

ICPSR 2760

Midlife in the United States (MIDUS 1), 1995-1996

Documentation of Coded Text Responses in
MIDUS 1

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Midlife in the United States (MIDUS 1), 1995-1996

Orville Gilbert Brim

*John D. and Catherine T. MacArthur Foundation. Research Network
on Successful Midlife Development*

Paul B. Baltes

Max-Planck-Institut für Bildungsforschung.

Larry L. Bumpass

University of Wisconsin

Paul D. Cleary

Harvard Medical School

David L. Featherman

University of Michigan

William R. Hazzard

Wake Forest University

Ronald C. Kessler

Harvard Medical School

Margie E. Lachman

Brandeis University

Hazel Rose Markus

Stanford University

Michael G. Marmot

University College, London. Medical School

Alice S. Rossi

University of Massachusetts at Amherst

Carol D. Ryff

University of Wisconsin

Richard A. Shweder

University of Chicago

Terms of Use

The terms of use for this study can be found at:
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2760/terms>

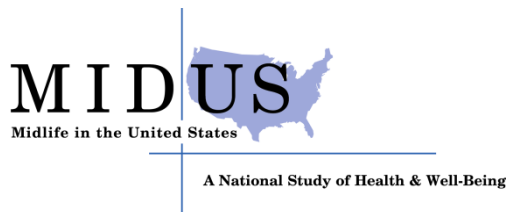
Information about Copyrighted Content

Some instruments administered for studies archived with ICPSR 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.



Documentation of Coding Text Responses in MIDUS 1

Background. This document describes how open-ended textual responses in the MIDUS 1 CATI and SAQ were transformed into categorical numeric codes. These codes are included in a stand-alone dataset (M1_P1_CodedTextData_N3950_20170329.sav). This dataset contains only those cases (N=3,950) that contained text data in their responses.

Several items in the MIDUS 1 Project 1 survey ask participants to answer a question by providing nominal, categorical response options. For example, one item in the telephone interview asks participants about what types of cancer they had. Participants are provided with a variety of response options, such as “brain cancer,” “lung cancer,” “liver cancer,” etc. With several exceptions, the instruments provide structured response options or metrics to capture answers to these types of questions. Building on the previous example, the phone instrument allows interviewers to record respondents’ answers to the “type of cancer” question using 10 possible categories. These categories are labeled “breast cancer,” “cervical cancer,” “colon or rectal cancer,” “lung cancer,” “lymphoma or leukemia,” “ovarian cancer,” “prostate cancer,” “skin cancer, melanoma,” “uterine cancer,” and “other ____ (specify).”

The instrument’s structured list of possible responses typically includes a response option of “other ____ (specify).” This response option allows the capture of responses not already provided by the instrument. During the telephone interview, when respondents provided a nominal response that did not belong to any of the provided categories, the interviewer would type the respondents’ answer into the “other ____ (specify)” field. For the SAQ, the respondent would write their response in the space provided under the “other ____” field.

Note that some questions were true open-ended measures without a structured list of response options. Item A35 in the phone interview, for example, asks respondents the treatment or therapy for high blood pressure. Respondents were simply asked to volunteer their own responses to this question.

To make the M1 text data more usable, researchers at the University of Wisconsin’s Institute on Aging categorized these text data. The remainder of this document explains the process by which text responses were coded.

The Process.

1. Preparing the Data for Coding. The verbatim text data from the M1 Project 1 phone interview and SAQ were delivered to the University of Wisconsin’s Institute on Aging from the University of Wisconsin Survey Center in Microsoft Excel files. These Excel files were cleaned

to ensure that participants with “other specify” responses were represented in only one row of the spreadsheet, and that empty or duplicate rows were removed.

2. Coding the data. The categorization of text responses was accomplished using a text mining software program called IBM SPSS Text Analytics for Surveys. The extraction function of the SPSS-TA program served as a starting point for coding the responses. This extraction function is the default tool in the software program and identifies identical words, synonyms, and themes throughout the text data. Using this extraction tool, responses were coded into discrete categories by operating under the following guidelines:

- (1) Each unit of information or theme from the text response field was coded once (i.e., was placed in one category only). However, in order to recognize that participants’ text responses could include more than one unit of information, responses could be coded in more than one category for a single question. Multiple responses were recorded in multiple variables. For example, most participants responded to question A35 with more than one theme such as “exercise and low sodium/sugar”. This response contains two units of information: (1) exercise, and (2) low sodium/sugar. This participant’s responses were accordingly assigned two codes in two variables: one variable contained a code for their response of “exercise” and another variable contained a code for their response of “low sodium/sugar.”
- (2) Categories were created such that their meanings would resemble the respondent’s original answer as precisely as possible. For example, for the SAQ item S1 regarding participants’ main ethnic background, several respondents listed specific countries in Northern Europe, such as Finland and Denmark. Even though these responses refer to nationalities and not ethnicities, separate categories for each of these countries were created (i.e., categories of “Denmark,” “Finland”).
- (3) Categories of responses that did not exceed the 2% of the total valid responses were categorized as “Other.”

Generally, we use the same categories, codes, and value labels for the same OS variables across datasets. This facilitates combining OS variables from M1, M3, or MR datasets. When creating new categories for a variable derived from the OS text responses, the category must meet the 2% rule, meaning that there should be a sufficient number of cases to warrant creating a new code to represent those responses. Any new codes that meet this 2% rule can be added to an existing SPSS TA variable scheme or dictionary. Any response that do not satisfy the 2% criteria should be coded as “Other.”

The only exception to this rule is whether a response option already exists for the question. For example, if less than 2% of responses volunteer “Austria” as an Other Specify response to an ethnicity question yet “Austria” is listed as a valid response option for that question, those cases’ responses can be coded into an “Austria” category for the coded-text variable. The reason is that any cases with an “Austria” code for the text variable can be combined with the extant “Austria” codes for the non-text variable.

At M2, content domain experts were consulted for help coding specific items. For example, the codes created types of cancer were reviewed by a biomedical researcher and a nurse practitioner. The codes created for items related to employment were reviewed by a sociologist with a background in labor force participation. The categories and codes derived from this process at M2 were used again for the M3 and M1 text data.

3. Packaging the newly coded information. Once all responses were coded in the SPSS-TA program, variables were exported to SPSS. These newly created “OS” variables were given variable names, variable labels, and value labels according to the following conventions:

- (1) The variable names for the newly created “OS” variables parallel those of the original variables, but if there is more than one newly created “OS” variable for a single item, the newly created “OS” variables can be distinguished from each other by their having an “A,” “B,” “C,” etc. at their end.
- (2) All “OS” variables have variable labels that begin with “OS-”. Variables for open-ended questions have variable labels that start with “OE-”.

For additional information regarding M1 text data, please contact Barry Radler at bradler@wisc.edu.