Variation in State COVID-19 Disease Reporting Forms on Social Identity, Social Needs, and Vaccination Status

Sunny C. Lin, PhD, MS; Anna Zhen, BS; Abram Zamora-Gonzalez; José Hernández; Steven Fiala, MPH; Aileen Duldulao, PhD, MSW

ABSTRACT
To assess variation in the way state health departments collect data on social identity, social needs, and vaccination status for people testing positive for COVID-19, we reviewed all state health department Web sites for COVID-19 disease reporting forms. We located forms for 39 states and coded each one for the presence and format of fields on race, ethnicity, disability, language, sexual and gender identity, comorbidities, pregnancy status, and social needs such as housing stability, occupation, and prior COVID-19 vaccination status. We find considerable variation in both type and format of data collected. For example, 10% of state forms have fields for race and ethnicity that are more disaggregated than the 5 US Census categories, 18% have fields on mental or physical ability, 37% have nonbinary or other fields for gender identity, 28% have fields on housing stability, and 15% have fields on vaccination status. Our findings suggest that national efforts to address COVID-19 may be limited by lack of standardization of data collection methods.

KEYWORDS: data collection, data equity, health disparities, public health informatics, social determinants of health

The collection of reliable, consistent data is key to an equitable national disease response: data guide policy priorities, promote community trust, combat misinformation, and inform resource allocation. Yet, the impact of the COVID-19 pandemic on marginalized communities in the United States is not well understood due to lack of consistent data collection practices. Lack of data can have detrimental effects on marginalized communities. For example, emerging evidence suggests that immigrants accounted for more than half of COVID-19–related deaths in San Diego in the first year of the pandemic, likely due to repeated exposure, higher stress, and limited access to health care brought on by structural inequalities. However, without data on the rates of mortality, morbidity, and breakthrough infections by migrant status, policy makers are unable to allocate resources to support health care access for these communities.

Currently, there is no agreed-upon standard for the collection of data related to social identity and needs in public health. In response to emerging research documenting the impact of COVID-19 on marginalized populations, many state health departments have mandated the reporting of social identity and needs information on COVID-19 disease reporting forms. However, similar to other infectious disease cases such as HIV/AIDS, social identity/needs data collection types and formats for COVID-19 are not standardized across states. Moreover, the Centers for Disease Control and Prevention (CDC) does not mandate the collection of information on social identity or needs in COVID-19 disease reports. If data collection methods are not comparable across states, it will be challenging to estimate and mitigate the impact of disease on marginalized communities at a national level. Currently, little is known about how comparable data collected on states’ COVID-19 disease report forms are to each other. We examined state health departments’ COVID-19 disease reporting forms to see what types of social identity, social needs, and vaccination data are collected and how.
Methods

From May to August 2021, we searched state public health department Web sites for disease reporting forms that providers used to report confirmed cases of COVID-19. We coded each form for the presence of specific data points related to social identity and social need. We chose to examine data points recommended in a 2015 Institute of Medicine (IOM) report on capturing data on social and behavioral domains in electronic health records, as well as those that were considered important after a review of the literature on populations most impacted by COVID-19. Data were then validated by 2 trained members of the study team and the study principal investigator.

For social identity, we examined whether forms contained fields on race and ethnicity, language, disability, sexual orientation, and gender identity. For race and ethnicity, we created 5 binary variables: (1) race and ethnicity in any format; (2) closed-ended options for the 5 Census categories for race and ethnicity (ie, White, Black, Asian/Pacific Islander, American Indian/Alaskan Native, and Hispanic); (3) more disaggregated race and ethnicity options; (4) open-ended fields; and (5) tribal affiliation. For language, we identified form fields for “language spoken at home” or “translation needed.” For disability, we identified fields for physical or mental disabilities, such as “disability status” or “physical or mental impairment.” For sexual orientation and gender identity, we created 4 binary variables for (1) sex; (2) gender nonbinary options such as “other” or “non-binary”; (3) transgender identity options; and (4) sexual orientation.

For social needs, we identified fields for occupation, housing, comorbidities, pregnancy status, COVID-19 vaccination, health insurance, and food security. For occupation, we identified open-ended questions for occupation or closed-ended questions about health care or childcare-related employment. For housing setting, we created 4 variables for residence in (1) congregate housing settings; (2) unstable housing such as homeless shelter, outside, or vehicle; (3) long-term care or other chronic care settings; and (4) correctional facilities. Finally, we created binary variables for pregnancy status, COVID-19 vaccination, health insurance, and food security.

This study was approved by the Portland State University Institutional Review Board.

