MIDUS 1 Project 2:

National *S*tudy of *D*aily *E*xperiences

Wave 1: Read Me File

MIDUS 1 Project 2
*N*ational *S*tudy of *D*aily *E*xperiences (NSDE)

Wave 1: Read Me File

**NOTE: Please read through this document carefully
prior to using the data and documentation.**

The purpose of this memo is to provide basic information about the MIDUS data and documentation files that are being made publicly available via ICPSR. Specifically, this memo pertains to the Time 1 data collection of daily dairy that was conducted as a component of MIDUS 1.

A. What Data File Is Available?

The M1P2 2 dataset:
Daily Data (Person-Day Dataset):   M1_P2_DAILY DAIRY_N=1499_20141103

(For details about Project 2, see M1_P2_ UserGuide).

B. What is the Structure of the MIDUS 2, Project 2 Dataset?

The dataset is a 'stacked' or 'person-day' dataset (see below) comprised of data for 1,488 days on 329 study variables from 1,499 participants. The dataset combines random digit dialed (RDD) participants, twins, siblings, and Milwaukee respondents. The variable M2ID identifies each individual respondents and can be used to merge M1P2 data with other MIDUS datasets. The variable SAMPLMAJ, from Project 1, identifies from which of the subsamples a respondent derives.

Variables have been named according to the MIDUS conventions (see M2_Variable Naming and Coding Conventions.doc). All variables include labels to aid interpretation. Value labels have been applied where appropriate. Details about Project 1 variables can be found in the following documents: M1_P2_ UserGuide

The Person-Day Dataset

The NSDE daily data file (M1_P2_DAILY DAIRY_N=1499_20141103) is structured as a "person-day" dataset such that each row of data corresponds to a single day for a given individual. Data regarding days are 'nested' within individuals over time. Micro-level information (data collected on a given day) comes from a macro-unit, in this case a single individual. In contrast to typical multivariate datasets, where each row of values corresponds to one single individual and that individual's values for some set of variables, each row of a person-day dataset corresponds to an individual's values for some set of variables on that day. Figures 1 and 2 depict the traditional multivariate and

person-day datasets, respectively. Both represent data from 5 participants, actually, the same 5 participants (note the *same* IDs). The figures below depict data from two components of a study. First, Figure 1 shows data about participants' age, gender, and depression scores (CESD, The Center for Epidemiologic Studies Depression Scale) taken at the baseline assessment (i.e., MIDUS Project 1). These data can be seen as characteristics that vary across individuals in the study (traditional individual differences research).

Figure 1. Traditional multivariate dataset.

ID	Age	Gender	CESD
101	45	M	4
102	86	F	7
103	37	F	11
104	72	M	8
105	66	M	18

Figure 2 depicts data collected from the same participants assessed on three consecutive days. Participants were asked whether they experienced a stressor on that day (variable 'Any Stress'; Yes = 1, No = 2), as well their negative affect (Neg. Affect) over the past 24 hours (sum of NA items). The day of assessment is indicated by the variable 'DAY'.

Figure 2. Person-day dataset (3 days of assessment).

ID	Day	Any Stress	Neg. Affect
101	1	1	7
101	2	2	5
101	3	1	7
102	1	1	11
102	2	1	15
102	3	2	10
103	1	1	9
103	2	1	9
103	3	1	8
104	1	2	5
104	2	2	6
104	3	1	9
105	1	2	5
105	2	2	5
105	3	2	6

Figure 2 shows participant 101 experienced a stressful event on the first and third days that they were assessed, whereas participant 105 did not report experiencing a stressor on any of the days. Figure 2 also shows that across all 5 participants, their negative affect varies from one day to the next. Figure 2 displays quite clearly the concept of the

'nested' data structure. Notice that there are multiple observations for each individual, and these observations are organized by ID **and** DAY of assessment.

