MATHEMATICA
Policy Research

# PRESENTATIONS

## Joint Statistical Meetings

***"The Extraordinary Power of Statistics"***

July 30-August 4, 2016 • Chicago, IL

## Sunday, July 31

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶4:00–5:50 pm | CC-W176c | **Raking, Post-Stratification, and Calibration Methods** (Survey Research Methods Section) | **Nancy Clusen, Chair** |
| ▶6:00–8:00 pm | CC-Hall F1 West | **Poster Session**<br><br>**Poster Presentation:** Analysis of Longitudinal Multi-Sequence MRI in Multiple Sclerosis<br><br>**Abstract:** Multiple sclerosis (MS) is an inflammatory disease of the central nervous system that is characterized by brain and spinal cord lesions. The formation of MS lesions in the brain is a complex process involving inflammation, tissue damage, and tissue repair—all of which are visible on structural magnetic resonance imaging (MRI). In this poster, I will explore the data that arises from multi-sequence structural MRI from longitudinal studies in patients with MS. I will present two statistical problems with clinical relevance that arise from these data: (1) longitudinal MS lesion incidence detection and (2) the development of a biomarker for MS lesion repair. Longitudinal MS lesion incidence is modeled using a historical functional Cox regression model and the repair biomarker is the result of dimension reduction through principal component analysis. | Jonathan Gellar (co-author) |

## Monday, August 1

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶8:30–10:20 am | CC-W184d | **Statistical Methods for Spatial Epidemiology** (Section on Statistics in Epidemiology) | **Nicholas Beyler, Chair** |
| ▶2:00–3:50 pm | CC-W175a | **If You Measure, They Will Publish: State and Federal Data on Sexual Orientation and Gender Identity in Action** (Social Statistics Section) | **Diane Herz, Discussant** |

## Monday, August 1 *(continued)*

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶2:00–3:50 pm | CC-W186a | **Regularization and Prediction in Time Series and Longitudinal Models** (Section on Statistical Learning and Data Science)<br>**Paper Presentation:** Longitudinal Network Prediction with Applications to Network-Based Interventions<br>**Abstract:** We present a statistical method to generate a series of predicted networks based on the historical evolution of social relations in a given population. This allows an investigator to incorporate and model uncertainty, which is inherent in prediction. The method, which is based on a novel and flexible procedure to sample dynamic networks given a probability distribution on evolving network properties, permits the use of a broad class of approaches in order to model trends, seasonal variability, uncertainty, and changes in population compositions. The proposed method is demonstrated on two dynamic networks; the first represents the sponsor/co-sponsor relationships between senators indicated from bills introduced in the US Senate from 2003-2012. The second is temporal network data-sampled every 20 seconds-representing interactions between participants of the ACM Hypertext 2009 conference. The proposed method enables investigators to make rapid and informed decisions regarding network interventions such as mechanisms to either encourage information diffusion or minimize the impact of a contagion; we present a network-based intervention for the ACM conference for demonstration. | **Ravi Goyal (presenter)** |

## Tuesday, August 2

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶10:30 am–12:20 pm | CC-W184d | **Nonprobability/Web Sampling and Data Analysis** (Survey Research Methods Section)<br>**Paper Presentation:** Exploration of Methods for Blending Unconventional Samples with Traditional Probability Samples<br>**Abstract:** With increasing survey costs, reduced participation rates, and a greater need for timely and domain specific estimates to inform treatment and policy decisions, interest has grown in the use of data obtained from unconventional sampling to supplement or replace traditional survey programs. Given these unconventional data sources may result in unknown levels of bias in their estimates, we explore the use of composite model-based estimation methods to blend data from an unconventional sample with parallel data from a traditional probability-based sampling strategy. We present a number of approaches for blending data from two sources. We then study their utility in a large simulation study based on public use data from the National Health Interview Survey (NHIS). We show that use of the proposed methods enables one to determine the fitness of use of an unconventional data source and to leverage such data appropriately to fulfill program objectives. | **Jonathan Gellar (presenter)**<br>Hanzhi Zhou and Michael Sinclair (co-authors) |

## Tuesday, August 2 *(continued)*

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶2:00–3:50 pm | CC-W176a | **Journal Clubs as a Teaching Venue for Health Science Students** (Section on Teaching of Statistics in the Health Sciences)<br>**Paper Presentation:** The Professional Melting Pot: Statisticians, Data Scientists, and Health Researchers Talk Shop to Improve Public Well-Being<br>**Abstract:** Statisticians in the private sector do not often have the same opportunities as government or academic statisticians to discuss statistical topics in an informal setting like a journal club or working group. In 2011, four statisticians at Mathematica Policy Research formed a working group to discuss statistical methods from their personal research, project work, and journal articles they found interesting and relevant to their work. Today the experimental working group has developed into the Data Scientific and Statistical Methods Working Group (DSSMWG) composed of over thirty statisticians, data scientists, health researchers, and survey researchers. DSSMWG meets once a month to discuss statistical methods and data science tools that help guide project work and improve public well-being, which is the core mission of Mathematica. In this paper we discuss DSSMWG's structure (past, present, and future) and give some examples of "success stories" the group has had. We also provide advice to colleagues interested in starting a working group or journal club at their own company or institution and highlight challenges they may face along the way and how to deal with them. | **Nicholas Beyler (presenter)**<br>Fei Xing (co-author) |
| | CC-W181b | **Methods and Measures: Including People with Disabilities in Surveys** (Committee on Statistics and Disability)<br>**Paper Presentation:** When Every Dollar Counts: Comparing Reported Earnings of Social Security Disability Program Beneficiaries in Survey and Administrative Records<br>**Abstract:** Like all federal agencies, the Social Security Administration is faced with the challenge of making the best use of administrative and survey data for research and program operations, though limited information exists on how estimates from these two data sources compare on key outcomes, such as employment. This paper uses data from the National Beneficiary Survey that are linked to administrative records on earnings to compare survey and administrative reports of annual earnings for a sample of Social Security Disability Insurance and Supplemental Security Income beneficiaries. We find that both administrative and survey data indicate that less than 20 percent of beneficiaries work during the previous year, though reports of both employment and earnings are higher in administrative data. Our findings indicate that administrative data will identify a much larger sample of beneficiaries who worked in the previous year, which is potentially important for initiatives that target these populations, though both data sources provide potentially plausible estimates for employment outcomes that can be used to depict overall patterns within the population. | **Holly Matulewicz (presenter)**<br>David Wittenburg, Lindsay Glassman, Lisa Schwartz, and Others (co-authors) |