Results

We located disease reporting forms for 39 of 50 US state health departments: 69% (n = 9) of Western states; 100% (n = 12) of Midwest states; 76% (n = 13) of Southern states; and 56% (n = 5) of Northeastern states. The remaining 11 states’ forms were either restricted to registered reporters (n = 10) or nonexistent (n = 1, report through e-mail text). Thirty-seven of the forms were accessible in .pdf files or Web-based forms, and 2 were restricted to registered users through a Web portal. For those 2, we accessed screenshots of the forms through training guides. The Table shows data types and formats present on state disease reporting forms.

Social identity

All 39 (100%) states’ forms contained at least one question about race and ethnicity. Type of race categories varied: 87% (n = 34) used at least the 5 minimum Census categories (ie, Black/African American, White, American Indian/Alaskan, Asian, and Native Hawaiian/Pacific Islander); 13% of states (n = 5) used race and ethnicity categories that were less disaggregated than the Census categories (eg, combining Asian and Pacific Islander categories); and 10% (n = 4) offered at least one option for race and ethnicity that was more disaggregated than the 5 minimum Census categories. Of note, 87% (n = 34) of states disaggregated the Asian and Pacific Islander categories and 5% (n = 2) of states included a “Middle Eastern” option. Form of race included check boxes (87%, n = 34), open-ended only (6%, n = 2; eg, “Race (Specify)”), and drop-down or radio buttons that limit options to one race only (8%, n = 3). Forty-one percent (n = 16) had an open-ended option for race (eg, “Other (Specify)”). 15 of those were tied to specific race categories (eg, “Asian, other (Specify)”). Ten percent (n = 4) asked about tribal affiliation. Twenty-three percent (n = 9) of states asked about language. Eighteen percent (n = 7) asked about disability.

All but one state (n = 38) collected information on sex or gender. We were unable to identify the response options for sex/gender for one Southern state, so we excluded this state in the statistics reported for nonbinary and transgender options. Thirty-seven percent (n = 14) of states included a closed-ended nonbinary option, and 21% (n = 8) of states asked whether individuals identified as transgender. Five percent (n = 2) of states collected data on sexual orientation.

Social needs and prior vaccination status

Sixty-two percent (n = 24) of states’ forms asked whether the individual lived or worked in congregate housing. Twenty-eight percent (n = 11) asked about housing stability. About half (n = 19) asked about residence in a chronic care setting, and 41% (n = 16) asked about residence in a correctional facility. Seventy-seven percent (n = 30) of forms asked about...
### Table: Percentage of States That Collect Data on Social Identity and Social Needs on COVID-19 Disease Reports, by Data Format and Region

| Social identity                              | United States (N = 39) | West (n = 9) | Midwest (n = 12) | South (n = 13) | Northeast (n = 5) |
|----------------------------------------------|------------------------|--------------|------------------|----------------|------------------|
| Race and ethnicity                           | 39 (100%)              | 9 (100%)     | 12 (100%)        | 13 (100%)      | 5 (100%)         |
| 5 Census categories                          | 34 (87%)               | 9 (100%)     | 8 (67%)          | 12 (92%)       | 5 (100%)         |
| Any open ended                               | 16 (41%)               | 5 (56%)      | 5 (42%)          | 3 (23%)        | 3 (60%)          |
| Disaggregated categories                     | 4 (10%)                | 3 (33%)      | 1 (8%)           | 0 (0%)         | 0 (0%)           |
| Tribal affiliation                           | 4 (10%)                | 2 (22%)      | 1 (8%)           | 1 (8%)         | 0 (0%)           |
| Language                                     | 9 (23%)                | 5 (56%)      | 3 (25%)          | 1 (8%)         | 0 (0%)           |
| Disability                                   | 7 (18%)                | 2 (22%)      | 2 (17%)          | 2 (15%)        | 1 (20%)          |
| Sex or gender                                | 38 (97%)               | 9 (100%)     | 11 (92%)         | 13 (100%)      | 5 (100%)         |
| Nonbinary or other option                    | 14 (37%)†              | 7 (88%)      | 3 (25%)          | 2 (17%)‡       | 2 (40%)          |
| Transgender option                           | 8 (21%)‡               | 4 (50%)      | 0 (0%)           | 3 (25%)‡       | 1 (20%)          |
| Sexual orientation                           | 2 (5%)                 | 2 (22%)      | 0 (0%)           | 0 (0%)         | 0 (0%)           |

