MIDUS 3 BIOMARKER PROJECT (P4) DATA FILE NOTES
(March 2023)

This document highlights aspects of the MIDUS 3 Biomarker data that analysts should be aware of prior to working with the data. In particular, Biomarker specific variable naming conventions and administrative variables are described along with issues that arise across multiple sections of the data (e.g. complex items that are represented using multiple variables, accommodating instances when respondents offer more information than our instruments provide room for). General issues and information that apply across multiple types of data are listed first, followed by issues that are unique to specific instruments.

This document provides general information. For more detailed information about specific sections of the data, see the documentation files for the following data types. The files are described in more detail in the README FIRST file:

- Blood, Urine and Saliva Assay Data
- Musculoskeletal Health and Function Data
- Medication Data
- Physical Exam Data
- Psychophysiology Protocol
- Psychosocial Constructs and Composite Variables
- Sleep Data (includes Actigraphy data)
- Gait Data
- Ankle Brachial Index Data

Also included among the documentation files are:

- Biomarker Project Summary - an overview of the Biomarker Project
- Documentation for PsychoSocial Constructs and Composite variables
- Documentation for Coded Text Responses
- Self-Administered Questionnaire (SAQ) – PDF of the SAQ booklet
- Medical History Interview – PDF of the CAPI (Computer Assisted Personal Interview).

Details about the data and links to the related documentation are also available via the MIDUS Colectica Portal (http://midus.colectica.org/) which houses interactive codebooks for all the publicly available MIDUS projects. The Portal includes search and explore functions, links to documentation, and a custom download function. A link to the portal is also available on the MIDUS website (http://midus.wisc.edu/) under QuickLinks.

Many of the issues described below apply to prior waves of the Biomarker project (MIDUS 2, MIDUS Refresher 1) as well as the MIDUS 3 data, differences are noted where appropriate.

Variable Names

MIDUS variable naming and coding conventions (See “Naming and Coding Conventions” in the general MIDUS documentation included with the Survey project documentation) specifies that for the M3 data the *third* character of the variable name be a letter that identifies the data type, collection method, 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 below:

Z = Biomarker (P4) Administrative Variables
S = Pittsburgh Sleep Questionnaire (PSQ)
Q = Self-Administered Questionnaire (SAQ)
H = Medical History
P = Physical Exam (Short & Long versions)
X = Medication Chart
B = Biomarker Assays (Blood, Urine, Saliva)
A = Daily Sleep Diary (used with the Actiwatch)
W = Actiwatch Watch data
D = Bone Densitometry, Body Composition & Muscle Function
V = Psychophysiology
G = Gait
I = Body Impedance
C = Ankle Brachial Index

Notes :

1. At MIDUS 2 questions about bone health and behaviors or experiences affecting bone health were included in a stand-alone instrument. These items were integrated into the Medical History at the Refresher and carried forward to M3, thus the designation 'O = Bone Questionnaire' is no longer used.

2. At M2 and MR1, the PSQ (Pittsburgh Sleep Questionnaire) and SAQ (Self-Administered Questionnaire) were administered as two separate questionnaires. At M3, some PSQ items were dropped and this abbreviated version was integrated into the first two pages of the SAQ. However, the question numbers and corresponding variable names were NOT integrated. The designation P and Q are continue used for PSQ and SAQ variables, respectively.

Administrative Variables

The data file includes the following administrative variables from the M3 Survey:

- M2ID – MIDUS Core sample ID included in all public files for the Core sample to facilitate linkage to data from other MIDUS Core projects
- SAMPLMAJ – the major sample identifier (i.e. Main sample or Milwaukee).
- C1PRSEX – participants gender
- C1PRAGE – participants age at the time of the M3 Survey Phone Interview

The file also includes the following administrative variables specific to the M3 Biomarker data:

- C4ZSITE – Biomarker data collection site
- C4ZCOMPM, C4ZCOMPY - Month and Year the Clinical Research Unit (CRU) visit was completed
- Time Zone variables: 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
    - C4ZRZONE - indicates the Time Zone the Respondent lived in at the time of the clinic visit
    - C4ZSZONE - time zone where the data collection Site is located.
- Lag Variables: indicating the time elapsed between the Biomarker (P4) Completion date and completion of other MIDUS 3 projects
    - C4ZC1PLG - Project 1 (Survey) Phone Interview

    - C4ZC1SLG - Project 1 (Survey) SAQ
    - C4ZC3CLG - Project 3 Cognitive Battery lag
- C4ZAGE - Respondent age when the Biomarker data was collected
- Tissue Sample Collection: variables indicating whether complete tissue samples were obtained. These variables appear with the assay data (C4B variables) immediately in front of the data related to a given tissue sample type
    - C4ZBLOOD – Blood sample completeness
    - C4ZURINE – Urine sample completeness
    - C4ZSALIV – Saliva sample completeness

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. In addition, one or more filter variables were created for use with individual data types (e.g. assay data, bone and muscle function, psychophysiology etc.) Details about filter variables are provided in the documentation for a given data type.

