

WHI Race and Ethnicity Language and Data Application and Interpretation Guide- Brief Overview

created by: The **WHI Race and Ethnicity Task Force** (see 9.1.2021 Guide)

A. Overarching principles:

Race and **Ethnicity** are distinct socio-political constructs that are **not** rooted in **biology**. **Ancestry** and **genetic admixture** are **not** interchangeable with “race” or “ethnicity”.

Ethnicity is distinct from **Race**. Both should appear in WHI characteristics tables.

B. WHI Race and Ethnicity Variables: Baseline WHI Form 2 [link] and Form 41 [link]

The **F41-mapped* dataset** (labeled 'f41_imputed_ctos_inv.dat' on the WHI website) should be used in all WHI analyses and papers. “**Other**” should only be used when a participant selected “other” as her identity. It should not be used to combine racial groups (or subgroups) for any purpose.

C. Race and Ethnicity Definitions and Terminology (Language)

A few basic language and writing guidelines are presented in the Guide, including terms which should and should not be used. Individuals should be referred to by their self-identified race and ethnicity. Terms used to group race and ethnicities should be well conceptualized in the manuscript.

D. WHI Cohort Representativeness of U.S. Women by Age and Race and Ethnicity

Authors should address the representativeness of the U.S. (older) female population, per Census data, of the WHI cohort when interpreting analytic results (Appendices B1 & B2 provide information to support this).

E. Geographic Origin and U.S. Regionalization Considerations

U.S. geographic diversity is particularly rich in WHI and may be relevant when focusing on race and ethnicity. Recruitment strategies varied across WHI Clinical Centers (see Appendices C1-C4).

F. Specific Considerations for Including Race and Ethnicity in WHI Analyses:

- Develop Questions & Methodological Strategies Informed by Conceptual Frameworks, e.g. Public Health Critical Race Methodology; National Institute of Minority Health and Health Disparities Research.
- Reporting of demographic data on race and ethnicity. WHI manuscripts should:
 - state that WHI participants self-identified their race and ethnicity (F41-mapped dataset);
 - provide rationale for use of race as a key variable; if race is the **primary exposure** of interest or where analyses are **stratified by race** and/or **ethnicity**;
- Analyses by Race and/or Ethnicity
 - Comparisons between race or ethnic groups should be informed by research questions.
 - Comparisons of race or ethnic groups to Non-Hispanic White participants is not required in WHI; this should only be done when supported by a research question; “all other race categories” should not be grouped (to increase sample size, for example) to be compared with White.
 - Within group analyses should acknowledge the heterogeneity within racial and ethnic groups.
- Data Interpretation & Reporting
 - **Statistical power for race and ethnicity subgroup analyses**, should be sufficient to detect differences by that group. **Authors** should acknowledge that sample selection limits interpretation of findings to the overall U.S. population or racial or ethnic subpopulations identified in the manuscript.
 - Over time, the **WHI sample composition has been influenced by selective drop-out** that can be investigated through the use of inverse probability weighting and other methods.
 - **Limitation:** Race and ethnicity are defined and interpreted within a socio-political framework as a proxy for both historical and ongoing differences in advantages arising from racism in social determinants of health, such as education, income, resilience and stressful life events (see WHI psychosocial constructs [link]). Other structural factors may be important for the data interpretation of racially disparate outcomes, but are not available in WHI.