

**Society for Prevention Research
18th Annual Meeting**

Pre-conference #4

**June 1, 2010
8:30 am -- 5:00 pm
Grand Hyatt Denver
Denver, Colorado**

Everything You Ever Wanted to Know About Secondary Data Analysis*

*But Were Afraid to Ask

Note: This workshop is sponsored by the ECPN Steering Committee

Chair

Richard Moser, Ph.D.
Kathy Etz, Ph.D.

Background:

This workshop continues the combined efforts of the SPR Database Taskforce and the Early Career Preventionist Network to promote re-analyses of existing datasets.

Numerous federal, state, and grant-related health data are available that are currently underutilized by research scientists, such as the National Health Interview Survey, the Health Information National Trends Survey, the California Health Interview Survey, and the Behavioral Risk Factor Surveillance System. In addition, many investigator initiated datasets are not mined to their full potential. These databases capture national estimates of health-related behavior, knowledge, and attitudes and/or provide information on the mechanisms and processes that account for behavioral or health outcomes as well as outcomes related to interventions. Secondary analyses of existing databases may serve as an economical alternative to expensive and time-consuming new data collection projects and may serve as a valuable tool to test complex statistical models, perform meta-analyses and develop and test empirical hypotheses.

This workshop, through lecture, applied demonstrations, and group discussion will inform participants about Internet-based and other health databases available for analysis and describe the utility and content of these databases, as well as explain how to access and analyze the data. Presentations will highlight analytic issues involved with utilizing these data, describe statistical software available to perform analyses, demonstrate the types of analyses that can be completed, and describe how two different datasets can be combined for analysis. In addition,

an early-career scientist will discuss her experience in doing secondary data analysis and how this has enhanced her career. The last part of the workshop will include a panel of people from both within and outside of the Federal government to describe and discuss relevant NIH funding mechanisms and discuss the process of obtaining funding. Audience members will be encouraged to participate in this discussion.

Learning Objectives:

1. Participants will be aware of Internet-based and other health survey databases available for analysis, the content and utility of available databases, and learn how to access the data.
2. Participants will learn about analytic issues involved with utilizing data that employ complex sampling designs and become familiar with the statistical software available to perform analyses of these data.
3. Participants will learn the value of analyzing these data through applied examples
4. Participants will learn how different datasets can be combined and analyzed.
5. Participants will become familiar with available NIH funding mechanisms to support secondary analyses.

Target Audience:

Researchers, particularly those in early career stages, who are interested in learning more about accessing and analyzing health data. It would be helpful if attendees had experience in data analysis and basic statistics.

Outline:

- I. Update on Database Taskforce and partnership with ECPN activities (Ty Ridenour)
- II. Background on secondary data analysis: Learning from an early-career researcher (Presenter: Bethany Bray)
 - a. Experiences with secondary data
 - b. Opportunities for collaboration
 - c. Role that secondary analysis played in her career
 - d. Ethical issues and tools for working with secondary data
- III. Accessing secondary data (Presenters: Richard Moser, Felicia LeClere)
 - a. Overview of existing databases
 - i. US/Canadian national and US state-level datasets
 - ii. Content
 - iii. Strengths/limitations
 - b. Inter-University Consortium for Political and Social Research (ICPSR)
 - i. Datasets available
 - ii. Using the download system

- IV. Analyzing secondary data (Presenters fore a-d: Richard Moser, Judith Gelertner, Felicia LeClere)
 - a. Basics
 - i. Linking to other data systems
 - 1. Linking keys and software systems
 - 2. Guidelines for merging
 - ii. Structuring data correctly
 - iii. Analytic issues
 - b. Survey data
 - i. Weights
 - 1. Final sample
 - 2. Replicate weights
 - ii. Stratification
 - iii. Clustering
 - iv. Variance estimation
 - c. Statistical software
 - i. Strengths/limitations of each
 - ii. How do you decide which one to use?
 - iii. Differences in results when using appropriate/inappropriate software
 - d. Combining datasets
 - i. Finding common variables across different datasets
 - ii. Linking to other data systems
 - 1. Linking keys
 - 2. Guidelines for merging
 - iii. Structuring data correctly
 - iv. Analytic issues of combining datasets
 - v. Applied examples
 - e. Applied examples (Presenters: Lila Finney Rutten, Kerry Keyes)
 - i. Using survey data
 - ii. Using etiology data
- V. Panel to discuss NIH funding for secondary data analysis (Presenters: Kathy Etz, Erik Augustson, Amy Yaroch, Marsha Lopez)
 - a. Mechanisms
 - b. Considerations
 - c. Group discussion

Presenters

Kathy Etz, Ph.D.
Program Director
National Institute on Drug Abuse
301-402-1749 (w)
301-443-2636 (fax)
ketz@nida.nih.gov

Qualifications:

Kathy directs a program of research that includes secondary data analysis approaches and has written funding announcements to support these types of analyses. Kathy is currently the project officer on the ICPSR NIDA Drug Abuse and HIV data archive.