Date and Time Variables

Date and time data often have proprietary or conflicting formatting characteristics that can create problems when moving between data file types (i.e. SPSS to Excel). Thus, the MIDUS Naming and Coding Conventions require that dates and times be converted to formats that allow them to be read by a wider array of software programs. In response to this requirement, dates and times are included in the M3 Biomarker data file as follows:

1. Date – consistent with the M2 and MR1 data, dates are presented as separate month and year variables
2. Time – in the MIDUS 2 Biomarker data times were reported in a 12-hour clock SPSS time format. *That convention was modified at the Refresher* such that all time variables were converted to a 24-hour clock (i.e. day runs from midnight to midnight and is divided into 24 hours, indicating the number of hours since midnight, from 0 to 23). The time variables are also formatted as 4 digit SPSS Restricted Numeric which allows leading zero's to be displayed. Thus, 1:00 a.m. appears in the data as '0100', while 10:00 p.m. as '2200'. Time variables in this format are found in:
    a. The Pittsburgh Sleep Questionnaire (PSQ) data
    b. Urine Collection Times (included with the Urine Assay data)
    c. Daily Sleep Diary and Actiwatch data
    d. Psychophysiology Flowsheet data

Added Variables

The following describes instances where multiple variables may be used to report more complex items.

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
    i. "Please estimate your daily calcium intake" (Q37).

The interviewer asks the respondent how many servings Milk (Q37a), Yogurt (Q37b), and Cheese (Q37c) 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)? (Q39a-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:

C4H37_F or C4H39_F – frequency
C4H37_T or C4H39_T – time frame categories

b. Medical History – Current Exercise. "….do you engage in regular exercise, or activity, of any type for 20 minutes or more at least 3 times/week?" Q73a-g. If the respondent says "Yes", the interviewer obtains information about the type of activity including the frequency with which the activity is performed.

i. 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:

C4H73_FD – how many times per day
C4H73_FW – how many days per week

ii. Sometimes respondents indicate that certain activities are seasonal (e.g. gardening, snow skiing, bicycling). To capture this information, we also assess 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 C4H73_S where the 6th character is a-g as appropriate.

Note: Health professionals have been promoting a broader understanding of exercise as movement of any type for at least 10 minutes at a time. Thus, many respondents report housework, gardening, and lifting & carrying on the job in response to this question. As a result, 73% of MIDUS 3 participants reported engaging in regular activity. Analysts interested in more traditional perceptions of exercise may want to review those responses prior to analysis.

c. 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:

C4X_DD_ – dosage
C4X_DDU_ – dosage units

C4X_F_ - frequency
C4X_FU_ – time frame categories

C4X_T_ - how long
C4X_TU_ – time frame categories

2) Additional Others, Events and Medications:

a) 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.

Count variables, and related Yes/No variables are created for the following sections:

- Symptoms & Conditions
- Family Medical History
- Other Major Events

Details about creation of new variables can be found in -

- Documentation of Psychosocial Constructs and Composite Variables
- Documentation for Coded Text Responses

b) The number of medications that a respondent takes, or the number of medication allergies reported, may exceed the space provided on the Medication Chart. Procedures have been developed to maximize inclusion of relevant data. Details can be found in the Documentation for Medication Data.

3) 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

During data collection and/or data cleaning we found instances where additional codes needed to be added to existing variables or new variables needed to be created to accurately represent respondent reports. The following describes these instances and how they were addressed.

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: Other Events – 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.

b) 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. Details about how these instances are handled can be found in the Medication Data documentation.

c) Sleep Diary: Nights where the respondent indicated they did not sleep (often due to an unusual work schedule) were coded with "96"s for most of the evening/bedtime questions.

2. 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 falls 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, 997 etc. = DON'T KNOW
8, 98, 998 etc. = MISSING
9, 99, 999 etc. = INAPP

2. Biomarker assay result data (variable name begins C4B…) 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 C4ZBLOOD, C4ZURINE and/or C4ZSALIV.
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.

Data Inconsistencies

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, as well as the Medical History CAPI interview, 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 the text responses are not included in the public data. Coding has been completed for most of these items. Coded text data are included in the public dataset in one of the following ways.

1) New categorical variables are created: Participants are often asked to report on specific events or describe situations in an open-ended format. When this text data is coded, a new categorical variable is created so that the data can be used in analyses. To distinguish the new variable from the source variable a single character is added to the end of the source variable name. For example, in the Medical History:
    a. Simple Falls section includes text variables describing 'Which' bone was broken (C4H3A1) and the 'Circumstances' of the fall in which the bone was broken (C4H3A2). The new categorical variables corresponding to these variables are named C4H3A1W and C4H3A2C, respectively.
    b. Current Health Screenings: in this section participant are asked if they have ever had cholesterol checks etc. and, if yes, asked to report the results. The results have been coded and the new categorical variables corresponding to the source variables have a 'C' at the end. For example, coded cholesterol results are reported as C4H25ARC.
2) Extant Codes for a related variable are expanded: this is done when specify "Other" options were included in a predefined set of categories. The text responses specified were reviewed and new categorical codes created if appropriate.
    a. For example, in the Medical History participants are asked how often they go to the dentist and several options, including an "Other Specify". Review of the specify revealed other categories of frequency (i.e. 3 times a year, 4 times a year etc.) that were incorporated into the existing categories. See C4H23A.

Details about coded text variables are provided in the documentation file 'M3_P4 Documentation for Coded Text Responses'.