

NESS 2017 Program

Contents

Welcoming Remarks	2
Keynote Speakers	3
Schedule	6
Detailed Program	7
NESS 2017 Committees	18
Abstracts of Invited Papers	19
Abstracts of Posters	20
NESS 2017 Participants	21

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

Morning sessions

1. New Vistas in Statistics with Applications

Organizer: **Aleksey Polunchenko**

1. **Aleksey Polunchenko**, Binghamton University
2. **Vasanthan Raghavan**, Qualcomm Flarion Technologies, New Jersey
3. **Zuofeng Shang**, Binghamton University
4. **Emmanuel Yashchin**, IBM

Oak Hall 235

2. Non-clinical in Pharmaceutical Industry

Organizer and Chair: **Chi-Hse Teng**

1. **Don Bennett**, Pfizer
2. **Jerry Lewis**, Biogen
3. **Ray Liu**, Takeda
4. **Chi-Hse Teng**, Novartis

Oak Hall 267

3. Space-Time Statistical Solutions at IBM Research

Organizer: **Yasuo Amemiya**

1. **Julie Novak**, IBM T. J. Watson Research Center
“Revenue Assessment in Large-Scale Businesses”
2. **Xiao Liu**, IBM T. J. Watson Research Center
“A Spatio-Temporal Modeling Approach for Weather Radar Image Data”

3. **Rodrigue Ngueyep Tzoumpe**, IBM T. J. Watson Research Center
“Spatial Segmentation of Spatial-Temporal Lattice Models for Agricultural Management Zoning”
4. **Yasuo Amemiya**, IBM T. J. Watson Research Center
“Spatio-Temporal Analysis for System Management”

Oak Hall 269

4. Graphical Models, Networks, Regulatome and Multivariate Analysis

Organizer and Chair: Yuping Zhang

1. **Forrest W. Crawford**, Yale
“Causal Inference for Network Epidemics”
2. **Zhengqing Ouyang**, Jackson Labs
3. **Sijian Wang**, University of Wisconsin Madison
4. **Kuang-Yao Lee**, Yale
“Learning Causal Networks via Additive Faithfulness”

Oak Hall 268

5. Big Data

Organizer and Chair: Haim Bar

1. **Jacob Bien**, Cornell University
“Learning Local Dependence in Ordered Data”
2. **Li Ma**, Duke University
“Fisher exact scanning for dependency”
3. **Pengsheng Ji**, University of Georgia
“Flexible Spectral Methods for Community Detection”
4. **Chihwa Kao**, University of Connecticut
“Large Dimensional Econometrics and Identification”

Laurel Hall 301

6. Bayesian Applications in High-Dimensional and Multivariate Modeling

Organizer and Chair: **Seongho Song**

1. **Seongho Song**, University of Cincinnati
“Bayesian Multivariate Gamma-Frailty Cox Model for Clustered Current Status Data”
2. **Xia Wang**, University of Cincinnati
“Scalable Massive Multivariate Data Modeling”
3. **Gyuhyeong Goh**, Kansas State University
“Bayesian Variable Selection using Marginal Posterior Consistency”
4. **Jian Zou**, Worcester Polytechnic Institute
“High Dimensional Dynamic Modeling for Massive Spatio-Temporal Data”

Laurel Hall 308

7. New Advances in Analysis of Complex Data: Heterogeneity and High Dimensions

Organizer and Chair: **Min-ge Xie**

1. **Dungang Liu**, University of Cincinnati
“Nonparametric Fusion Learning: Synthesize Inferences from Diverse Sources using Confidence Distribution, Data Depth and Bootstrap”
2. **Dan Yang**, Rutgers University
“Bilinear Regression with Matrix Covariates in High Dimensions”
3. **Pierre Bellec**, Rutgers University
“Slope Meets Lasso in Sparse Linear Regression”
4. **Yiyuan She**, Florida State University
“On cross-validation for sparse reduced rank regression”

Laurel Hall 206

8. Machine Learning and Big Data Analytics

Organizer and Chair: **Jinbo Bi**

1. **Sanguthevar Rajasekaran**, University of Connecticut
“The closest pair problem: Algorithms and applications”

2. **Renato Polimanti**, Yale University
“Resources to Investigate the Genetic Architecture of Complex Traits: Large-Scale Datasets and Summary Association Data”
3. **Sheida Nabavi**, University of Connecticut
“Statistical machine learning to identify candidate drivers of drug resistance in cancer”
4. **Michael Kane**, Yale University
“A First Look at Using Human Mobility Data to Assess Community Resilience”