| Social need                                 |                        |              |                  |                |                  |
| Congregate housing setting                  | 24 (62%)               | 5 (56%)      | 8 (67%)          | 8 (62%)        | 3 (60%)          |
| Housing stability                           | 11 (28%)               | 3 (33%)      | 2 (17%)          | 4 (31%)        | 2 (40%)          |
| Long-term care                              | 19 (49%)               | 3 (33%)      | 4 (33%)          | 9 (69%)        | 3 (60%)          |
| Correctional facility                       | 16 (41%)               | 3 (33%)      | 4 (33%)          | 7 (54%)        | 2 (40%)          |
| Occupation                                  | 30 (77%)               | 6 (67%)      | 10 (83%)         | 10 (77%)       | 4 (80%)          |
| Comorbidities                               | 16 (41%)               | 5 (56%)      | 5 (42%)          | 3 (23%)        | 3 (60%)          |
| Pregnancy status                            | 28 (72%)               | 7 (78%)      | 10 (83%)         | 8 (62%)        | 3 (60%)          |

| COVID-19 vaccination status                 |                        |              |                  |                |                  |
| History of prior vaccination                | 6 (15%)                | 2 (22%)      | 0 (0%)           | 2 (15%)        | 2 (40%)          |

*The authors’ analysis of publicly available state COVID-19 disease reports (N = 39): 69% (n = 9) of Western states; 100% (n = 12) of Midwest states; 76% (n = 13) of Southern states; and 56% (n = 5) of Northeastern states.

†Excludes one southern state whose closed-ended options for “sex assigned at birth or sex” were not accessible; states were assigned to regions using Census Bureau designations (https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf).

### Limitations

Findings from this study should be interpreted with limitations in mind. We did not assess the quality of these data (eg, how often were specific data points missing?), nor did we examine other efforts to collect and/or standardize data collection (eg, did all local health departments in a state collect data the same way?). In addition, the generalizability of our study is limited. First, we were missing forms from 4 states from the West, 3 states from the South, and 4 states from the Northeast. Second, in some states, especially those with decentralized public health departments, forms we examined may be different from forms used by local health departments.

### Discussion

This work has important public health implications. First, we found that many state health department disease reporting forms are missing important data fields on social identity and needs such as disability and vaccination status. This is concerning, given the increasing body of evidence suggesting that social identity and needs are important predictors of COVID-19–related severe morbidity and mortality and the paucity of data on racial disparities in vaccine effectiveness. Second, we found that the types of data collected vary across states. For example, while data on race and ethnicity were collected on all states’ forms, race and ethnicity options that were more disaggregated than the 5 minimum Census categories were collected on only 10% of forms. This is important because lack of disaggregation standardization prevents state-reported COVID-19 data from being compared across state lines, making it impossible to
infer the impact of the pandemic on marginalized communities nationally. For example, a September 2020 report found that Filipino-Americans make up 4% of nurses in the United States, yet account for 32% of COVID-19–related deaths. This disparity has largely gone unnoticed and unaddressed because of lack of data disaggregation in public reporting systems—Filipino individuals are often grouped with all other Asian ethnicities when reported to the public, obscuring the existence and severity of this disparity.

Our findings suggest that coordination and guidance is needed to help local and state health departments standardize data collection practices on social identity and social needs. Although no consensus currently exists on what data points are considered essential for public health efforts, a 2014 report from the IOM suggested collecting the following sociodemographic data in electronic health records: race and ethnicity, country of origin, education, employment, financial resources strain, health literacy, and sexual orientation. And in a 2021 report, The Urban Indian Health Institute recommended that race and ethnicity data be collected according to the 1997 OMB standards for race and ethnicity, to avoid catch-all categories such as ‘other’ and ‘multiracial,’ and to allow respondents to select multiple races if applicable.

In addition, several states have made efforts to ensure that standardization of data collection efforts is responsive to needs of the local community. For example, in 2021, Oregon mandated the collection of REALD (race, ethnicity, language, or disability) and SOGI (sexual orientation or gender identity) data, using a tool developed in close partnership with community organizations. In Massachusetts, the Massachusetts Racial Equity Data Road Map was created to help guide organizations in discussions with communities about data collection. Both efforts stress the importance of collecting race and ethnicity data in order to document the effects of structural and institutional racism, rather than using it as a predictive risk factor (eg, using race as a factor in medical decision making), and tailoring efforts to identify community groups that are prevalent in the geographic area.

The Coronavirus Aid, Relief, and Economic Security (CARES) Act provided the CDC with US $500 million to support public health data infrastructure. While this investment is much needed to support our nation’s public health efforts, moving forward, it will be important for policy makers to create data collection standards that allow for comparison across communities, as well as customization to meet the needs of the local geographic area. For example, it might make sense for Minnesota to offer “Hmong” as a more granular “Asian” category for race and ethnicity, as 25% of the US Hmong population resides in the state, while allowing Maine, which has fewer Hmong Americans, to forego that option. In addition, future research is needed to understand how members of the community view and respond to social identity and need options on public health reporting forms. For example, community members in some localities may not feel safe publicly identifying as a transgender person and may be more responsive to data collection efforts that forego this question.

