

ICPSR 29282

**National Survey of Midlife
Development in the United States
(MIDUS II): Biomarker Project,
2004-2009**

Carol D. Ryff
University of Wisconsin-Madison

Teresa Seeman
University of California-Los Angeles

Maxine Weinstein
Georgetown University

Project 4 Readme Data File Notes

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

Terms of Use

The terms of use for this study can be found at:
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/29282/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.

MIDUS II BIOMARKER PROJECT (P4) DATA FILE NOTES

June 1, 2012

This document highlights aspects of the MIDUS Biomarker data that analysts should be aware of prior to working with the indicated section of data. For example, there are instances when variables were added, or sections of an instrument were expanded for data entry purposes to accommodate additional information provided by the respondent. General issues and information that apply across multiple types of data are listed first. Issues that are unique to specific instruments appear at the end of this document.

Detailed information about specific sections of the data can be found in the documentation files describing the:

- Blood, Urine and Saliva Data
- Bone Health Data
- Medication Data
- Physical Exam Data
- Psychophysiology Protocol
- Psychosocial Constructs and Composite Variables
- Sleep Data

Also included among the documentation files are an overview of the Biomarker Project (P4) protocol (Biomarker Project Summary) and two data collection instruments; the Self-Administered Questionnaire (SAQ) and the Medical History.

Variable Names

MIDUS variable naming and coding conventions (See “Naming and Coding Conventions” in the general MIDUS documentation) specifies that the third character of the variable name be a letter that identifies the type, or name, of the instrument used to collect the data. The Biomarker (P4) data types, instruments, and data collection methods are designated by the indicated letters:

Z = Biomarker (P4) Administrative Variables
Q = Self-Administered Questionnaire (SAQ)
S = Pittsburgh Sleep Questionnaire (PSQ)
H = Medical History
P = Physical Exam (Short & Long versions)
X = Medication Chart
O = Bone Questionnaire
B = Biomarker Assays (Blood, Urine, Saliva)
T = Tissue Specimen Collection Form
A = Daily Sleep Diary
W = Actiwatch Watch data
D = Bone Densitometry Scan data
V = Psychophysiology

Administrative Variables

The following administrative variables appear at the beginning of the file:

- Data Collection Site (B4ZSITE)
- Month and Year the General Clinical Research Center (GCRC) visit was completed (B4ZCOMPM, B4ZCOMPY)
- Variables indicating whether complete Blood (B4ZBLOOD), Urine (B4ZURINE), and Saliva (B4ZSALIV) samples were collected.
- Variables indicating the Time Zone the respondent lived in at the time of the P4 visit (B4ZRZONE) as well as the Time Zone for the data collection site (B4ZSZONE). These were created to provide a means of controlling for variation in saliva cortisol level that could be due to changes in time zone for data collection.
- Variables indicating the time elapsed between the Biomarker (P4) Completion date and the Project 1 (Survey) Phone Interview, SAQ, and Cognitive Battery dates (B4ZB1PLG, B4ZB1SLG, B4ZB1CLG).
Note: Individuals from the African American sample, in contrast to the original MIDUS sample, were eligible for P4 regardless of whether they completed the Project 1 (Survey) SAQ. P4 Participants from the African American sample who had not previously completed the Project 1 SAQ were asked to do so at the time of the clinic visit. Thus, these individuals have a value of “0” for B4ZB1SLG.
- Respondent age when the Biomarker data was collected (B4ZAGE)

Filter Variables

Filter variables can be used to identify subsets of cases to include or exclude from analysis. Some of the administrative variables above can be used as filter variables. We also created a few filter variables for use with specific sections of the data as follows. Additional information about filter variables can also be found in the documentation files identified above.

