WHAT YOU NEED TO KNOW: DATA RELEASE DOCUMENTATION

This document is designed to help data users navigate the NYC HANES analytic guidelines and dataset specific documentation. Analytic Guidelines and all sections of data documentation should be used concurrently to facilitate analysis of the NYC HANES data.

Analytic Guidelines

The Analytic Guidelines provide guidance on the use of NYC HANES datasets and sample weights, as well as other analytic issues. Additionally, sample programs are provided to illustrate the concepts presented in this document.

Data Structure

The Data Structure outlines general descriptions and sections of each of the following 6 datasets that comprise the NYC HANES public release data:

- **SPfile (Study Participant File)**
- **CAPI (Computer Assisted Personal Interview)**
- **ACASI (Audio Computer-Assisted Self-Interview)**
- **CIDI (Composite International Diagnostic Interview)**
- **EXAM**
- **LABS**

Data Documentation

Data documentation is provided for each dataset listed above and includes the following:

- **Variable List**
  The Variable List is the list of variables that are included in the dataset. Please note that the public release data do not necessarily include all of the questions asked in a section. Variables may have been removed due to confidentiality, quality, or other considerations. Thus, it is possible that a dataset does not completely match all of the questions asked in a questionnaire section.

- **Questionnaire**
  Questionnaires are provided for CAPI, ACASI and CIDI. Please note the questionnaire skip patterns and check items which provide information about targets and eligibility for a particular item.

- **Dataset-specific documentation**
  Each dataset is accompanied by documentation detailing interview setting and mode of administration, quality assurance and quality control techniques, data processing and editing, and analytic notes.

- **Codebook**
  The codebook contains question text, response codes and frequency distributions for each variable available in the public release file. It is strongly suggested to use the codebook in conjunction with the questionnaire. The following table provides a

description of the codebook content:

Codebook Field	Description
Variable Name	Name of variable as it appears in the dataset
SAS Format Name	Used to apply user-defined formats for variables, which are provided in the SAS program 'SAS variable formats.sas'.
SAS Label	Brief description of the variable
Question	Actual question text NOTE: The LABS and EXAM components do not have questions associated with them; in cases where questions are not applicable, see Variable List for description.
Instruction	Specialized instructions for the interviewer or examiner while conducting the interview or examination. NOTE: The instructions are for the interviewer or examiner ONLY and are not read as part of the question to the respondent / examinee.
Value	The possible responses for a variable.
Value Label	Description of the possible responses for a variable. NOTE: The Value Label for continuous variables is referred to as "Range".
Frequency	Frequency of occurrence.
Cumulative Frequency	Cumulative frequency (or accrued total) of occurrence.
Percent	Reflects the occurrence (unweighted) per 100 of the study population.
Weighted Percent	Reflects the weighted occurrence per 100 of the weighted population.
Skip to	Skip to field indicates the item to which participants with specific responses skip. Skip items included either of the following: 1. Another variable, as indicated by the variable name. 2. A 'Check Item', as indicated by the 'Check Item' name and 'CI' in parentheses. A 'Check Item' incorporated previously reported information about responses given in prior sections or component items. This information is checked by a computer algorithm that automatically executes a skip pattern. 3. End of Section, as indicated by EOS. If the response resulted in EOS, the participant moved on to the next section within the component.
Not asked or missing	Indicates the number of study participants who were not asked the question (due to a skip pattern or check item instruction) or who had a missing response.

Recodes and Formats

A program (*SAS Formats for SPfile.sas*) that includes the SAS formats for the variables in the NYC HANES study participant file (SPfile) dataset is provided. Additionally, SAS code and corresponding SAS formats are available for recodes that have been previously used in analyses of the NYC HANES analyses. Before using the recodes, analysts MUST consider whether the recode is relevant to their analysis by reviewing relevant literature and clinical guidelines for appropriate 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 of interest. Users may find that cutoffs or categories should be redefined using similar variables.