

DATA DOCUMENTATION

CAPI

(Computer-Assisted Personal Interview)

Interview Setting and Mode of Administration

The information for this component was collected using a computer-assisted personal interview (CAPI) system through a face-to-face interview conducted at the health clinic or in the home.

Pre-translated interviews were conducted in English and Spanish. As needed, interviews were conducted in additional languages using a staff member or family-member proxy or a telephone-based translation service.

Information was recorded using a computer-assisted data entry method. Some of the questions in this section required the use of printed hand cards. In most cases, the respondent selected a response choice after looking at the hand card.

Please refer to the [CAPI survey questionnaire](#) and [general descriptions of each section](#).

Quality Assurance and Quality Control

The computer CAPI system had built-in consistency checks, including range checks, valid response checks and checks for logical relationships, an online information screen that provided the interviewers with information about the terms used in the questionnaires, and hard and soft edit checks to reduce the number of keying entry errors. Questionnaire data was systematically reviewed for logical, processing, and data input errors on a periodic basis during the process of data collection. The data have been further edited for completeness, consistency, illogical values and analytic usefulness.

Data Processing and Editing

Occupation (OCQ)

- Open-ended responses (not released) were collected on the type of work done (**OCQ240**) and the specific duties performed (**OCQ250**). US Census Bureau coding specialists classified responses into the Census 2003 Occupation code categories (<http://www.census.gov/hhes/www/ioindex/overview.html>) which were developed to be consistent with the Standard Occupational Classification (SOC) Manual: 2000 (<http://www.bls.gov/soc/>). This recoded occupation variable (**OCCREC**) is included in the dataset.

Physical Activity and Fitness (PAQ)

Data Documentation - CAPI

- Respondents who answered 'Yes' to having done any vigorous activity for at least 10 minutes (**PAQ206**), but indicated 0 times in **PAQ280** or less than 10 minutes per time in **PAQ300** were recoded to 'No' in **PAQ206** and missing in **PAQ280** and **PAQ300**.
- Respondents who answered 'Yes' to having done any moderate activity for at least 10 minutes (**PAQ326**), but indicated less than 10 minutes per time in **PAQ420** were recoded to 'No' in **PAQ326** and missing in **PAQ420**.

Analytic Notes

Analytic Notes are available for several variables in CAPI and are provided below by section. 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 [CAPI 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). Additionally, analysts may find that CAPI variables should be utilized in conjunction with variables from other survey components (e.g., Exam, Labs).

'[Previously Used Recodes](#)' are provided for some measures 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.

Occupation (OCQ)

- The recoded occupation variable **OCCREC** (see Data Processing and Editing) can be combined into 7 commonly used groups (see 'Previously Used Recode' OCCGRP):
 - Managerial and professional specialty
 - Sales
 - Technical and administrative
 - Service
 - Farming, forestry, and fishing (though only one SP in this group)
 - Precision production, craft, and repair
 - Operators, fabricators, and laborers

Smoking and Tobacco Use (SMQ)

- **SMQ020** and **SMQ040** can be combined to form a recode for smoking status with three categories including Never Smoker, Former Smoker or Current Smoker. See 'Previously Used Recode' SMOKER.

Blood Pressure (BPQ)

- Based on published literature, a commonly used definition of hypertension incorporates blood pressure exam measurements (**BPXSAR** and **BPXDAR** from the Exam component) with prescribed anti-hypertensive medication use (**BPQ051a**). Hypertension awareness, treatment and control can be estimated among individuals with hypertension using **BPQ020**, **BPQ051a** and the blood pressure exam measurements. See 'Previously Used Recodes' for hypertension.
- Based on published literature, a commonly used definition of hypercholesterolemia incorporates total blood cholesterol lab measurements (**LBXTC** from the Labs component) with prescribed cholesterol-lowering medication use (**BPQ100d**). Hypercholesterolemia awareness, treatment and control can be estimated among individuals with hypercholesterolemia using **BPQ080**, **BPQ100d** and the total cholesterol lab measurement. See 'Previously Used Recodes' for hypercholesterolemia.