- 1) Urine assay results may be affected if the sample was collected for fewer than 11 hours or more than 13 hours. They may also be affected if voids were missed. Analysts can identify cases that may have questionable catecholamine or urine cortisol values using the following variables:
 - B4BUPROB - Urine collection period less than 11 hours or greater than 13 hours.
 - B4BUSTRT - Urine collect: Start time.
 - B4BUEND - Urine collect: End time.
 - B4BUVDYN - Any voids not collected.
 - B4BUVDN - Number voids not collected
- 2) Some respondents from UW-Madison are missing all or part of the Actiwatch and Daily Sleep Diary data. The following filter variables were created to identify respondents with partial or missing data, as well as those cases where watch values were imputed or based in part on event marker button times.

B4AWAVL – categorical variable indicating whether daily sleep diary and/or Actiwatch® data are available for a given case.

B4AWIMPU – dichotomous Yes/No variable indicating whether there is missing data in the Daily Sleep Diary which required that rest intervals be marked by using imputed values.

B4AWMARK – dichotomous Yes/No variable indicating whether there is a discrepancy between the daily sleep diary (self-reported) and Actiwatch® data that required rest intervals be marked by using imputed values.

B4AWDAYS – indicates the number of days of sleep data available for a given respondent. All participants are expected to have 7 days of sleep data.

B4AWPART – dichotomous Yes/No variable indicating whether there is there is partial watch data. A case is flagged as having partial data if the respondent did not wear the watch for the full 7 days of the collection period because they: 1) took it off for a few hours, 2) forgot to put it on until a day or two into the collection period, 3) or took it off a day or two early.

B4AWIDIO – dichotomous Yes/No variable indicating whether or not the respondent had an idiosyncratic sleep pattern. Sometimes this is due to the respondents work schedule, but may be due to an illness (self or family member), or traveling.

B4AWTMZN –dichotomous Yes/No variable indicating whether or not the respondent traveled and slept outside their usual time zone during the data collection period or if the watch data collection period include a Daylight Savings Time change.

B4AWLAG – indicates the number of days between the date the GCRC visit was completed and the date the watch data collection began.

Details about the protocol for processing Actiwatch data can be found in the Sleep Data documentation file.

Added Variables

1) One Item – Multiple Variables:

The format of the following questions in the indicated instruments requires that multiple variables (as noted) be created in order to adequately capture the complexity of the data.

a. Medical History: Current Health Practices: Diet & Exercise (p. 12 & 14)

i. “Please estimate your daily calcium intake” (Q17).

The interviewer asks the respondent how many servings Milk (Q17a), Yogurt (Q17b), and Cheese (Q17c) s/he consumes on a daily basis.

ii. “On an average DAY, how many 8 ounce cups or glasses do you drink of

Coffee with caffeine, Tea with caffeine, Other beverages with caffeine (e.g. Coke)? (Q18a-c).

Many respondents consume the above on a daily basis, but others consume them on a weekly, monthly, and even yearly basis. Thus, responses to these questions are

entered as two variables. The first variable is the number of times in a given time frame that a person consumes the indicated item. The second variable is a categorical variable reflecting the commonly mentioned time frames (day, week, month, year).

These variables have the following format:

B4H17_F or B4H18_F – frequency

B4H17_T or B4H18_T – time frame categories

iii. “....do you engage in regular exercise, or activity, of any type for 20 minutes or more at least 3 times/week?” Q25a-c. If the respondent says “Yes”, the interviewer obtains information about the type of activity including the frequency with which the activity is performed.

1. Sometime respondents report performing activities more than once in a given day several times per week (e.g. walking in the morning and evening 5 days per week). This information is entered using 2 variables for each type of exercise, one reflecting the number of times per day and the other reflecting the number of days per week. These variables have the following format:

B4H25_FD – how many times per day

B4H25_FW – how many times per week

2. Sometimes respondents indicate that certain activities are seasonal (e.g. gardening, snow skiing, bicycling). To capture this information, we added a set of variables reflecting seasonality of an activity. The options include “Not Seasonal” for year round activities, to codes for individual seasons or combinations of seasons (e.g. Spring & Summer) for each activity reported. These variables have the format B4H25_S where the 6th character is a-j as appropriate.

b. Medication Chart: The following columns require two variables to capture the relevant information.

Drug Dosage - the drug dosage has two components the quantity and the unit (e.g. 200 mg or 3 tablets)

Frequency – the frequency with which a respondent takes a given medication has two components “how often” and a time frame (e.g. 2 times per day or every other day).

