”

<sup>91</sup>Louisiana Department of Health, “COVID-19 Information.”

<sup>92</sup>As with the CRDT team, the RDT team had difficulty finding data for any of the U.S. territories apart from the District of Columbia.

<sup>93</sup>Cities, counties, and Continuums of Care may have other data on houselessness that are not publicly available, and may use different collection mechanisms in addition to the data the RDT was able to locate. The RDT team contacted city statistics departments for data, but found little information that way.

<sup>94</sup>“What Is a Continuum of Care?,” National Alliance to End Homelessness (NAEH), last modified January 14, 2010, <https://endhomelessness.org/resource/what-is-a-continuum-of-care/>. Not all homeless service providers are part of a Continuum of Care, but all Continuums of Care are federally designated. 24 C.F.R. §§ 578 *et seq.*

<sup>95</sup>NAEH, “What Is a Continuum of Care?.”

<sup>96</sup>NAEH, “What Is a Continuum of Care?.”

<sup>97</sup>Although all Continuums of Care (CoCs) are federally designated HUD constructs, CoCs may choose not receive funding from HUD, and those that opt out of such funding are not required to (and therefore generally do not) report data, including race and ethnicity data, to HUD.

<sup>98</sup>Kasia Klasa et al., “Continuums of Care Survey 2020: Results and Recommendations,” (New York: Public Health Governance Lab, December 2021), <https://cornell.app.box.com/v/CoCSurveyBrief2021>.

<sup>99</sup>Bruce D. Meyer et al., “Learning about Homelessness Using Linked Survey and Administrative Data,” *National Bureau of Economic Research*, no. W28861, (2021): 14, [https://www.nber.org/system/files/working\\_papers/w28861/w28861.pdf](https://www.nber.org/system/files/working_papers/w28861/w28861.pdf).

<sup>100</sup>In 2007, eleven more CoCs reported than in 2020, and four more reported in 2016 and 2013 than in 2020. See Appendix 2 for more details. To our knowledge, no single source shows how many CoCs are failing to report race and ethnicity data on homelessness and the reasons for non-reporting.

<sup>101</sup>Bruce D. Meyer et al., “Learning about Homelessness,” 3.

<sup>102</sup>Shelter individuals are classified into further categories: emergency sheltered, transitional sheltered, rapid re-housing, safe haven, and individuals in permanent supportive housing.

<sup>103</sup>For example, some shelters may record specific unique users over the course of one year, while others record any user per day.

<sup>104</sup>The most prominent report on homelessness in the United States is the Annual Homeless Assessment Report (AHAR), which is produced by HUD by using both PIT counts and HMIS data. AHAR has been presented to Congress yearly since 2007. Bruce D. Meyer et al., “Learning about Homelessness,” 3. AHAR is limited by the data quality problems described above for PIT counts and HMIS data. Moreover, AHAR presents aggregated information of the state of homelessness by race and ethnicity, but only at the national level and by categories of CoC, such as suburban and urban. The report categorizes CoCs by major city CoC, other largely urban CoC, largely suburban CoC, and largely rural CoC. See, e.g., “The 2020 Homeless Assessment Report to Congress,” *U.S. Department of Housing and Urban Development, Office of Community Planning and Development*, (Washington, D.C., 2020), <https://www.huduser.gov/portal/sites/default/files/pdf/2020-AHAR-Part-1.pdf>.

<sup>105</sup>Amy Symens Smith et al., “The Emergency and Transitional Shelter Population,” (U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2012). This effort was a continuation of previous efforts from 2000 and 1990 to count this population. Bruce D. Meyer et al., “Learning about Homelessness,” 11–12. The 2020 Census data were collected between April 9 and

May 4, 2020. Collection sites included emergency and transitional shelters, soup kitchens, mobile food vans and targeted non-sheltered outdoor locations.

**106** Bruce D. Meyer et al., “Learning about Homelessness,” 1.

**107** These data are made available to qualified researchers through the Federal Statistical Research Data Center (FSRDC) network, and can be requested here: <https://www.census.gov/programs-surveys/ces/data/restricted-use-data.html>.

**108** The Criminal Justice Administrative Records System (CJARS), founded in 2016, is working to create a “nationally integrated data repository designed to transform research and policymaking on the United States criminal justice system.” Keith Finlay and Michael Muller-Smith, “CJARS Data Documentation,” *Criminal Justice Administrative Records System (CJARS)*, (March 22, 2021) 8, <https://cjars.isr.umich.edu/data-documentation-download/>. Currently, however, among the sixteen states that have volunteered statewide conviction data to CJARS, 36.7% of cases are missing valid race. Finlay and Muller-Smith, “CJARS Data Documentation,” 9, 36.