Physical Activity (PAQ)

- The series of questions (**PAQ206-PAQ420**) about how often participants do vigorous or light/moderate physical activity of at least 10 minutes duration and for about how long these sessions generally last can be used to categorize adults into physical activity status. Based on the definition of physical activity used in the National Center for Health Statistics (NCHS) [Health, United States, 2005 With Chartbook on Trends in the Health of Americans](#) (page 522), vigorous and moderate activities can be combined to classify adults as having done one of the following (see 'Previously Used Recodes' for physical activity):
 - Regular leisure-time activity if they reported at least three sessions per week of vigorous leisure-time physical activity lasting at least 20 minutes in duration or at least five sessions per week of light or moderate physical activity lasting at least 30 minutes in duration.
 - Some leisure-time activity if they reported at least one session of light/moderate or vigorous activity but did not meet the requirement for regular leisure-time activity.
 - No leisure-time activity (inactive) if they did not report any sessions of moderate or vigorous leisure-time physical activity of at least 10 minutes duration or reported they were unable to perform leisure-time physical activity.
 - Outliers are considered to be those who reported having done vigorous or moderate activity greater than or equal to 28 times/wk OR greater than or equal to 720 min each time (Healthy People 2010 Operational Definitions). Users should carefully inspect the data for other values they may consider too high.

Hospital Utilization and Access to care (HUQ)

- **HUQ030** and **HUQ040** can be combined to include response of 'Hospital emergency room' as having no routine place for care. See 'Previously Used Recode' RTNPLCE.

Diabetes

- Based on published literature, diabetes estimates are generally presented as diagnosed diabetes and undiagnosed diabetes (See 'Previously Used Recodes' for diabetes):
 - Diagnosed diabetes, or self-reported diabetes, is based on participant report that a health care provider had ever told them that they had diabetes (other than during pregnancy for women) (**DIQ010**).
 - Participants without a prior diagnosis but whose fasting plasma glucose (**LBXGLU** from the Labs component) is greater than or equal to 126 mg/dL were considered to have undiagnosed diabetes.
 - Total diabetes is the combination of the measures for diagnosed and undiagnosed diabetes (i.e., participants who self-reported diabetes or whose fasting plasma glucose was greater than or equal to 126 mg/dL).

Medical Conditions (MCQ)

- For colonoscopy screening, consider both **MCQ470** and **MCQ480**. See 'Previously Used Recodes'.

Health Insurance (HIQ)

- **HIQ030** and **HIQ200** should be used in combination to determine participant's type of health insurance coverage. Individuals may have reported two types of insurance coverage. Analysts can use a chronological determination to categorize participants' coverage. For instance, if interested in classifying private insurance first, Medicare second, then government, participants reporting any private health insurance would be classified as having private insurance. Otherwise, persons who reported having Medicare would be classified as having Medicare, and other government coverage would include all remaining participants who reported Medicaid or other government program. See 'Previously Used Recodes'.
 - Some participants ages 20-64 reported Medicare as their health insurance. An option is to include these participants in the government category and only including persons aged > 65 who reported having Medicare in the Medicare group.

Alcohol Use (ALQ)

- **ALQ110**, **ALQ120**, and **ALQ130** can be used to determine 'heavy drinking' which can be defined as an average of more than 2.0000 drinks per day for men and an average of

Data Documentation - CAPI

more than 1.0000 drinks per day for women. Never drinkers can be a separate category or included in the 'non-heavy drinking' group. See 'Previously Used Recodes' for heavy drinking.

- **ALQ120** and **ALQ130** are used to calculate the average number of drinks per day using the following formula:

$$\text{Drinks} = \frac{(\# \text{of days drank}) * (\text{average number of drinks on those days})}{(\text{number of days in reported unit})}$$

- **ALQ110**, **ALQ120**, and **ALQ140** can be combined to determine if participants had 5 or more drinks on one or more days in the past 12 months. Never drinkers can be excluded or put in a separate category. See 'Previously Used Recodes' for binge drinking.
- **ALQ110**, **ALQ120**, and **ALQ140** can be combined to determine an indicator for 'binge drinking' by creating a continuous variable for the number of times the participant had 5 or more drinks on one day in the past 12 months. Never drinkers can be excluded or put in a separate category. See 'Previously Used Recodes' for binge drinking.

Income (INQ)

- An income screener question was asked (**INQ640**) to query if the total family income was < \$20,000 or ≥ \$20,000. If the respondent answered **INQ640**, a follow-up question asked the respondent to select an income range from a list of income ranges on a printed hand card; the midpoint of the income range value was used as the total family income value (**INQ650**). **INQ650** is separated into two variables with letter response categories. See 'Previously Used Recodes' for income. Note that because data are not available on the number of individuals in a family, a poverty index cannot be determined.