Richard P. Moser, Ph.D.
Research Psychologist
National Cancer Institute
301-496-0273 (w)
301-435-7547 (fax)
moserr@mail.nih.gov

Qualifications:

Richard works at the National Cancer Institute and has many years of experience performing statistical analyses using survey data that employ a complex sampling design and has numerous related publications. In addition, he is part of the team that manages one of these surveys, the Health Information National Trends Survey (HINTS) so he understands all of the issues in regards to accessing and analyzing survey data. He has also taught other seminars and given training classes to NIH staff on this topic.

Bethany C. Bray, Ph.D.
Assistant Professor, Department of Psychology
Virginia Tech University
540-231-4287 (w)
540-231-3652 (fax)
bcbray@vt.edu

Qualifications:

Bethany is an early-career researcher interested in developing methods for prevention and treatment research. Almost all of her work has been conducted using secondary data, both public and private, and so can discuss the impact that this type of research has had on her career.

Felicia B. LeClere, Ph.D.
Director, Data Sharing for Demographic Research
ICPSR
University of Michigan
734-615-7333 (w)

734-647-8700 (fax)

Qualifications:

Felicia is the Director of the new National Institute on Drug Abuse (NIDA) data archive called National Drug Abuse and HIV Data Program at the ICPSR and is very knowledgeable about datasets that are available for analysis. She is familiar with many of the issues that arise when one performs secondary data analysis.

Ty A. Ridenour, Ph.D., M.P.E.
University of Pittsburgh
Center for Education and Drug Abuse Research
412-383-5245 (w)
412-624-1929 (fax)
tar27@pitt.edu

Qualifications:

Ty currently is the Chair of both the Early Career Preventionist Network and the SPR Database Taskforce. Many of his peer-reviewed research publications and NIH grants include re-analyses of existing datasets as well as innovations in statistical techniques. His research has been continuously funded by NIH since 1996.

Erik Augustson, Ph.D., M.P.H.
Health Science Administrator/Behavioral Scientist
National Cancer Institute

Qualifications:

Erik has expertise on funding mechanisms available for secondary data analysis and consults with extramural researchers on this subject. In addition, Erik has more than 40 scientific publications—many that involved secondary data analysis of health survey data-- and has made more than 100 scientific presentations.

Judith Gelernter, Ph.D.
School of Computer Science
Carnegie Mellon University

Qualifications:

Judith's research concerns the application of artificial intelligence methods for knowledge discovery from secondary data and the design of easy-to-use tools that exploit that data effectively. She works in the Language Technology Institute of Carnegie Mellon's School of Computer Science.

Kerry Keyes

Qualifications:

Katherine has worked with a number of large scale publically available datasets (including the NESARC, NHSDUH, NCS, and NHANES) and is currently completing her dissertation work using the Monitoring the Future (MTF) surveys. MTF is a yearly cross-sectional survey conducted among a nationally representative sample of 8th, 10th, and 12th graders in the United States. In her dissertation work, Katherine is confronting a number of complex issues using these data, including three sources of clustering in the sample design, complicated sets of several hundred codebooks, and numerous analytic issues in generating valid inferences from a data of over 1 million adolescents who are clustered in space and time.

Amy Yaroch, Ph.D.
Executive Director
Center for Human Nutrition
Omaha, NE

Qualifications:

Amy has developed questions, analyzed data, and co-authored papers on various national-level secondary datasets, including the communication and health behavior-focused Health Information National Trends Survey (HINTS), the California Health Interview Survey (CHIS), and the HealthStyles and YouthStyles Surveys. Prior to her current position, Amy was a Program Director/Behavioral Scientist at the National Cancer Institute, Health Promotion Research Branch and so is very familiar with Federal funding mechanisms for secondary data analysis.

Lila Finney Rutten, Ph.D., M.P.H.
Behavioral Scientist
SAIC Inc./Frederick
National Cancer Institute

Qualifications:

Lila's current responsibilities including managing/coordinating the Health Information National Trends Survey (HINTS) and supporting research activities in the Behavioral Research Program at the National Cancer Institute. Lila has extensive experience analyzing secondary data and has successfully authored many scientific journal articles using secondary data.