Laurel Hall 306

9. Statistical Approaches in Modeling and Incorporating Dependence

Organizer and Chair: Ting Zhang

1. **Mengyu Xu**, University of Central Florida
“Pearsons Chi-Squared Statistics: Approximation Theory and Beyond”
2. **Kun Chen**, UConn
“Robust Dimension Reduction of Correlated Multivariate Data”
3. **Liliya Lavitas**, Boston University
“Unsupervised Self-Normalized Change-Point Testing for Time Series”
4. **Buddika Peiris**, Worcester Polytechnic Institute
“Constrained Inference in Regression”

Laurel Hall 309

10. Survival Analysis

Organizer and Chair: Sy Han Chiou

1. **Daniel Nevo**, Harvard
2. **Bella Vakulenko-Lagun**, Harvard
3. **Jing Qian**, UMass
4. **Sangwook Kang**

Laurel Hall 302

11. Extremes

Organizer and Chair: **Richard Davis, Phyllis Wan**

1. **John Nolan**, American University
“Mvevd: An R Package for Extreme Value Distributions”
2. **Jingjing Zou**, Columbia University
“Extreme Value Analysis without the Largest Values: What can be Done?”
3. **Karthikey Murthy**, Columbia University
“Distributionally Robust Extreme Value Analysis”
4. **Tiandong Wang**, Cornell University
“Asymptotic Normality of Degree Counts in the Preferential Attachment Network”

Laurel Hall 305

12. Feinberg Memorial Session: Bayesian Statistics with Applications

Organizer and Chair: **Dipak Dey**

1. **Edoardo Airoldi**, Harvard University
“Bayesian Methods for Protein Quantification”
2. **Bani Mallick**, Texas A&M University
“Fast Sampling with Gaussian Scale-Mixture Priors in High Dimensional Regression”
3. **Sudipto Banerjee**, UCLA
“High-Dimensional Bayesian Geostatistics”

Laurel Hall 307

Afternoon sessions

1. Panel Discussion on Careers in Statistics

Organizer and Chair: **Naitee Ting**

1. **Birol Emir**, Pfizer
2. **Chun Wang**, University of Connecticut
3. **Yasuo Amemiya**, IBM T. J. Watson Research Center

4. **Minge Xie**

Oak Hall 235

2. **Statistical Applications in Finance and Insurance**

Organizer and Chair: Guojun Gan

1. **Liang Peng**, Georgia State University
“Inference for Predictive Regressions”
2. **Fangfang Wang**, University of Connecticut
“A Common Factor Analysis of Stock Market Trading Activity”
3. **Oleksii Mostovyi**, University of Connecticut
“Sensitivity analysis of the expected utility maximization problem”
4. **Kun Chen**, University of Connecticut
“Towards differential pricing in auto insurance via large-scale predictive modeling: a partnership between Travelers and UConn”

Oak Hall 267

3. **Application of Statistical/Predictive Modeling in Health Related Industry**

Organizer and Chair: Nan Shao

1. **Xiaoyu Jia**, Icahn School of Medicine at Mount Sinai
2. **Zhaonan Sun**, IBM T. J. Watson Research
“Exploiting Convolutional Neural Network for Risk Prediction with Medical Feature Embedding”
3. **Victoria Gamerman**, Boehringer Ingelheim Pharmaceuticals, Inc.
4. **Nan Shao**, New York Life Insurance
“Statistical Modeling in the Life Insurance Industry”

Oak Hall 268

4. Biopharmaceutical session

Organizer and Chair: **Adina Soaita**

1. **Abidemi Adeniji**, EMD Serono
2. **Bushi Wang**
3. **Joseph C Cappelleri**, Pfizer
“Meta-Analysis of Safety Data in Clinical Trials”
4. **Qiqi Deng**, Boehringer Ingelheim
5. **Birol Emir**, Pfizer

Oak Hall 269

5. Complex Data/Network Modeling

Organizer and Chair: **Yuan Huang**

1. **Yize Zhao**, Weill Cornell Medical College, Cornell
“Hierarchical Feature Selection of the Complex Biomedical Data”
2. **Heather Shappell**, Biostatistics, Boston University
“Methods for Longitudinal Complex Network Analysis in Neuroscience”
3. **Krista Gile**, Math and Statistics, UMASS
“Inference from Link-Tracing Network Samples”
4. **Xizhen Cai**, Temple
“Variable Selection for Dynamic Networks”
5. **Xuan Bi**, Department of Biostatistics, Yale University
“Genome-Wide Mediation Analysis of Psychiatric and Cognitive Traits in the Philadelphia Neurodevelopmental Cohort”

