

## Considerations for Capturing of Race, Ethnicity, and Gender Data

**Developed by:** Northwestern University's Department of Preventive Medicine's Working Group on Response to Structural Racism\*

### Background

The Department of Preventive Medicine's (DPM) Working Group on Response to Structural Racism in Research surveyed multiple data dictionaries within DPM and beyond (e.g., Feinberg School of Medicine [FSM], United States Census, and the National Institutes of Health [NIH] reporting requirements). We assessed common practice and **aggregated ideas from these sources to develop these considerations for best practices** when capturing relevant demographic variables in Human Subjects' Research. These considerations **serve as guidelines and recommendations**. They are not binding in any way. We recognize research studies each have unique needs, and there is no single set of best practices that would apply to all studies.

### On Capturing Race and Ethnicity

1. **Consider the audience and target study population.** Depending upon country, culture, or age group of interest, these data may not apply or the study participants may not understand the social construct or categories of race or ethnicity.
2. **Explain the purpose** of this data collection to study participants (either within the consent or in the instructions of the relevant survey).
3. **Consider the goal** of collecting these data.
  - a. Will these data be used for **summarization or reporting** only? If so, consider the minimum requirements of the funder or regulatory body as a starting point for discrete categories.
  - b. Will they be **used for analyses**? If so, consider how race/ethnicity variables will be parameterized in analyses (Refer to Tips on Considerations for Analysis).
4. If the study is short-term (e.g., less than one year in length), these data will likely only require collection at "baseline." If the study is longer-term, investigators should **recognize that racial/ethnic identity from a participant's perspective may change over time**. Therefore, the investigators could consider allowing participants to provide updates on these variables as their time in the study progresses.
5. We **always** suggest obtaining these data points **via participant self-report**, according to the categories in which they identify personally.
  - a. In all cases, **participants must be allowed the option to select more than one racial/ethnic category**. This can be accomplished through checkbox options and the instructions for participants to "Select all that apply." Another alternative would be to use a matrix of (yes/no) fields for those that prefer not to use checkboxes.
  - b. Participants **must be allowed** to either **skip** questions related to race and ethnicity **OR select a "Prefer not to answer" option**.
  - c. Investigators should **consider an "Other" option, with the ability to "Specify."**
  - d. Participants must also be allowed to select "Unsure" or "Unknown."
6. Although **certain countries and reporting entities recognize a difference between race and ethnicity**, study participants may not. **Consider whether it is necessary to break these two constructs apart into separate segments of data capture of a demographics form**.
7. Regarding the number of racial and ethnic categories, **more options are better** to a degree; however, **participant burden and overwhelm should be considered**. The following are examples from the United States Census and NIH.
  - a. **Census** (allows for more than one selection)
    - i. **Race:** White, Black or African American, American Indian or Alaskan Native, Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, Other Asian, Native Hawaiian, Samoan, Chamorro, Other Pacific Islander, Some other race.
    - ii. **Ethnicity:** Hispanic, Latino, or Spanish origin as a yes / no option.
  - b. NIH Reporting Categories (as of January 2022)
    - i. **Race:** American Indian / Alaska Native, Asian, Native Hawaiian / Pacific Islander, Black or African American, White, More than one Race.

1. We recommend allowing participants to select as many / as few options as they choose of the above. If participants select more than one option, they may be grouped into the “More than one Race” category for reporting purposes.
- ii. **Ethnicity:** Hispanic or Latino vs. Not Hispanic or Latino.
8. **Avoid forcing participants to identify with broad categories** to the extent possible. It will be easier to collapse categories for reporting and analytic reasons after data collection, rather than risk participant difficulty in response.
9. Recommendations of phrasing toward study participants:
  - a. “Which of these best describe your race / ethnicity? (Select all that apply)”
  - b. “Do you identify with any of the following racial/ethnic categories?”
  - c. “Indicate the racial / ethnic categories with which you identify below. (Select all that apply)”
  - d. In each case, followed by a series of checkboxes or a matrix of yes / no fields as outlined above.

### **Sex and Gender Identity**

1. Determine whether the goals of the study **require collection of sex as a biological variable, gender identity, or both.**
2. Given **NIH often requires considerations for sex as a biological variable**, most demographic data collection tools should **include a field for “Sex assigned at birth”** with the options: Male or Female; in some cases it may be appropriate to include an “unknown” or “other” category for sex as assigned at birth.
3. In cases where **gender identity** is of interest for meet study goals, many of the ideas for race and ethnicity apply:
  - a. Response should always be via **participant self-report.**
  - b. Participants must be allowed **flexibility** in response (i.e., we cannot force them into man or woman categories).
  - c. Participants must be **allowed not to answer** if they so choose.
  - d. Participants must be **provided an “Unsure” option.**
  - e. Participants should be provided an opportunity to update their gender identity if it seems reasonable within the context of the study (e.g., lengthy follow-up time, etc.).
4. **Recommendations** on capturing data on gender identity:
  - a. Ask the participants “With which gender do you identify the most?”
  - b. Potential categories: Man, Woman, Transgender Man, Transgender Woman, Gender Queer or Gender Nonconforming, Choose not to disclose, Unsure.

### **\*Working Group Members:**

Kiarri Kershaw (Chair); Mercedes Carnethon; Jody Ciolino; Frank Granata; Elizabeth Gray; Mark Huffman; Molly Jones; Monica Rodriguez; Leah Neubauer; Denise Scholtens