Sex 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

Definition

Sex is defined as the totality of characteristics of reproductive structure, functions, phenotype, and genotype, differentiating a male from a female person (adapted from the National Library of Medicine’s Medical Subject Headings definition for sex).

This data element is not recommended for the capture of the concept of “gender,” which is the assemblage of properties that distinguish people on the basis of the societal roles expected for the two sexes (NCI Thesaurus definition for gender).

Sex (select one)

- Male
- Female
- Unknown or not reported

Additional data elements for sex and also gender may be justified by individual study protocol objectives and proposed analyses.

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](https://www.nih.gov).

### 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.