

# NESS 2017 Program

## Contents

Welcoming Remarks	2
Keynote Speakers	3
Schedule	6
Detailed Program	7
NESS 2017 Committees	8
Abstracts of Invited Papers	9
Abstracts of Posters	10
NESS 2017 Participants	11

# Welcoming Remarks

## Keynote Speakers

### Hypothesis Testing for Weak and Sparse Alternatives With Applications to Whole Genome Data

**Dr. Xihong Lin**, Harvard University

Massive genetic and genomic data generated using array and sequencing technology present many exciting opportunities as well as challenges in data analysis and result interpretation, e.g., how to develop effective strategies for signal detection using massive genetic and genomic data when signals are weak and sparse. In this talk, I will discuss hypothesis testing for sparse alternatives in analysis of high-dimensional data motivated by gene, pathway/network based analysis in genome-wide association studies using arrays and sequencing data. I will focus on signal detection when signals are weak and sparse, which is the case in genetic and genomic association studies. I will discuss hypothesis testing for signal detection using variable selection based penalized likelihood based methods, the Generalized Higher Criticism (GHC) test, and the Generalized Berk-Jones test, and the robust omnibus test. I will discuss the challenges in statistical inference in the presence of both between-observation correlation and signal sparsity. The results are illustrated using data from genome-wide association studies and sequencing studies.

---

Xihong Lin is Chair and Henry Pickering Walcott Professor of Department of Biostatistics and Coordinating Director of the Program of Quantitative Genomics at the Harvard T. H. Chan School of Public Health, and Professor of Statistics of the Faculty of Art and Science of Harvard University.

Dr. Lin's research interests lie in development and application of statistical and computational methods for analysis of massive genetic and genomic, epidemiological, environmental, and medical data. She currently works on whole genome sequencing association studies, genes and environment, analysis of integrated data, and statistical and computational methods for massive health science data.

Dr. Lin received the 2002 Mortimer Spiegelman Award from the American Public Health Association and the 2006 COPSS Presidents' Award. She is an elected fellow of ASA, IMS, and ISI. Dr. Lin received the MERIT Award (R37) (2007–2015), and the Outstanding Investigator Award (OIA) (R35) (2015–2022) from the National Cancer Institute. She is the contacting PI of the Program Project (PO1) on Statistical Informatics in Cancer Research, the Analysis Center of the Genome Sequencing Program of the National Human Genome Research Institute, and the T32 training grant on interdisciplinary training in statistical genetics and computational biology. Dr. Lin was the former Chair of the COPSS (2010–2012) and a former member of the Committee of Applied and Theoretical Statistics (CATS) of the National Academy of Science. She is the former Chair of the new ASA Section of Statistical Genetics and Genomics. She was the former Coordinating

Editor of Biometrics and the founding co-editor of Statistics in Biosciences, and is currently the Associate Editor of Journal of the American Statistical Association. She has served on a large number of statistical society committees, and NIH and NSF review panels.

---

## Honest Learning for the Healthcare System: Large-scale Evidence from Real-world Data

**Dr. David Madigan**, Columbia University

(joint work with Martijn J. Schuemie, Patrick B. Ryan, George Hripcsak, and Marc A. Suchard)

In practice, our learning healthcare system relies primarily on observational studies generating one effect estimate at a time using customized study designs with unknown operating characteristics and publishing—or not—one estimate at a time. When we investigate the distribution of estimates that this process has produced, we see clear evidence of its shortcomings, including an over-abundance of estimates where the confidence interval does not include one (i.e. statistically significant effects) and strong indicators of publication bias. In essence, published observational research represents unabashed data fishing. We propose a standardized process for performing observational research that can be evaluated, calibrated and applied at scale to generate a more reliable and complete evidence base than previously possible, fostering a truly learning healthcare system. We demonstrate this new paradigm by generating evidence about all pairwise comparisons of treatments for depression for a relevant set of health outcomes using four large US insurance claims databases. In total, we estimate 17,718 hazard ratios, each using a comparative effectiveness study design and propensity score stratification on par with current state-of-the-art, albeit one-off, observational studies. Moreover, the process enables us to employ negative and positive controls to evaluate and calibrate estimates ensuring, for example, that the 95% confidence interval includes the true effect size approximately 95% of time. The result set consistently reflects current established knowledge where known, and its distribution shows no evidence of the faults of the current process. Doctors, regulators, and other medical decision makers can potentially improve patient-care by making well-informed decisions based on this evidence, and every treatment a patient receives becomes the basis for further evidence.

---

David Madigan is the Executive Vice-President for Arts & Sciences, Dean of the Faculty, and Professor of Statistics at Columbia University in the City of New York. He previously served

as Chair of the Department of Statistics at Columbia University (2008–2013), Dean, Physical and Mathematical Sciences, Rutgers University (2005–2007), Director, Institute of Biostatistics, Rutgers University (2003–2004), and Professor, Department of Statistics, Rutgers University (2001–2007). He received his bachelor’s degree in Mathematical Sciences (1984, First Class Honours, Gold Medal) and a Ph.D. in Statistics (1990), both from Trinity College Dublin.

Dr. Madigan has over 160 publications in such areas as Bayesian statistics, text mining, Monte Carlo methods, pharmacovigilance and probabilistic graphical models. In recent years he has focused on statistical methodology for generating reliable evidence from large-scale healthcare data. From 2011 to 2014 he was a member of the FDA’s Drug Safety and Risk Management Advisory Committee.

Dr. Madigan is a fellow of the American Association of the Advancement of Science (AAAS), the Institute of Mathematical Statistics (IMS) and the American Statistical Association (ASA), and an elected member of the International Statistical Institute (ISI). He served as Editor-in-Chief of Statistical Science (2008–2010) and Statistical Analysis and Data Mining, the ASA Data Science Journal (2013–2015).

# Schedule

## Friday, April 21, 2017

08:30am—05:00pm      NESS short courses at Rome Ballroom

## Saturday, April 22, 2017

All activities will be held in Rome Ballroom except where otherwise noted

08:30am—09:15am      Registration & Refreshment & Poster Session

09:15am—09:30am      Welcoming Remarks

09:30am—10:30am      Keynote Presentation:

**David Madigan**, Columbia University

10:30am—10:45am      Coffee Break

11:00am—12:45pm      Parallel Invited Sessions (**Laurel / Oak Halls**)

12:45pm—02:00pm      Lunch, Poster Session (continued)

01:00pm—02:00pm      Poster Session (continued)

02:10pm—02:40pm      Special Session: New England Statistical Society

02:40pm—03:40pm      Keynote Presentation:

**Xihong Lin**, Harvard University

03:40pm—03:55pm      Coffee Break

04:10pm—05:55pm      Parallel Invited Sessions (**Laurel / Oak Halls**)

05:55pm—06:30pm      Travelers Reception, Student Paper and Poster Awards Ceremony

07:00pm—09:00pm      NESS Dinner (signing up required with limited space; held at **Sichuan Pepper in Vernon**.)

## Detailed Program

## NESS 2017 Committees



## Abstracts of Invited Papers

## Abstracts of Posters

## NESS 2017 Participants