## Tuesday, August 2 *(continued)*

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶2:00–3:50 pm | CC-W181b | **Methods and Measures: Including People with Disabilities in Surveys** (Committee on Statistics and Disability)<br>**Paper Presentation:** A Feasibility Study of Recruiting and Maintaining a Web Panel of People with Disabilities<br>**Abstract:** Online panels are a cost-effective tool for collecting survey data. They also offer compelling advantages for researchers (e.g. maintaining contact with respondents) and respondents (e.g. completing surveys at their convenience). However, little is known about feasibility of collecting data from them using panel surveys. On one hand, older research documents a "digital divide" in which people with disabilities lack access to the Internet. On the other hand, recent technological developments have made computers more accessible to people with disabilities. We present results from a pilot study testing the feasibility of recruiting and retaining an online panel of people with disabilities. Vocational rehabilitation program applicants who previously participated in a phone survey were invited to participate in two follow up web surveys. Those without a means of completing an online survey were provided with a tablet and Internet access for the duration of the study. Response rates, demographic characteristics of respondents, and lessons learned from setting up the panel will be discussed. | **Jesse Chandler (presenter)** |
| | CC-W183c | **Adaptive/Innovative Survey Design and Survey Cost** (Survey Research Methods Section)<br>**Paper Presentation:** Implementing Adaptive Design on a Longitudinal Survey of At-Risk Youth: Empirical Evidence Based on a Deep-Dive Analysis<br>**Abstract:** There are a number of challenges associated with conducting longitudinal survey research with at-risk youths who are both highly-mobile and difficult to engage. This paper looks at an empirical example where innovative solutions were implemented to engage and retain our sample across survey waves for an impact evaluation study. In particular, we explore two essential ingredients in the mathematical framework of adaptive survey design: 1) alternative survey strategies for sample retention, and 2) statistical tools for dynamic monitoring of the former during the data collection process. Reliant on multiple available data sources including baseline information and paradata, we calculate quality indicators and construct prediction models to assess: 1) the representativeness of our data; 2) how our engagement and retention strategies influence the representativeness of the data; and 3) how we can adapt our data collection efforts to maximize the representativeness of our data. Our goals are to: 1) increase the knowledge base regarding the use of alternative measures of quality in survey practice;and 2) expand the research on innovative survey strategies such as the use of social media.<br>**Paper Presentation:** Selecting a Sample from a Changing Frame<br>**Abstract:** Probability samples are usually selected from a fixed sample frame, which is a close facsimile of the target population, and the probability of selection can be easily quantified. The sample design is necessarily more complicated when the sample frame is not fixed, but changes over time. In this paper, we describe an adaptive sample design for a target population that are members of a disabled population who had successful work experience. The highest priority in this design is to minimize the length of time between the interview date and the period of successful work. We describe a design that was intended to do this, while at the same time accommodating the changing frame. | **Jillian Stein (presenter)**<br>Hanzhi Zhou (co-author)<br>**Eric Grau (presenter)**<br>Others (co-authors) |
| | CC-W184a | **Missing Data, Imputation, and Calibration** (Government Statistics Section) | **Donsig Jang, Chair** |

## Wednesday, August 3

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶8:30–10:20 am | CC-W194b | **Recent Advances in Functional Data Analysis** (Section on Statistical Learning and Data Science)<br>**Paper Presentation:** Historical Functional Cox Regression, with an Application to Prediction of Multiple Sclerosis Lesions<br>**Abstract:** Historical functional regression was originally developed as a form of function-on-function linear regression: the responses and predictors are defined on a common time domain and the response can be explained by values of the predictor only up to the present time, leading to a bivariate coefficient function supported on a triangular region. Recently, coefficient functions of this type have formed the basis of a novel historical functional Cox model for time-to-event data. The historical Cox model relates the log hazard to not only the present values but also earlier values of a time-varying biomarker. We demonstrate how, by means of this approach, one or more time series of magnetic resonance imaging sequences can serve as predictors of lesion incidence in multiple sclerosis. This application entails sharing information among separate historical models for a set of brain regions, and sheds light on the time lag between early warning signals and clinical manifestation. | Jonathan Gellar (co-author) |
| ▶10:30 am–12:20 pm | CC-W185bc | **Sirken Award Session** | **John Czajka, Chair** |
| ▶2:00–3:50 pm | CC-W190a | **Expanding Capacity for the Measurement of Sexual Orientation and Gender Identity in Federal Surveys** (Social Statistics Section) | **Diane Herz, Chair** |

## Thursday, August 4

| Time | Room | Session | Presenter/Author(s) |
|---|---|---|---|
| ▶10:30 am–12:20 pm | CC-W176c | **Testing for Data Quality: Challenges in Practice** (Government Statistics Section) | **Donsig Jang, Discussant** |
| | CC-W179b | **Methods for Item Missing and Unit Nonresponse** (Survey Research Methods Section) | **Barbara Carlson, Chair** |