Laurel Hall 301

6. Spatial Analysis of Public Health Data

Organizer and Chair: **Beth Ziniti**

1. **Harrison Quick**, Dornsife School of Public Health, Drexel University
“Spatiotemporal Trends in Heart Disease Mortality”

2. **Joshua Warren**, Yale School of Public Health
“A Bayesian Spatial Kernel Smoothing Method to Estimate Local Vaccine Uptake using Administrative Records”
3. **Gavino Puggioni**, University of Rhode Island
“Spatiotemporal Analysis of Vector-Borne Disease Risk”
4. **Chanmin Kim**, Harvard T. H. Chan School of Public Health
“Public Health Impact of Pollutant Emissions”

Laurel Hall 308

7. Network Data Analysis

Organizer and Chair: **Edoardo M. Airoldi**

1. **JP Onnela**, Harvard University
“Inference and model selection for mechanistic network models”
2. **Vishesh Karwa**, Harvard University
“Estimating average treatment effects under interference: Modes of failure and solutions”
3. **Xinran Li**, Harvard University
“Randomization Inference for Peer Effects”

Laurel Hall 206

8. Statistical Approaches to Data Modeling and Analysis

Organizer and Chair: **Erin Conlon**

1. **Evan Ray**, University of Massachusetts Amherst
“Feature-Weighted Ensembles for Probabilistic Time-Series Forecasts”
2. **Daeyoung Kim**, University of Massachusetts Amherst
“Assessment of the Adequacy of Asymptotic Theory in Statistical Inference”
3. **Patrick Flaherty**, University of Massachusetts
“A Deterministic Global Optimization Method for Variational Inference”
4. **Matthias Steinruecken**, University of Massachusetts Amherst
“Unraveling the Demographic History of Modern Humans using Full- Genome Sequencing Data”

5. **Zheng Wei**, University of Massachusetts Amherst
 “On Multivariate Asymmetric Dependence Using Multivariate Skew-Normal Copula-Based Regression”

Laurel Hall 306

9. Social Networks and Causal Inference

Organizer and Chair: Daniel Sussman

1. **Daniel Sussman**, Boston University
 “Optimal Unbiased Estimation of Causal Effects under Network Interference”
2. **Alex Volfovsky**, Duke University
 “Causal Inference in the Presence of Networks: Randomization and Observation”
3. **Dean Eckles**, Massachusetts Institute of Technology
 “Estimating Peer Effects in Networks with Peer Encouragement Designs”
4. **Hyunseung Kang**, University of Wisconsin at Madison
 “Peer Encouragement Designs in Causal Inference with Partial Interference and Identification of Local Average Network Effects”

Laurel Hall 309

10. Statistical Innovations in Genomics

Organizer and Chair: Zhengqing Ouyang

1. **Hongkai Ji**, Johns Hopkins Bloomberg School of Public Health
2. **Pei Wang**, Mount Sinai School of Medicine
 “Constructing Tumor-Specific Gene Regulatory Networks Based on Samples with Tumor Purity Heterogeneity”
3. **Yuping Zhang**, University of Connecticut
4. **Kai Wang**, Columbia University
 “Long Read Sequencing to Study Human Genome Variation”

Laurel Hall 302

11. Recent Developments on High-Dimensional Statistics and Regularized Estimation

Organizer and Chair: **Kun Chen**

1. **Ethan Fang**, Penn State
“Blessing of Massive Scale: Spatial Graphical Model Estimation with a Total Cardinality Constraint Approach”
2. **Cheng Yong Tang**, Temple University
“Sufficient Dimension Reduction with Missing Data”
3. **Sahand Nagahban**, Yale University
“Restricted Strong Convexity Implies Weak Sub-Modularity”
4. **Ting Zhang**, Boston University
“A Thresholding-Based Prewhitened Long-Run Variance Estimator and Its Dependence-Oracle Property”

Laurel Hall 305

12. Subgroup Analysis

Organizer and Chair: **Xiaoqing Wang**

1. **Yanxun Xu**, Johns Hopkins University
“A Nonparametric Bayesian Basket Trial Design”
2. **Lynn Lin**, Pennsylvania State University
“Clustering with Hidden Markov Model on Variable Blocks”
3. **Jared Huling**, University of Wisconsin-Madison
“Heterogeneity of Intervention Effects and Subgroup Identification based on Longitudinal Outcomes”
4. **Wai-Ki Yip**, Foundation Medicine, Inc.
“STEPP Analysis for continuous, binary, and count outcomes and other recent STEPP development”

Laurel Hall 307

NESS 2017 Committees

Abstracts of Invited Papers

Abstracts of Posters

NESS 2017 Participants