**109** These data were available by way of a fixed-length, unpacked database, which requires expertise with programming software to access.

**110** The downloadable database has information on ethnicity, but the format available to the public requires the assistance of a data scientist to disentangle ethnicity information.

**111** While the collection effort did not focus on other arrest data, the RDT team did identify that race data (and not ethnicity data) was available for arrest of various person crimes, property crimes, societal crimes, drug abuse crimes, gambling crimes, and prostitution crimes in the CDE.

**112** Data are current up to 2019.

**113** Data on victim and offender race and ethnicity are available for download for 2018 and 2019, but only at the national level. The racial categories used for that dataset are also not compliant with OMB specifications (white, Black, Hispanic, Asian, and other). Datasets for criminal victimization for other years are available through an account with the Inter-University Consortium for Political and Social Research (ICPSR) and include racial and ethnic information at the national level (the categories of race vary over time and per question).

**114** Kenny Lo, “How to Address Concerns about Data on Deaths in Custody,” *Center for American Progress* (May 24, 2021), <https://www.americanprogress.org/issues/criminal-justice/reports/2021/05/24/499838/address-concerns-data-deaths-custody/>; Ethan Corey, “How the Federal Government Lost Track of Deaths in Custody,” *Appeal* (June 24, 2020), <https://theappeal.org/police-prison-deaths-data/>; Evaluation and Inspections Division 19-01, “Review of the Department



of Justice’s Implementation of the Death in Custody Reporting Act of 2013,” (Washington, D.C.: U.S. Department of Justice Office of the Inspector General, 2018), <https://oig.justice.gov/reports/2018/e1901.pdf>.

**115**See the Crime Data Explorer from the FBI for more details at <https://crime-data-explorer.fr.cloud.gov/pages/home>.

**116**These include Oregon, California, Nevada, Utah, Colorado, Texas, Hawaii, Minnesota, Missouri, Alabama, Illinois, Tennessee, North Carolina, Virginia, West Virginia, Maryland, New Jersey, New York, Connecticut, Vermont, and Maine.

**117**Alabama, California, Colorado, Connecticut, District of Columbia, Maryland, Missouri, New Jersey, New York, North Carolina, Oregon, Tennessee, Texas, Vermont, and Virginia request reporting on the race of the victim. Colorado, Connecticut, District of Columbia, Maryland, Missouri, New Jersey, New York, North Carolina, Oregon, Texas, and Virginia require reporting on the ethnicity of the victim.

**118**“Use of Force Data and Transparency,” National Conference of State Legislatures (NCSL), last modified November 2020, <https://www.ncsl.org/research/civil-and-criminal-justice/use-of-force-data.aspx>.

**119**Data variables include victim names, location, police department, cause of death, and other informational variables on the circumstances of the killing.

**120**This database includes informational variables concerning the name of the victim, date of the shooting, manner of death, armed/unarmed status, age, gender, race, city, state, signs of mental illness, threat level, flight, and body camera use.

**121**Prevalence counts are the proportion of individuals in a defined population that have an outcome/characteristic of interest.

**122**The CDC’s National Violent Death Reporting System (NVDRS) contains such data dating back to 2003, though at that time, data were only available for seven states. Its public interface contains data on “deaths from legal intervention,” which is a misleading euphemism for police killings. Indeed, the NVDRS technical documentation defines “deaths from legal intervention” as “a subtype of homicide where the victim is killed by or died as a result of law enforcement acting in the line of duty.” See National Violent Death Reporting System, accessed February 2, 2022, <https://wisqars.cdc.gov/nvdrs/>.

**123**See, e.g., Duren Banks et al., “National Sources of Law Enforcement Employment Data,” (Washington, D.C.: U.S. Department of Justice Bureau of Justice Statistics, October 4, 2016), <https://bjs.ojp.gov/content/pub/pdf/nsleed.pdf>.

<sup>124</sup>For a discussion of who collects racial data in the public health context and how their practices can be improved, see Geoffrey S. Holtzman et al., “The Racial Data Gap: Lack of Racial Data as a Barrier to Overcoming Structural Racism,” *American Journal of Bioethics* 22, no. 3 (March 8, 2022), <https://doi.org/10.1080/15265161.2022.2027562>.

