

ICPSR 4652

**National Survey of Midlife  
Development in the United States  
(MIDUS II), 2004-2006**

Naming, Coding, and Formatting Conventions

Inter-university Consortium for  
Political and Social Research  
P.O. Box 1248  
Ann Arbor, Michigan 48106  
[www.icpsr.umich.edu](http://www.icpsr.umich.edu)

# **National Survey of Midlife Development in the United States (MIDUS II), 2004-2006**

Carol Ryff

*University of Wisconsin-Madison*

David M. Almeida

*Pennsylvania State University*

John S. Ayanian

*Harvard University*

Deborah S. Carr

*University of Wisconsin-Madison*

Paul D. Cleary

*Harvard University*

Christopher Coe

*University of Wisconsin-Madison*

Richard Davidson

*University of Wisconsin-Madison*

Robert F. Krueger

*University of Minnesota*

Marge E. Lachman

*Brandeis University*

Nadine F. Marks

*University of Wisconsin-Madison*

Daniel K. Mroczek

*Purdue University*

Teresa Seeman

*University of California-Los Angeles*

Marsha Mailick Seltzer

*University of Wisconsin-Madison*

Burton H. Singer

*Princeton University*

Richard P. Sloan

*Columbia University*

Patricia A. Tun

*Brandeis University*

Maxine Weinstein

*Georgetown University*

David Williams

*University of Michigan*

# Terms of Use

The terms of use for this study can be found at:  
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4652/terms>

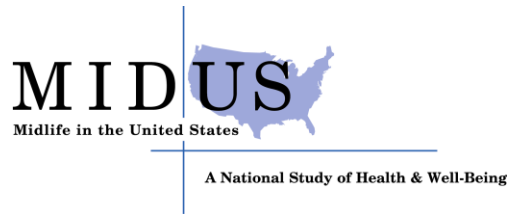
## Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

### NOTICE

#### WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.



## Guidelines for MIDUS II Datasets

### Naming, Coding and Formatting Conventions

We have developed guidelines for naming and coding conventions for the MIDUS II data. Such conventions are necessary because we will need to merge records across waves of data collection (MIDUS I and II) as well as across Projects 1-5. Without standardized conventions, this task will be unmanageable.

The attached pages provide specific guidelines for how variables will be named in MIDUS II. 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 (or less), CAPITALIZED

### 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.
- 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. Thus, “item identifiers” are limited to 5 characters.
- The letters/numbers in **bold** font, in the list below, represent the characters to be used in the variable name.

1<sup>st</sup> Character: **B** (“B” is being used to designate the variable belongs to the second wave of MIDUS – all variables must start with this letter).

2<sup>nd</sup> Character: Number (Project Identifier: i.e., **1, 2, 3, 4, 5** – the second character of each variable must identify the source Project). The Milwaukee subsample will use the character **A** to distinguish it from other Project 1 data.

3<sup>rd</sup> Character: Letter (Instrument Identifier: *type* of instrument, or, *name* of instrument; i.e., **Phone Interview, Computer Assisted Personal Interview (CAPI), Self-Administered Questionnaire, Cognitive Battery, Diary, Physical Examination, Biology, EEG**; Letters that can be mistaken for numbers should be avoided, i.e., “I”, “L”, and “O”. Note that these Letters must be mutually exclusive *within* Projects.

4<sup>th</sup> Character: Letter/Number (Item Identifier/Scale Name)

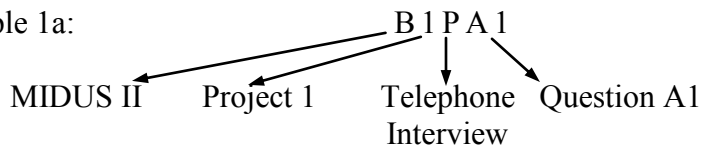
5<sup>th</sup> Character: Letter/Number (Item Identifier/Scale Name)

6<sup>th</sup> Character: Letter/Number (Item Identifier/Scale Name)

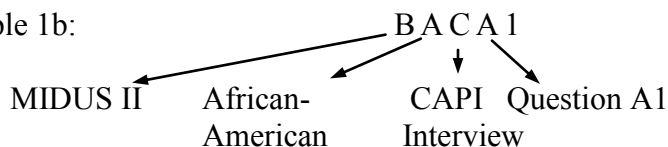
7<sup>th</sup> Character: Letter/Number (Item Identifier/Scale Name)

8<sup>th</sup> Character: Letter/Number (Item Identifier/Scale Name)

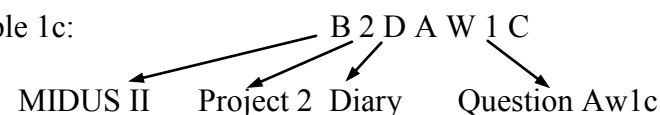
Example 1a:



Example 1b:



Example 1c:



## **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).