




OCTOBER 2018

# RECOMMENDATIONS FOR GRANULAR RACE & ETHNICITY DATA COLLECTION

PRESENTED TO THE SIM PROGRAM STAFF, OFFICE OF HEALTH STRATEGY

HEALTH EQUITY SOLUTIONS  
175 Main Street, Hartford, CT 06106



## About the Project

Health Equity Solutions (HES), a 501(c)(3) organization in the state of Connecticut, was contracted by the Office of Health Strategy (OHS) to facilitate a discussion with three Wave 1 Community and Clinical Integration Program (CCIP) Participating Entities (PEs) on the importance of collecting granular race and ethnicity data to inform population health strategies that reduce health disparities. The following report is a summary of the information collected and reported before, during, and after the facilitated meeting.

## Background

Race and ethnicity are often confused but are two significant and separate concepts. Race is defined a social construct that is linked to perceived biological differences, often demarcated along the lines of phenotypical or somatic characteristics, i.e. skin color, hair type, eye shape, etc. On the other hand, ethnicity refers to shared beliefs, culture, ancestry and language, that uniquely and closely relevant to an individual, group or population. While vast scientific research has concluded that race has no biological value, it remains a critical point of analysis given the impact it has in the lived experiences of individuals in society.

The effort to standardize and collect race, ethnic, and language (REAL or REL)<sup>1</sup> data is not a new one. Rules framing standards on how racial and ethnic data (R/E) have primarily been promulgated at the federal level and have set the basic standard for data collection. The most popular R/E data collection standard was developed by the Office of Management and Budget (OMB) in 1977 and updated in 1997, setting five racial categories (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian other Pacific Islander, and White) and two broad categories of ethnicity (Hispanic/Latino or Not Hispanic/Latino).<sup>2</sup> While this directive set a federal policy standard minimum, collective implementation and use of the data collection standard has not been uniform.

In 2002, the Institute of Medicine (IOM)<sup>3</sup> issued its groundbreaking report entitled *Unequal Treatment*.<sup>4</sup> In the report, the IOM emphasized the importance of collecting R/E data as a strategy to address health disparities. In 2009, the IOM expanded on its previous recommendations and called for standardization in R/E data collection and called for the collection of more granular ethnicity data beyond the OMB Directive 15 (OMB-5).<sup>5</sup> Subsequently, when the Affordable Care Act (ACA) was passed in 2010, it contained section 4302, which centered on data collection, analysis and reporting standards in service to understanding health disparities.<sup>6</sup> With OMB and IOM efforts laying the foundation, Section 4302 set OMB's directive as the floor of R/E data collection and set no limitation on the

---

<sup>1</sup> While race, ethnicity, and primary language are each significant to advancing health equity and eliminating health disparities, this paper focuses squarely on race and ethnicity only.

<sup>2</sup> [OMB Directive 15, 1997](#)

<sup>3</sup> IOM is a division of the National Academies of Sciences, Engineering, and Medicine. In 2016, the IOM changed its name to the Division of Health and Medicine (HMD).

<sup>4</sup> <http://nationalacademies.org/hmd/reports/2002/unequal-treatment-confronting-racial-and-ethnic-disparities-in-health-care.aspx>

<sup>5</sup> <http://www.nationalacademies.org/hmd/Reports/2009/RaceEthnicityData.aspx>

<sup>6</sup> Section 4302 mandates all federal agencies gather information on race, ethnicity, sex, primary language, and disability status.

granularity of R/E data that could be collected, analyzed, and reported in service of reducing health disparities.

In 2014, the State of Connecticut was awarded a State Innovation Model (SIM) grant from the Center of Medicare and Medicaid Innovation (CMMI) that expanded the Triple Aim to include a focus on promoting health equity in the state's work to develop and implement healthcare payment and service delivery reform. In furthering the work, the need for R/E data to measure progress on this goal became readily apparent. Stakeholders quickly understood the limitations of the current R/E data collection efforts and availability of such data and the need to improve upon and expand data collection if the state was truly going to monitor and improve health disparities<sup>7</sup>.