<sup>125</sup>See, e.g., Bruce G. Link and Jo C. Phelan, “Social Conditions as Fundamental Causes of Disease,” *Journal of Health & Social Behavior* No. 80-94 (1995) <https://pubmed.ncbi.nlm.nih.gov/7560851/>; Jo C. Phelan, Bruce G. Link, and Parisa Tehranifar, “Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications,” *Journal of Health & Social Behavior* 51 (2010), <https://pubmed.ncbi.nlm.nih.gov/20943581/>.

<sup>126</sup>See, e.g., Julia Angwin et al., “Machine Bias: There’s Software Used across the Country to Predict Future Criminals. And It’s Biased Against Blacks,” *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; Latanya Sweeney, “Discrimination in Online Ad Delivery,” *Communications of the ACM* 56, (May 2013), <https://dl.acm.org/doi/10.1145/2447976.2447990>.

<sup>127</sup>Nancy López, “‘What’s Your Street Race-Gender?’ Why We Need Two Separate Questions on Hispanic Origin and Race for the 2020 Census and Implications for the African Diaspora and Urban Communities,” *Smithsonian Showcase* (Washington, D.C.: 2015), <http://www.ustream.tv/recorded/58998514>.

<sup>128</sup>For example, the Affordable Care Act requires the U.S. Department of Health and Human Services to develop data collection “standards” for race and ethnicity, using the OMB categories as a floor, but including additional racial and ethnic categories that can be “rolled up” into the OMB categories. The following options are “rolled up” into the Asian category: “Asian Indian,” “Chinese,” “Filipino,” “Japanese,” “Korean,” “Vietnamese,” and “Other Asian.” 42 U.S.C. § 300kk; Office of the Assistant Secretary for Planning and Evaluation, “HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status,” (Washington, D.C.: U.S. Department of Health and Human Services, October 30, 2011), <https://aspe.hhs.gov/reports/hhs-implementation-guidance-data-collection-standards-race-ethnicity-sex-primary-language-disability-0>.

<sup>129</sup>Arlene Bierman et al., “Addressing Racial and Ethnic Barriers to Effective Health Care: The Need for Better Data,” *Health Affairs* 21, no. 3 (May/June 2002), <https://pubmed.ncbi.nlm.nih.gov/12026007/>.

<sup>130</sup>Kimberlé Crenshaw, “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics,” *University of Chicago Legal Forum* 1989, no. 8 (1989): 139, <http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>.

- 131**Crenshaw, “Demarginalizing the Intersection,” 140. See Angela P. Harris, “Race and Essentialism in Feminist Legal Theory,” *Stanford Law Review* 42, no. 3 (February 1990), <https://www.jstor.org/stable/1228886> (challenging “gender essentialism—the notion that a unitary, ‘essential’ women’s experience can be isolated and described independently of race, class, sexual orientation, and other realities of experience”).
- 132**López, “What’s Your ‘Street Race?’”
- 133**American Sociological Association, “The Importance of Collecting Data.” See Victor Rubin et al., “Counting a Diverse Nation: Disaggregating Data on Race and Ethnicity to Advance a Culture of Health,” (Policy Link, 2018): 46, [https://northsoundach.org/wp-content/uploads/2019/11/Counting\\_a\\_Diverse\\_Nation\\_08\\_15\\_18\\_sized.pdf](https://northsoundach.org/wp-content/uploads/2019/11/Counting_a_Diverse_Nation_08_15_18_sized.pdf).
- 134**Kimberlé Crenshaw, “Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color,” *Stanford Law Review* 43, no. 6 (July 1991): 1252–53, <https://doi.org/10.2307/1229039>.
- 135**Merlin Chowkwanyun and Adolph L. Reed, “Racial Health Disparities and Covid-19 — Caution and Context,” *New England Journal of Medicine* 383, no. 3 (July 16, 2020): 202, <https://www.nejm.org/doi/full/10.1056/nejmp2012910>.
- 136**See generally Khalil Gibran Muhammad, *The Condemnation of Blackness* (Cambridge: Harvard University Press, 2011).
- 137**Chowkwanyun and Reed, “Racial Health Disparities and Covid-19,” 202.
- 138**Chowkwanyun and Reed, “Racial Health Disparities and Covid-19,” 202.
- 139**See generally LaGarrett J. King, Gregory D. Simmons, and Barry Thomas, “Different Names, Same Chains: Black Political Disenfranchisement from the Three-Fifths Compromise to the Prison Industrial Complex,” *Black History Bulletin* 82, no. 2 (Fall 2019): 27–28, [https://www.jstor.org/stable/10.5323/blachistbull.82.2.0026#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/10.5323/blachistbull.82.2.0026#metadata_info_tab_contents); Muhammad, *The Condemnation of Blackness*.
- 140**Kendi, “We Still Don’t Know”; Robert A. Hahn and Donna H. Stroup, “Race and Ethnicity in Public Health Surveillance: Criteria for the Scientific Use of Social Categories,” *Public Health Reports* 109, no. 1 (1994): 13, <https://pubmed.ncbi.nlm.nih.gov/8303018/> (“[C]ategories such as race and ethnicity have been and continue to be important determinants of access to societal resources.”).
- 141**American Sociological Association, “The Importance of Collecting Data.”

