MIDUS Refresher
Project 2:

*N*ational *S*tudy of *D*aily *E*xperiences

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 Refresher Project 2 (Daily Stress) data and documentation files.

A. **What Data File Is Available?**

The NSDE 2 dataset:
Daily Data (Person-Day Dataset): MR_P2_DailyData_N782_20200702.sav

(For details about Project 2, see MR_P2_StudyDescription_20200706).

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

The dataset is a "stacked" or "person-day" dataset (see below) comprised of data for days on 604 study variables from 782 participants. The dataset contains random digit dialed (RDD) participants. The variable SAMPLMAJ, from Project 1, identifies from which of the subsamples a respondent derives.

The Person-Day Dataset

The NSDE daily data file (MR_P2_DailyData_N782_20200702.sav) 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 1 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 as 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. **What additional files are available?**

Survey instruments. The complete Daily Diary instrument is available:
> ***MR_P2_DailyDiaryInstrument_20160906***

Documentation. Constructed variables and scales can be found immediately following the cortisol variables at the end of the data file (see MR_P2_Scales_20200706 and MR_P2_Saliva_20200706).

D. **Other Important Information.**

ID Systems. A 5-digit respondent identification variable (MRID) has been created for MIDUS Refresher cases. This ID system is implemented to help maintain confidentiality of respondents and can be used throughout to merge and link the publicly available MIDUS Refresher datasets.

DDI Codebook. A DDI codebook (in both XML and PDF) is available for the MR Daily Diary project data on the MIDUS Portal (http://midus.colectica.org/). This Portal can also be used to access all the publicly-available MIDUS data and created customized data extracts using variables from across all MIDUS datasets.

2019 Updates. Variable labels were fixed for the following variables: RA2DA1010, RA2DA1011, RA2DA1012, RA2DA1013, RA2DA1014, RA2DA1015, RA2DA1016, RA2DF12, RA2DF13.

2020 Updates. Thirteen additional variables were added to this dataset that represent levels of salivary Alpha Amylase.