Over the course of 2016-2017, HES set out to develop a proposed solution for collecting more granular level R/E data. Our goal was to offer a solution that was pragmatic, met the goals of multiple stakeholders, was dynamic rather than static in order to change over time as the population changes, and ultimately improve the health and health care of Connecticut residents.

## Methods

The research methods employed for this project involved reviewing and examining publicly available information. The research began with examining existing mandates for federal reporting and guidance for racial and ethnic categorization. The research led to the following:

- The base standard racial and ethnic data collection standard for most, if not all health-related federal data follows the 1997 OMB Directive 15;
- The most comprehensive coding of racial and ethnic categories is the Centers for Disease Control and Prevention (CDC) Health Level 7 (HL7) Race and Ethnicity Code Set, which contains over 900 race codes grouped under each of the OMB-5 racial categories and 43 detailed ethnicity codes.<sup>8</sup>

Next, we researched the approach other states and locales used to narrow the field of categories used to code R/E on vital records. This research led to the conclusion that the states and cities examined (California, Massachusetts, Minnesota, North Carolina, Oregon, and the City of New York) all used a variety of methods to determine the categories and to execute the mandate. Ultimately, all of the research, including the federal standards led to the following consistent themes:

- R/E data collection must be self-reported
- Individuals must have the option to select more than one racial or ethnic group
- Detailed racial and ethnic groups should follow the hierarchical mapping to align with the OMB-5 standard

The possible solutions to data collection issues were discussed with various communities impacted by R/E data. Specifically, representatives from Southeast Asian groups, West Indian/Caribbean groups, Latinx groups, and a representative from the Muslim community.

---

<sup>7</sup> <http://www.ctvoices.org/publications/data-promote-health-equity-children-and-families-CT>

<sup>8</sup> <https://phinivads.cdc.gov/vads/ViewValueSet.action?id=67D34BBC-617F-DD11-B38D-00188B398520>

Structured, ad hoc conversations with providers were conducted on how this could be used to improve innovation in research and to identify potential barriers in execution. There were conversations with data specialists as well. We reviewed and discussed recommendations from a for-profit health equity firm in Connecticut to test the efficacy of using the granular data. This process led to recommending a granular race and ethnicity data collection standard that was grounded in the current population of the state.

Finally, the overall list was presented to the Wave 1 CCIP PEs for feedback on October 4, 2018.

**Findings & Recommendations**

After careful review and synthesis of the research and other data from the Census Bureau’s American Community Survey (ACS), the CDC, and publicly available data on federal data websites, the following standards and mapping list was created based on the “Top 40+” ethnic groups in the state.<sup>9</sup>

*Standards*

- Self-reported over observed must be the method by which individual R/E data is collected
- Time of collection varies and should remain flexible based on workflow; follow best practices from the federal government or the Connecticut State Medical Society
- Individuals should be able to select more than one race or ethnicity
- Each category should have an “Other” option and allow for manual entry of an answer
- Where applicable & available use predictive data completion for patients (automated/tablet/online/computer-based)

*Granular Collection Mapping*

Below are the tables detailing the hierarchical mapping for the proposed race/ethnic data collection categories. Each table presents the categories in alphabetical order (not population based).

**Table I. Hierarchical Mapping for American Indian/Alaska Native (n=6)\*<sup>10</sup>**

American Indian/Alaska Native	Alaska Native	Mashantucket Pequot
	Cherokee	Mohegan
	Iroquois	Other American Indian/Alaska Native:

<sup>9</sup> The top 40 idea originated as a way to determine a practical cut-off for racial and ethnic categories that would be collected in the state. After reviewing the data, 43 ethnic groups.

<sup>10</sup> The (n=) always includes the category “other” for that table.