Our study shows that public health departments need more guidance and coordination to standardize data collection practices. Specifically, we found that data types collected for most states were minimal, with only 10% of states collecting disaggregated data on race and ethnicity, 23% collecting data on language, 18% disability, 5% sexual orientation, and 15% prior vaccination. Drawing on the work of the IOM, Urban Indian Health Institute, Oregon Health Authority, and Massachusetts Department of Public Health, we suggest that efforts to standardize data collection focus first on developing best practices for the process of engaging data collection with an equity-minded approach. Otherwise, the outsized impact of COVID-19 and other diseases on marginalized populations will continue to be rendered invisible and unaddressed because of lack of data.

**Implications for Policy & Practice**

- Most state health departments collect minimal data on social identity and social needs, limiting national efforts to understand the impact of the pandemic on marginalized communities.
- Only 15% of states collected data on prior vaccination status on disease reports. As breakthrough cases continue to rise, it is imperative to collect data on prior vaccination status in a systematic way to understand vaccine effectiveness among different populations.
- Moving forward, policy efforts to address health equity will continue to be limited by poor data collection practices, driven, in part, by lack of equity-minded approaches to the process of data collection for public health.

**References**

1. Blackstock O, Skinner A, Raifman J. Opinion: racial disparities may be emerging in breakthrough infections. We must track them better. The Washington Post. https://www.washingtonpost.com/opinions/2021/10/18/racial-disparities-may-be-emerging-breakthrough-infections-we-must-track-them-better. Published 2022. Accessed October 19, 2021.
2. Cahill S, Grasso C, Keuroghlian A, Scioritone C, Mayer K. Sexual and gender minority health in the COVID-19 pandemic: why data
collection and combating discrimination matter now more than ever. Am J Public Health. 2020;110(9):1360-1361.
3. Miller S, Wherry L, Mazumder B. Estimated mortality increases during the COVID-19 pandemic by socioeconomic status, race, and ethnicity. Health Aff (Millwood). 2021;40(8):1252-1260.
4. Krieger N. Enough: COVID-19, structural racism, police brutality, plutocracy, climate change—and time for health justice, democratic governance, and an equitable, sustainable future. Am J Public Health. 2020;110(11):1620-1623.
5. Huntsberry W, Marx J, Ross B. The first year of COVID: half of San Diego pandemic deaths were immigrants. Voice of San Diego. https://voiceofsandiego.org/2021/12/01/the-first-year-of-covid-half-of-san-diego-pandemic-deaths-were-immigrants. Published 2021. Accessed December 2, 2021.
6. Centers for Disease Control and Prevention. COVID-19 message mapping guide. Secondary COVID-19 message mapping guide. https://ndc.services.cdc.gov/mmgpage/covid-19-message-mapping-guide/. Published 2021. Accessed February 24, 2022.
7. Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records, Board on Population Health and Public Health Practice, Institute of Medicine. Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2. Washington, DC: National Academies Press; 2014.
8. Srikrishnan M. The first year of COVID: Filipinos were among hardest hit, but hidden by data. Voice of San Diego. https://voiceofsandiego.org/2021/12/06/the-first-year-of-covid-filipinos-were-among-hardest-hit-but-hidden-by-data. Published 2021. Accessed February 2, 2022.
9. Gold R, Cottrell E, Bunce A, et al. Developing electronic health record (EHR) strategies related to health center patients’ social determinants of health. J Am Board Fam Med. 2017;30(4):428-447.
10. Urban Indian Health Institute. Data genocide of American Indians and Alaska Natives in COVID-19 data. https://www.uihi.org/projects/data-genocide-of-american-indians-and-alaska-natives-in-covid-19-data. Published 2021. Accessed February 17, 2022.
11. McGee M. Summary: Assessment of Race, Ethnicity, Language and Disability (REALD) Data Quality in the Oregon Health Plan ONE System. Portland, OR: Oregon Health Authority Equity and Inclusion Division; 2018.
12. Manning S, Blinn A, Seik S, et al. The Massachusetts Racial Equity Data Road Map: data as a tool toward ending structural racism. J Public Health Manag Pract. 2022;28(suppl 1):S58-S65.
13. Centers for Disease Control and Prevention. Public health data modernization initiative. Secondary public health data modernization initiative. https://www.cdc.gov/budget/documents/covid-19/COVID-19-Data-Modernization-Initiative-Fact-Sheet.pdf. Published 2020. Accessed February 24, 2022.