

Race/Ethnicity Data Standard

From the NIH Collaboratory Phenotypes, Data Standards, and Data Quality Core

Available at: <https://www.nihcollaboratory.org/Pages/Knowledge-Repository.aspx>

See [disclaimer](#)

Using self-reporting or self-identification to collect an individual's data on ethnicity and race, investigators should use two separate questions with ethnicity information collected first followed by the option to select more than one racial designation.

Ethnicity (select one)

- Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can also be used in addition to "Hispanic or Latino."
- Not Hispanic or Latino

Race (select all that apply)

- American Indian or Alaska Native: A person having origins in any of the original peoples of North, Central, or South America, and who maintains tribal affiliations or community attachment.
 - Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. (Note: Individuals from the Philippine Islands have been recorded as Pacific Islanders in previous data collection strategies.)
 - Black or African American: A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."
 - Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
 - White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
-

Justification

The NIH has adopted the 1997 Office of Management and Budget (OMB) [revised minimum standards](#) for maintaining, collecting, and presenting data on race and ethnicity for all grant applications, contract and intramural proposals, and for all active research grants, cooperative agreements, contracts, and intramural projects.

Further, the NIH requires investigators to report ethnicity/race and sex/gender in all clinical research. Investigators are instructed to provide plans for the total number of subjects proposed for the study and to provide the distribution by ethnic/racial categories and sex/gender, with the categories presented above. Requirements and enrollment tables are available from the [NIH website](#).

Additionally, Stage 2 Meaningful Use requirements for eligible hospitals, critical access hospitals, and eligible professionals specify that race and ethnicity codes must follow current federal standards as published by OMB:

[Eligible Hospital and Critical Access Core Requirements](#)

[Eligible Professional Core Requirements](#)

Disclaimer

- The material presented within this document is provided as part of the Collaboratory's commitment to disseminate information and knowledge acquired from the Collaboratory project as soon as possible.
- Conditions such as “diabetes” can be characterized using EHR data in many different ways. The computable phenotype definition for one purpose may not be a good fit for a different purpose. Any existing computable phenotype definition should be evaluated to determine whether or not it is a good fit for a particular use case.
- The material presented here has not been fully vetted or endorsed by the NIH, the Collaboratory Steering Committee, or all Collaboratory members.
- The information presented is continually evaluated and updated as new use cases, phenotype definitions, and phenotype validation results become known.

Prepared by: Rachel Richesson, Monique Anderson, Michelle Smerek

Reviewed by: The Phenotypes, Data Standards, and Data Quality Core of the NIH Collaboratory

Version: 1.1, last updated August 25, 2014