<sup>142</sup>Will Douglas Heaven, “Predictive Policing Algorithms Are Racist. They Need to Be Dismantled,” *MIT Technology Review*, July 17, 2020, <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>; Kristian Lum and William Isaac, “To Predict and Serve?,” *Significance* 13, no.5 (October 2016): 15–16, <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x>.

<sup>143</sup>See generally Peter Wagner, “Breaking the Census: Redistricting in an Era of Mass Incarceration,” *William Mitchell Law Review* 38, no. 4 (2012), <https://open.mitchellhamline.edu/wmlr/vol38/iss4/9/>.

<sup>144</sup>Talia B. Gillis and Jann L. Speiss, “Big Data and Discrimination,” *University of Chicago Law Review* 86, no. 2 (2019): 472, <https://lawreview.uchicago.edu/publication/big-data-and-discrimination>. Bipartisan conversations about regulating artificial intelligence and data are ongoing. See, for example, the provisions concerning artificial intelligence in the National Defense Authorization Act of 2021, as summarized by Stanford University. “Summary of AI Provisions from the National Defense Authorization Act 2021,” *Stanford University Human-Centered Artificial Intelligence*, last modified January 1, 2021, <https://hai.stanford.edu/policy/policy-resources/summary-ai-provisions-national-defense-authorization-act-2021>.

<sup>145</sup>Holt, “Legal Considerations.” Federal data collection is subject to laws such as the Paperwork Reduction Act of 1995, the Privacy Act of 1974, the Federal Records Act of 1950, the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the E-Government Act of 2002, the Genetic Information Nondiscrimination Act of 2008, and the Public Health Service Act (PHSA), as well as Institutional Review Board requirements.

<sup>146</sup>Institute of Medicine, *Unequal Treatment*. For a deeper discussion of state-imposed restraints, see Melnick and Perrin, *Improving Racial and Ethnic Data on Health*.

<sup>147</sup>*Printz v. United States*, 521 U.S. 898, 925 (1997); accord *New York v. United States*, 505 U.S. 144 (1992).

<sup>148</sup>Makram B. Jaber, “Unfunded Federal Mandates: An Issue of Federalism or a ‘Brilliant Sound Bite?’,” *Emory Law Journal* 45, no. 1 (1996): 287; Robert W. Adler, “Unfunded Mandates and Fiscal Federalism: A Critique,” *Vanderbilt Law Review* 50, no. 5 (1997): 1168, <https://scholarship.law.vanderbilt.edu/vlr/vol50/iss5/2/>.

<sup>149</sup>Jaber, “Unfunded Federal Mandates,” 303; Adler, “Unfunded Mandates and Fiscal Federalism,” 1154–75; Catherine Lovell and Charles Tobin, “The Mandate Issue,” *Public Administration Review* 41, no. 3 (1981): 318–31.

150 U.S. Const. Art. I, § 8, cl. 3.

151 U.S. Const. Art. I, § 8, cl. 1.

152 Jaber, “Unfunded Federal Mandates,” 303. “One of the ‘oldest questions of constitutional law’ . . . is whether the Tenth Amendment constitutes such a limitation on Congress’s commerce power. On this question, the Supreme Court has vacillated in its views.” Jaber, “Unfunded Federal Mandates,” 303. While federal mandates are often challenged on the ground that they are costly to implement, “even justices holding very different views of the Tenth Amendment have been able to agree that the financial impact of a federal mandate, by itself, is not a determinative factor in the constitutional analysis.” Patricia T. Northrop, “The Constitutional Insignificance of Funding for Federal Mandates,” *Duke Law Journal* 46, no. 7 (1997): 915.