Taken for how long? – the period of time in which the respondent has been taking a given medication has two components quantity and time frame (e.g. 6 months, 10 years)

These three pieces of information are entered using 2 variables each. The first variable is the amount or frequency, the second variable is categorical and specifies the valid responses for the unit (e.g. mg, tablet, puff etc.) or time frame (e.g. day, week, etc) respectively. These variables have the following format, where the fourth character indicates the medication type and the final character indicates the medication number:

B4X_DD_ – dosage
B4X_DDU_ – dosage units

B4X_F_ - frequency
B4X_FU_ – time frame categories

B4X_T_ - how long
B4X_TU_ – time frame categories

2) Additional Others or Events:

There are several sections in the Medical History in which respondents have the opportunity to report “Other” symptoms or conditions or are asked about “events” of a certain type. There are a few instances in which respondents regularly report more conditions, or events than space allows for. Thus, at the end of the following sections of data additional variables are included to facilitate inclusion of this information in analyses. Specifically, a variable that reflects the total # of ‘other’ or ‘events’ was created. If this number is greater than the number of spaces provided in the medical history, data staff record all the information about these additional occurrences in open-ended string variables. Coding of these data is pending. The new count variables specified below immediately follow the original variables in the data set.

- a. Symptoms & Conditions (p. 2-3): Have you had any of the following conditions/illnesses? Items Q1a- Q1w list specific conditions (e.g. heart disease, cancer, diabetes etc.). Items Q1x –y (“Other? Please specify:”) allow additional symptoms and conditions to be listed. The new variable has the following format:

B4H1TOTH - # of ‘other’ conditions specified

- b. Family Medical History (p. 10-11): Has anyone in your immediate family (blood relatives only) had ...? Items Q15a-q list specific conditions (e.g. high blood pressure, stroke, alcoholism). Items Q15r-s (“Other? Please specify:”) allow additional conditions to be listed. The new variable has the following format:

B4H15OTH - # of ‘other’ conditions specified

- c. Other Events (p.24): Are there any other things, either positive or negative that have happened to you or your family or close friends since you completed the MIDUS Phone Interview in (_____) that stand out in your memory? If the respondent says “Yes” to this question (Q57) the interviewer asks “What has happened and when?” The interview booklet allows space to enter information for up to 5 events (Q57a1-5). The new variable has the following format:

B4H57TOT - # of events listed

3) Responses Exceeded the Space Provided:

- a. Medication Chart: The number of medications that a respondent takes, or the number of medication allergies reported, may exceed the space provided. Rather than create

a new set of variables for each type of medication, a set of variables was created at the end of the file for recording information about all additional medications or medication allergies. Specifically, a Yes/No variable is included to indicate if the respondent has additional medications. If 'Yes' then the information about the additional medications or allergy medications is recorded in open-ended string variables.

Effective June 2012 coding of the additional medications has been completed. As part of that process the data file now includes detailed quantitative information on up to:

- 15 prescription medications (increased from 11)
- 12 over-the-counter medications (increased from 10)
- 12 alternative medications (increased from 8)

Uncoded text data about additional medications still exists for a few cases reporting more than 15 prescription medications and/or more than 12 over-the-counter or alternative medications. The variable B4XXM has been modified to include multiple categories indicating the type of medication for which there is additional uncoded text data.

- b. Medical History: Current Health Practices: Exercise (p. 14) – “.... do you engage in regular exercise, or activity, of any type for 20 minutes or more at least 3 times/week?” (Q25)

Respondents self-define activities that fall into one of the three categories specified. They often report more than 3 types of exercise or activity. In this section a supplemental “Additional Activities” form was created that allows staff to record information about up to 10 additional forms of regular exercise/activities (B4H25a-j). Thus, the data about additional exercise are entered in a format paralleling that of the items appearing in the booklet.