Linking Traditional Multivariate and Person-Day Datasets

It may be of interest to link the data from the baseline assessment to the daily assessments to answer certain research questions (e.g., Are age and depressive symptomatology related to the likelihood of experiencing a stressor on any given day?). To answer these questions, we would need to merge the two datasets into one. Let us assume that the data from the traditional multivariate dataset (baseline variables, including age, gender, and CESD scores) are in a dataset called 'baseline', and the data from the daily component of the study are contained in a dataset called 'daily'. SAS and SPSS codes to merge these two datasets into one are shown below:

SAS Code

```
Data all; *Create/name a dataset;
Merge baseline daily; *Tells SAS that the new dataset 'all' is going to be a combination of
the 'baseline' and 'daily' datasets;
By ID;*Tells SAS that 'ID' is the relevant variable by which the data should be linked;
Run;
```

SPSS Code

```
MATCH FILES /FILE=*
  /TABLE='C:\NSDE\aggr2.sav'
  /BY ID
EXECUTE.
```

When these commands are executed, the resulting dataset should look similar to Figure 3.

Figure 3. Combined multivariate (baseline) and person-day (daily) datasets.

ID	Day	Any Stress	Neg. Affect	Age	Gender	CESD
101	1	1	7	45	M	4
101	2	2	5	45	M	4
101	3	1	7	45	M	4
102	1	1	11	86	F	7
102	2	1	15	86	F	7
102	3	2	10	86	F	7
103	1	1	9	37	F	11
103	2	1	9	37	F	11
103	3	1	8	37	F	11
104	1	2	5	72	M	8
104	2	2	6	72	M	8

104	3	1	9	72	M	8
105	1	2	5	66	M	18
105	2	2	5	66	M	18
105	3	2	6	66	M	18

Figure 3 shows that age, gender, and CESD variables from the 'baseline' dataset have been appended to the 'daily' dataset. Furthermore, notice that each participant's age, gender, and CESD score have been appended to EACH row. Because age, gender, and CESD only were assessed once, each variable has a single value for each participant, and that value is included at each observation ('DAY') for each participant ('ID'). Although age, gender, and CESD only were assessed once, the fact that they appear at each observation after merging the datasets is normal. Because these measures were not time-varying (i.e., collected more than once), in the context of this example, we would not expect them to have differing values across DAYs for any given participant. If a value did vary from day to day (say CESD for participant 103 was 11, 14, 11, for days 1 to 3, respectively), something went wrong in the data merging process.

C. <u>Constructed Variables and Scales</u>

Constructed variables and scales can be found at the end of the data file (see M1_P2_UserGuide.pdf).

E. <u>New and Revised Materials</u>
The following changes have been made to the dataset since its original public release:
- Administrative variable changes:
  - o A2DM2ID was renamed M2ID.
  - o A2DSAMP was replaced with SAMPLMAJ.
  - o M2FAMNUM, A1PAGE_M2, and B1PGENDER were added.
- 15 rows of data were removed because they contained no data. One case (M2ID=17342) was completely removed for lack of data. The M2IDs for the remaining affected cases are: 10235 10430 10923 11159 11553 11692 11692 12054 12451 14545 16686 17524 18155 18196
- For longitudinal consistency, variable labels from the MIDUS 2 daily experiences data were applied to the MIDUS 1 data.
- The original survey instrument and dataset used a coding system for non-response that is different from MIDUS conventions. In the original data, 8=Don't Know and 9=Refused. All variables were recoded to fit the M2 conventions: 7=Don't Know, 8=Refused, 9=Innap. This dataset also recoded sysmis into 8s and 9s where appropriate, applied appropriate value labels, and designated them as missing values in SPSS. This dataset also recoded Yes/No variables from the original (1=Yes, 5=No) to M2 conventions (1=Yes, 2=No). One important point is because of these changes, the instrument is incongruent with the data (i.e. the instrument lists 1=Yes, 5=No, and has different non-response labels).
- Three variables were dropped from the data (A2DEDUCP A2DAGE A2DHHTOT) since about 1/3 of the cases had missing data for these variables. Complete data on these variables (education, age,

household income) can be found in the M1 Project 1 dataset and merged with the M1 Project 2 data (via the M2ID variable) if desired.