**Table II. Hierarchical Mapping for Asian (n= 19)**

Asian	Asian Indian	Laotian
	Bangladeshi	Malaysian
	Burmese	Nepalese
	Cambodian	Pakistani
	Chinese	Sri Lankan
	Filipino	Taiwanese
	Hmong	Thai
	Indonesian	Vietnamese
	Japanese	Other Asian:
	Korean	

**Table III. Hierarchical Mapping for Black or African American (n=8)**

Black or African American	Black or African American	Jamaican
	African	West Indian
	African American	Other:
	Dominican <sup>11</sup>	
	Haitian	

**Table IV. Hierarchical Mapping for Hispanic, Latino/a, or Spanish (n=20)**

Hispanic, Latino/a, or Spanish	Argentinian	Nicaraguan
	Chilean	Panamanian
	Columbian	Peruvian
	Costa Rican	Puerto Rican
	Cuban	Salvadorian
	Dominican	Spaniard
	Ecuadorian	Spanish
	Guatemalan	Uruguayan
	Honduran	Venezuelan
	Mexican, Mexican American, Chicano/a	Other Spanish:

**Table V. Hierarchical Mapping for Native Hawaiian or Other Pacific Islander (n=4)**

Native Hawaiian or Other Pacific Islander	Guamanian or Chamorro
	Native Hawaiian
	Samoan
	Other Pacific Islander:

<sup>11</sup> Dominican are listed in the CDC racial and ethnic codes as both a race and an ethnicity.

**Table VI. Hierarchical Mapping for Native Hawaiian or Other Pacific Islander (n=4)**

White	Arab
	European
	Middle Eastern or Northern African
	Portuguese

**Table VII. Hierarchical Mapping for Other Race (n=1)**

Some other Race:
------------------

### **Other Considerations**

In reviewing the practices of other states, we note that Minnesota uses country of origin as a data collection standing in collecting R/E data. We considered this option and concluded that country of origin alone could be a misleading question and would not capture vital disparities that could exist in American born individuals whose parents and/conception originate from another country. The research led us to conclude that if country of origin was asked, R/E data would still be necessary.

### **Limitations**

These recommendations are primarily based on ACS data, which uses estimates based on survey data. While this is a limitation, the ACS is a highly used data source to track and understand trends.

The recommendations were not publicly distributed or socialized with the widest possible audience. The project relied on a key informant method based on convenience sampling.

## **Appendix A: Frequently Asked Questions**

Most of the questions asked by the CCIP PEs can be answered by the Frequently Asked Questions document found in the [Race and Ethnicity Data Improvement Toolkit](#) produced by the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP).

### **What is the capacity of the various electronic health record vendors to collect more granular data?**

The capability of the various EHRs to collect more granular data is still under review. To date, we know that eClinicalWorks and Epic both have the ability to collect detailed race and ethnic data.

### **What existing guides or tools are available to assist with preparing to ask the R/E/L questions?**

See the FAQ document attached to this report. Also see the Connecticut State Medical Society's [Health Equity Data Collection Guide](#).

### **Allowing individual responses to the category “other” seems challenging. If this is in paper format, all we can do is scan the paper in and the other becomes free form text. What can we do about this?**

The work currently underway by the Health Information Technology Officer will allow for the extraction, reading and recording of free form text. In the near term, more detailed review of the EHR's ability to extract free form text in a systematic manner is underway.

### **How much additional time is required to ask more granular data?**

See the FAQ document attached to this report.

## References

Baker, David, et.al. March 2006. A System for Rapidly and Accurately Collecting Patient's Race and Ethnicity. *American Journal of Public Health*.

Centers for Disease Control and Prevention. PHIN VADS. 2018. Value Set, General Race.

Centers for Disease Control and Prevention. PHIN VADS. 2018. Value Set, Ethnicity.

Data Innovations. Healthcare Cost and Utilization Project (HCUP). April 2017. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/datainnovations.jsp](http://www.hcup-us.ahrq.gov/datainnovations.jsp)

Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/10260>.

Institute of Medicine. 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/12696>.

Office of Management and Budget. October 1997. The Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity were last revised in 1997 ([62 FR 58782](#))

Siegel, Karen. 2018. Data to Promote Health Equity for Children and Families in Connecticut. Connecticut Voices for Children. New Haven, CT. <http://www.ctvoices.org/publications/data-promote-health-equity-children-and-families-CT>