Note: In recent years health professionals have been promoting a broader understanding of exercise as movement of any type for at least 10 minutes. Thus, many respondents report housework, gardening, and lifting & carrying on the job in response to this question. As a result 76% of participants report engaging in regular activity. Analysts interested in more traditional perceptions of exercise may want to review those responses prior to analysis.

4) Respondent Can't Recall Date:

Respondents are not always able to recall when an event occurred, but are able to provide verbal references (e.g. summer of 89 or in the mid 90's). In most instances, the appropriate missing value code was entered and text about the event was entered into marginal comments. At data cleaning the reference text was reviewed and the “missing” data was recoded to a valid response as appropriate.

Recodes/Additional Codes

1. There are a few questions for which enough respondents gave responses that were sufficiently ambiguous that a valid code could not be assigned or a specific number entered. As it would be inappropriate to assign the “Don’t Know” Missing Value code, the code “96” is utilized. The meaning of this code changes according to the item or set of items where it is specified.

- a) Medical History: Immune Function: Immunizations (Q11a-d) – “Have you had....?, If Yes, how old were you?” While many respondents report the age at which they had an illness or were immunized, many are only able to say something like ‘when I was a child’ or ‘before started school’ or something other phrase indicating it occurred during childhood. In such instances, the age is coded as “96”
- b) Medical History: Other Events (p.24) (Q57) – respondents report about long and short term ongoing events/experiences (family member going through major ongoing health event, ongoing preparations for a wedding etc.) as well as acute events. If the event is “ongoing” a “96” is entered for the month, along with the year the medical history was completed for the year.
- c) Medication Chart: Some medications have two or more active ingredients having different dosages. Others contain a single active ingredient, but are taken in different quantities throughout the day or on alternate days. In these instances, the dosage is recorded as “99999996”, “Other” is selected as the unit, and the dosage information is recorded in an open-ended format.

2. Medical History: Immune Function: Allergies (p. 8): “Do you have any allergies that have been diagnosed by a doctor or allergist?” (Q10). Many respondents report allergies that have not been diagnosed by a physician. Staff thus record information about all allergies reported and also note whether or not the allergy had been diagnosed by a physician. The codes for Q10 (B4BH10) were modified from Yes/No to 4 codes reflecting whether the allergies reported were diagnosed and/or undiagnosed.

3. Medical History Current Health Practices: Diet and Exercise: “On an average Day how many...

- glasses of water do you drink? (Q19)
- sugared beverages do you drink? (Q20)
- servings of fruits and vegetables do you eat? (Q21)
- servings of whole grain do you eat? (Q22)

The response options listed in the booklet (p.12-13) for these questions do not allow for respondents who consume the indicated item less than once per day. To accommodate this omission, a fifth category “Other – Less than once a day” is added to the response options at data entry.

4. Bone Questionnaire:

- a. Page 4, Q22 “When did you first notice irregularity in your menstrual cycle length (cycle length variability 7 days or more)? Women are not always able to recall exactly when they first noticed irregularity in their menstrual cycle and often provide

an open ended response that cannot be collapsed into the Months and Years format. The following codes are used at variable B4O22Y to represent common responses:

- 94 = Never Irregular
- 95 = Always Irregular
- 96 = Surgery No Irregularity

- b. Page 5, Q 26: “Do you have any rods, plates, or screws, or pins in your bones or joints?” Data cleaning revealed that the response “Elsewhere” had been excluded from Item 26. Since responses to this item appear in the data as a series of dichotomous Y/N variables, 2 existing variables were modified and a new variable was added to this series as follows:

B4O26D: relabeled to “Rod/plate location: elsewhere”

B4O26E: relabeled to “Pin location: hands or feet”

B4O26F: NEW variable labeled “Pin location: elsewhere”

5. Cholesterol Panel: LDL cholesterol levels are estimated using the Friedewald formula from direct measurements of total cholesterol (TC), Triglycerides (TG), and HDL cholesterol as follows:

$$\text{LDL} = \text{TC} - \text{HDL} - \text{TG}/5$$

This equation is unreliable if Triglycerides exceed 350 mg/dL. Meriter, the reference lab running the cholesterol panel assay, will not compute the LDL if the Triglyceride level is greater than 400 mg/dl. In those instances MIDUS computes the LDL value by substituting 400 mg/dL for the reported TG level.

