Guidelines for MIDUS 2 Milwaukee Datasets

Naming, Coding and Formatting Conventions

We have developed guidelines for naming and coding conventions for the MIDUS 2 Milwaukee data. The attached pages provide specific guidelines for how variables will be named in MIDUS 2 Milwaukee. We have also included coding conventions for "yes/no" variables as well as for "don't know" responses, "refused/missing" data, and "inapplicable" codes. Guidelines for variable formats and construction of files for repeated measures (e.g., diary data) are also provided.

Note: All project leaders will be responsible for delivering cleaned, coded SPSS data files to the Administrative Core. Each project is responsible for creating constructed variables. These variables should be included in early releases of the data as well as the final data file. We recommend sending an early draft or preliminary dataset to the Administrative Core for review before you make final data deliveries.
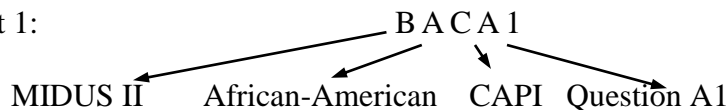
I. Variable Naming Conventions

### A. Short Variable Names (SVNs): First 8 Characters

Notes:
- The first 8 characters (or less) must be unique for each variable.
- If you are doing analysis and/or programming code, using SVNs early in the process will prevent you from having to rename the variable names in your work.
- Because the Milwaukee data collection for Project 1 used a different data collection protocol, the variable names, sequence, and locations do not parallel the longitudinal MIDUS variables. Thus, Milwaukee data from Project 1 (and only Project 1) use a slightly different naming convention in which an "A" is substituted for a "1" to indicate Project 1 Milwaukee data:

Project 1:   B A C A 1

        MIDUS II   African-American   CAPI   Question A1

- For Milwaukee data gathered in Projects 2 through 5, the standard MIDUS naming conventions apply. The first 3 characters of each variable name will identify the longitudinal wave in which the data were collected, the specific project, and the instrument used to collect the data. Characters 4 through 8 will identify the specific item or scale score variable that is represented by the variable name.

Projects 2 thru 5:   B 4 Q A 1

        MIDUS II   Project 4   SAQ   Question A1
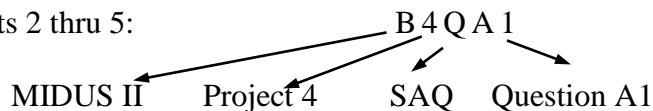
- See "M2_Variable Naming and Coding Conventions_10-26-05.doc" for more details.

II. Variable Labeling Conventions

We decided upon a *40 character limit for labels* (considering the limitations across statistical applications) and we are using mixed case text for more sentence-like descriptions.

Example: **How is your physical health?**

III. Variable Formats

1. Utilize "Numeric" whenever possible; avoid using raw string variables. Open-ended responses, text and verbatim data should be numerically coded. Raw text or qualitative data can be formatted as a text file or spreadsheet.

2. Variable formats should be precise - they should not exceed the maximum number of digits possible for a response. Thus, if a response will be 1-2 digits (i.e., something on a scale of 1-10) then the variable should be formatted as 2 digits.

3. Decimals: Limit to 2 places after the decimal point, unless important details of the data require more places be utilized.

4. Date/time formats: Dates or times recorded by *project staff* for administrative purposes (e.g. date & time Medical History was completed) can be submitted to the Core in the date and time formats specified in the established conventions for MIDUS (e.g., mm/dd/yyyy, or hh:mm:ss, etc.). Dates or times reported by *respondents*, especially in self-administered materials, cannot be submitted in date and time formats due to problems in applying missing value codes in SPSS. The individual components of a date or time must be recorded separately in their own variables. For example, dates will be broken into month, day and year variables, while times will be broken into hours, minutes and meridian (am/pm etc.)

IV. Value Labeling Conventions

A. Format

All value labels will be **UPPER CASE**.

> Example:
> **1 = YES**
> **2 = NO**
> **7 = DON'T KNOW**

B. Coding Conventions for Yes/No Responses

> **YES = 1**
> **NO = 2**

C. Coding Conventions for "Non-response" (Don't Know, Missing Data, and Inapplicable)

> 9's will be used as place holders (i.e., if a variable is 3 digits, use 9 to fill in the first 2 places, then use 7, 8, and 9 for the final digit as displayed below).
>
> Variables should have system-missing only if a R did not complete an instrument as requested. E.g., in Project 1 individuals who do not complete the SAQs are assigned system-missing values for all SAQ variables. This is consistent with procedures followed in MIDUS I.
>
> DON'T KNOW (7's) - 7, 97, 997, 9997, etc.
> > Used to indicate explicit "Don't know" responses (where a specific response option of "Don't know" was offered).
>
> REFUSED/MISSING (8's) - 8, 98, 998, 9998, etc.
> > Used to indicate R did not provide a response to a particular question.

INAPPLICABLE (9's) - 9, 99, 999, etc.
> Used to indicate that R was not asked a particular question. This will occur most often because of skip patterns that the R is asked to follow, or questions that R determines do not apply to him or her.

V. File Formats for Repeated Measures (e.g. diary data)

We recommend the construction of horizontal data files (rather than vertical) for repeated measures (i.e., salivary cortisol samples collected on 4 different days, 4 times each day, would result in 16 different variables, each with a unique name).