MTF Workshop –
Exploring and Analyzing Monitoring the Future Data: A Primer

July 23-25, 2018

ICPSR Summer Program
Ann Arbor, Michigan

Instructors: Pat Berglund, Deb Kloska, Austin McKitrick

The overall goal of this workshop is to train participants to access and analyze MTF cross-sectional and panel data. We will cover use of documentation tools and public release and restricted data available from the ICPSR/NAHDAP archives. In addition, we will discuss planning and data management prior to analysis and demonstrate a variety of analysis techniques. Descriptive and regression analyses of selected drug variables will utilize probability weights and complex sample design features such as stratification and clustering for proper variance estimation. SAS will be the only software used in the workshop, but we anticipate that users of other software can generalize to their statistical software of choice.

The format of the workshop will be lecture/demonstration of key topics in addition to supervised computer exercises done in the NAHDAP enclave using “workshop data sets” designed for this training. All instruction will be provided by experienced MTF analysts.

Instructor Bios

**Patricia A. Berglund** is a Senior Research Associate in the Survey Methodology Program at the Institute for Social Research. She has extensive experience in the use of computing systems for data management and complex sample survey data analysis. She works on research projects in youth substance abuse, adult mental health, and survey methodology using data from Army STARRS, Monitoring the Future, the National Comorbidity Surveys, World Mental Health Surveys, and other data sources. In addition, she is involved in development, implementation, and teaching of analysis courses and computer training programs at the Survey Research Center-Institute for Social Research. She also lectures in the SAS® Institute-Business Knowledge Series. Email: pberg@umich.edu

**Deborah D. Kloska** is a Senior Research Associate with the Youth and Social Issues Program at the Institute for Social Research. She has worked with Monitoring the Future data for more than 20 years, training students and assisting researchers in understanding and analyzing the MTF data. With a degree in Applied Statistics, her analytical interests focus on applying up-to-date knowledge of statistical methods and procedures in the area of substance use, especially as related to longitudinal data analysis and complex survey design. Email: ddkloska@umich.edu

**Austin W. McKitrick** is a Research Associate with the Monitoring the Future (MTF) project. He attended West Virginia University, graduating with degrees in Biology and Psychology. His data management experience includes an internship with ICPSR, along with managing the MTF data and documentation for the restricted-use panel file. Austin works closely with researchers who require access to the restricted-use panel data. Email: mtfdataaccess@umich.edu
Workshop Outline

Day 1 (July 23, 2018)

8:30-9am
Registration/Orientation given by ICPSR staff

9am-4pm with 1 hour break between Noon-1pm

Welcome to Workshop

Topics
1. Overview of MTF project, history of project, data sets available to public, examples of basic reports/detailed papers.

2. How to access the ICPSR site from internet browser, where documentation and other tools are located, how to use the “workshop data sets” and extract needed variables and set up analysis data set.

3. Planning for analysis project (what data and years, questionnaires, key topics, selection of variables, analysis goals)

4. Hands-on Computer Work
Have students access the enclave via web browser (using their own laptops or one supplied by ICPSR if needed), become familiar with the site and “workshop” data sets, how to use tools such as questionnaires and codebooks for planning project and analyzing data.
Day 2 (July 24, 2018)

9am-4:00pm with lunch break noon-1pm

Topics

1. Data management and analysis of 12th grade public-use and restricted-use merged data, 2010-2012. Analysis of 12th grade data using both public release and restricted form. Show how to merge data sets together and do descriptive analysis of one drug using weights and complex sample design variance estimation (PROC SURVEYMEANS/PROC SURVEYFREQ). The demonstration will also include a typical figure of drug trends over time, 2010-2012. (9-11am)

2. Demonstration of regression analyses using the same data set as used for descriptive analysis. These analyses will use SAS PROC SURVEYREG/SURVEYLOGISTIC, depending on the type of outcome variable. (11am-noon)

3. Computer Exercises
   
   **Note: all demonstrations and computer work will be done using SAS 9.4 (available in the enclave). Students will be encouraged to generalize exercises to software of their choice but we cannot support multiple software for the exercises in this workshop.**

   Using “workshop data sets” on VDE, have students do a pre-defined analysis of 12th grade data similar to that presented earlier. We will provide starter code for this exercise to help people get started with SAS. All three instructors will be available to help students complete this exercise, and we will go through results together during lab session. (1pm-4pm)
Day 3 (July 25, 2018)

9am-2:30pm with lunch break from 11:30am-12:30pm

Topics

1. Data management for analysis of panel data. Selection of variables and samples for analysis, data structures needed for analysis, selection of correct weights, and other important features of MTF panel data. (9-9:45am)

2. Selection of appropriate analysis technique for follow-up data. How to handle repeated measures and correlation within respondents, and use of attrition weights. Discussion of complex sample design features. Use of PROC MIXED/PROC GENMOD and other SAS procedures for longitudinal data analysis. (9:45-10:30)

3. Demonstration of analysis of selected years of panel data, typical tables/figures of selected drug(s) use over time (descriptive) and selected regression analyses appropriate for longitudinal data analysis. (Exact variables needed will be determined later). (10:30-11:30)

4. Computer Exercises
   Using “workshop panel data sets” on VDE, have students do a pre-defined analysis of panel data such as marijuana use or binge drinking at age 18 and Follow-up 1-3. As in day 2, we will provide starter code and help in doing this exercise and go through results together during lab. (12:30-2pm)

5. Wrap-Up
   Summary of workshop, what might we have covered if we had more time, what other topics are of interest to participants?

   Discussion of other software that can perform correct analysis (Mplus, Latent Gold, Stata, etc.) handle data of this type (not used in workshop).

   General questions/answers. (2-2:30pm)

Instructors will be available until 3:00 pm for those with questions about how to apply for use of the restricted-use data (if needed).