6. Sometimes lab values are reported as “>” or “<” some value. When this occurs, the reported value is replaced with a value ‘one unit’ below the minimal or maximal detectable score. For example:

If the lower limit was <1.0, then we could change all of those scores to .9

If the highest possible score was 120 for a particular test, and was not normally reported with decimal values, then all the >120 would be converted to 121.

Extremely high and low values, therefore, are curtailed. Consequently the variance on the tails of the distribution is truncated. This is only problematic if a high percent of values fall in this category.

Missing Values

1. Throughout all the data files Missing Values appear as follows (see MIDUS Naming and Coding Conventions for details):

7, 97 etc. = Don’t Know

8, 98 etc. = Missing or Refused

9, 99 etc. = Respondent self-defined question as “INAPP”

2. Physical Exam (Short Version): “What is the tallest you’ve been measured in your life?” (Q2d). This question was added to this protocol in October/November 2004, thus this item is INAPP for respondents who completed P4 before the question was added.

3. Bone Questionnaire: Questions (Q23-27d) regarding metal in the body were added to the questionnaire in June 2007 when a Whole Body scan was added to the Bone Density protocol. Due to site specific differences, this change was implemented in June 2007 at Site 2, in August 2007 at Site 1, and in February 2008 at Site 3. These items are missing or INAPP for respondents who completed P4 before this change was implemented.

4. Biomarker assay result data (variable name begins B4B...) may be INAPP or MISSING for the following reasons:

- a) The assay could not be performed because a partial blood, urine or saliva sample was partial or the sample was not collected for some reason. This can be determined by looking at the B4ZBLOOD, B4ZURINE and/or B4ZSALIV.
- b) If the final evening void before bed, or the first void on rising are missed, or a significant portion of the urine sample were missed for some other reason the 12 hour catecholamine values (Norepinephrine, Epinephrine, Dopamine) were not computed.
- c) Saliva Duplicates: Initially, each saliva sample was assayed in duplicate. In January 2007 we stopped running assays in duplicate after it was determined that that the cortisol assay provided high quality, reliable results. Thus, the first 621 cases have saliva cortisol data for 2 samples, while the remaining cases have data for a single sample. An “average” value was created for all cases to provide a single consistent variable for each sampling time that could be used in analysis.

Data Inconsistencies

1) Pittsburgh Sleep Questionnaire: “If you have a roommate or bed partner, ask him or her how often during the past month you have had...Other restlessness while you sleep?” (Q10e). If a response other than “Not during the past month” was circled, then a response should have been recorded on the “Please describe:” line. Sometimes it appears as if the respondent, rather than the respondent’s Bed Partner/Roommate, provided this information. For example, the following responses raised this question for data staff.

“unable to remain sleeping for long periods of time”

“Sore shoulders - can't seem to find comfortable position”

“overactive bladder, tossing and turning - having to sleep w/husband”

Staff preparing data for entry flag responses to this item that appear problematic by instructing data entry staff to enter ‘7- Don’t Know’ at B4S10E. During data cleaning, decisions about these cases are made on a case by case basis. In some instances B4S10E is recoded from “7” back to the original response circled by the respondent, in others B4S10E will be recoded to “1 –Not during the past month’. Coding is in progress on responses recorded at “Please Describe”

2) Physical Exam Vision Assessment (Item 3b): Visual acuity (corrected and uncorrected) is measured from a single distance. Consequently, if an individual wears glasses to correct for distance vision, their corrected vision could be worse than their uncorrected vision. If an

individual has no vision or light perception these variables are coded as “INAPP”. For details about protocol for measuring visual acuity see the Physical Exam documentation.

Other Specify & Open-Ended Items

All of the paper and pencil instruments include places to record responses to open-ended questions or requests to ‘Please Specify’ or ‘Please Describe’ when the respondent gives a response in the category “Other”. This information was entered but is not included in the public data. Coding is in progress or pending for these items.