153 Jaber, “Unfunded Federal Mandates,” 316.

154 In the context of public health surveillance, “both the state and federal governments often have concurrent jurisdiction in promoting the best use of public health resources” through what is known as “police powers.” Christopher Ogolla, “Will the Use of Racial Statistics in Public Health Surveillance Survive Equal Protection Challenges? A Prolegomenon for the Future,” *North Carolina Central Law Review* 31, no. 1, (October 2008): 6, <https://core.ac.uk/download/pdf/234101483.pdf>. See *Bond v. United States*, 572 U.S. 844, 854 (2014); Holt, “Legal Considerations.” But the powers of the federal government are more limited than those of the states. *Bond*, 572 U.S. at 854. For this reason, “CDC’s involvement in state and local public health investigations usually is intended to assist the state or local investigator rather than exercise a specific federal authority.” Holt, “Legal Considerations.”

155 Some unfunded federal mandates are also subject to the 1995 Unfunded Mandates Reform Act (UMRA), 2 U.S.C. §§ 1501 *et seq.* (1995), which enacted certain procedural requirements to guide Congressional deliberations regarding such mandates. Daniel H. Cole and Carol S. Comer, “Rhetoric, Reality, and the Law of Unfunded Federal Mandates,” *Stanford Law & Policy Review* 8, no. 2 (1997): 104, <https://www.repository.law.indiana.edu/facpub/651>.

156 *South Dakota v. Dole*, 483 U.S. 203, 206–07 (1987).

157 *Dole*, 483 U.S. at 206–07 n.2 (internal quotations omitted).

158 Jaber, “Unfunded Federal Mandates,” 345.

159 L.M. Hunt and M.S. Megyesi, “Genes, Race and Research Ethics: Who’s Minding the Store?,” *Journal of Medical Ethics* 34, no. 6 (2008): 496, <http://www.jstor.org/stable/27720117>.

<sup>160</sup>See, e.g., C. Matthew Snipp, “Racial Measurement in the American Census: Past Practices and Implications for the Future,” *Annual Review of Sociology* 29, no. 1 (2003): 574, <https://www.jstor.org/stable/30036980>; M. Barton Laws and Rachel A. Heckscher, “Racial and Ethnic Identification Practices in Public Health Data Systems in New England,” *Public Health Reports* 117, no.1 (January–February 2002): 50–61, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1497401/>.

<sup>161</sup>See Tina J. Kauh, Jen’nan Ghazal Read, and A.J. Scheitler, “The Critical Role of Racial/Ethnic Data Disaggregation for Health Equity,” *Population Research and Policy Review* 40, no. 4 (August 2021), <https://link.springer.com/article/10.1007/s11113-020-09631-6>.

<sup>162</sup>Hansi Lo Wang, “1 in 7 People Are ‘Some Other Race’ on the U.S. Census. That’s a Big Data Problem,” *National Public Radio* (September 30, 2021), <https://www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic>.

<sup>163</sup>Nicholas Jones et al., “2020 Census Illuminates Racial and Ethnic Composition of the Country,” *United States Census Bureau* (August 12, 2021), <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>.

<sup>164</sup>For example, in San Diego County, California, “the zip codes with the highest rates of COVID-19 cases, hospitalizations, and deaths overlap with the area’s Arab American enclaves.” Farah Kader and Clyde Lanford Smith, “Participatory Approaches to Addressing Missing COVID-19 Race and Ethnicity Data,” *International Journal of Environmental Research and Public Health* 18, no. 12 (June 2021): 3, <https://www.mdpi.com/1660-4601/18/12/6559/htm>. Without a separate category for Middle Eastern/North African people, inequities like this cannot be adequately tracked and addressed.

<sup>165</sup>Snipp, “Racial Measurement in the American Census.”

<sup>166</sup>Office of the Assistant Secretary for Planning and Evaluation, “HHS Implementation Guidance on Data Collection Standards.”

<sup>167</sup>Mora and Rodríguez-Muñiz, “Latinos, Race, and the American Future.”

<sup>168</sup>In 2009, an Institute of Medicine subcommittee recommended “combining

the use of granular ethnicity categories with the broad OMB categories, as well as an assessment of a patient’s language need” as part of a larger, standardized racial and ethnic demographic health data collection scheme. Cheryl Ulmer et al., “Implementation.” The Satcher Institute also recommends that “all states should require the reporting of racial/ethnic data using the OMB Race and Ethnic Standards for Federal Statistics.” Douglas et al., “Variation in Reporting of the Race and Ethnicity,” e6.

**169**For a discussion of criteria for meaningful public health surveillance and their application to racial and ethnic categories, see, e.g., Hahn and Stroup, “Race and Ethnicity in Public Health Surveillance,” 7–15.