DATA DOCUMENTATION
SPfile

The SPfile data file contains selected demographic variables, sample weights and administrative data. The SPfile dataset does not represent a separate interview component of the NYC HANES survey, though it includes demographic variables (variables with a DMQ or SFQ prefix) that were asked during CAPI.

Data Processing and Editing

Race/Ethnicity

- Racial and ethnic classifications were based on the OMB-designated Hispanic ethnicity category (**DMQ240**) and five minimum race categories [e.g., White, Black/African American, Indian (American)/Alaska Native, Native Hawaiian/Other Pacific Islander and Asian (**DMQ265A-E**)][1]. Responses not captured in one of these categories were recorded verbatim in a 'Some Other Race' category. Free-text responses for Some Other Race **(DMQ265OS)** were recoded *where possible* into one of the race categories **(DMQ265A-E)** and/or the Hispanic ethnicity **(DMQ240)** category based on definitions in the OMB 1997 Standards[1,2] and Census SF1[3]:

  - Free-text responses for Some Other Race indicating one or more races/countries of origin that are already classified as Hispanic or one of the standard races were first recoded into the relevant race and/or ethnicity category or categories. When individuals marked only Some Other Race and provided a write-in of two or more distinct races, the response was recoded to the race categories where possible and a primary race was assigned in **DMQ275Q** according to the first write-in[4]; **DMQ275G** was also recoded to indicate a primary race was chosen.

  - For responses of 'Hispanic', 'Latino' or a country indicating Hispanic origin (recoded to a 'Yes' response for Hispanic origin in **DMQ240** unless already indicated), the response was left in 'Some Other Race' if no other race category was specified. Responses indicating Hispanic ethnicity and a single race are not considered to be multiracial and were recoded out of **DMQ275G** and **DMQ275Q** if necessary[1]. The response of 'India'/'Indio' was understood to be a colloquial term indicating someone with skin color comparable to that of a person of Spanish American Indian descent, and in most cases confirmed by answers to Hispanic origin, country of origin or language of interview. These responses were recoded to Hispanic origin in **DMQ240** unless already indicated and responses remained in 'Some Other Race'.

  - **DMQ265F** was recoded to missing in instances where free-text responses were removed from **DMQ265OS**.

Analytic Notes

Analytic Notes are available for several variables in the SPfile dataset and are provided below by variable. Analysts are encouraged to also refer to NHANES Data, Documentation, and Codebooks on specific sections by component for other notes and documentation (http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/nhanes03_04.htm). In general, analysts are advised to carefully review question wording, skip patterns and check items to understand targets and eligibility for a particular item. Also note that questions involving a quantity and unit of measure have separate variables for each as indicated in the SPfile Variable List and should be integrated accordingly (these questions, if mentioned below, are referred to by the main variable name as it appears on the questionnaire).

'Previously Used Recodes' are provided for some demographics and outcomes if available. However, before using the recodes analysts must consider whether the recode is appropriate to their analysis and should review relevant literature and clinical guidelines for standard definitions of specific health outcomes. The list of 'Previously Used Recodes' is not exhaustive; the recodes are intended to help analysts understand how variables can be used in defining outcomes. Users may find that cutoffs or categories can be redefined using similar variables.

Analysts should refer to all materials under 'Using the Data' before analyzing the data. These materials provide guidance on how to use the data as well as the documentation available on the website. The Analytic Guidelines describe the specific weights in detail and when to use each one. The sample programs also illustrate examples for using each type of weight.

Race/Ethnicity

- This race/ethnicity variable **RACE_ETH** is derived by combining responses to questions on race (**DMQ265A-E**)] and Hispanic origin (**DMQ240**)[1]. The recode category "5" (Other race – includes multiracial) includes all remaining single non-specific race responses, the 'Multiracial' group who indicated more than one race but did not select a main race, those who indicated a verbatim response to non-specific multiracial heritage (e.g., mixed, Mulatto), and those with missing values on race.[5]

Education

- Note that **0** (not 1) is the lowest value.
- Consider combining 'GED or equivalent' (Value 14) with 'High school graduate' (Value 13).
- **DMQ140** can be recoded into three categories: 1) Less than high school; 2) High school and general equivalency diploma (GED); 3) More than high school. See 'Previously Used Recode' EDUC.

Marital Status

- The variable **SFQ180** can be recoded into a new variable with three categories: 1) Never married; 2) Married or living with partner; 3) Widowed, divorced, or separated. See 'Previously Used Recode' MARSTAT.

US Born vs Foreign Born

Data Documentation – SPfile

- **DMQ105** can be used to recode into US Born and Foreign Born. The US territories reported, including Puerto Rico, the Virgin Islands, and Guam, are available separately and can be included in US Born. See 'Previously Used Recode' US_BORN.

- **DMQ161** can be used to create a variable that indicates the length of time that a person who is considered foreign born has lived in the United States. See 'Previously Used Recode' FOR_BORN.

---

[1] Executive Office of the President/Office of Management and Budget (OMB) 1997 Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.
http://www.census.gov/population/www/socdemo/race/Ombdir15.html.
[2] US Census Bureau Racial and Ethnic Classifications Used in Census 2000 and Beyond.
http://www.census.gov/population/www/socdemo/race/racefactcb.html, 2000.
[3] US Census Bureau Summary File 1 2000 (SF1) appendix documentation of race codes.
http://www.census.gov/prod/cen2000/doc/sf1.pdf, Appendix G.
[4] US Census Bureau Questions and Answers for Census 2000 Data on Race. http://www.census.gov/Press-Release/www/2001/racegandas.html, 2001.
[5] NHANES 2001-2002 Data Files – Demographics\Documentation.
(http://www.cdc.gov/nchs/data/nhanes/nhanes_01_02/demo_b_doc.pdf)