INTERNATIONAL CHINESE STATISTICAL ASSOCIATION
YEAR 2010 APPLIED STATISTICS SYMPOSIUM

泛華統計協會

The 19th Annual ICSA

Applied Statistics Symposium

June 20 – 23, 2010

Hyatt Regency Downtown Indianapolis, IN, USA

WELCOME AND ACKNOWLEDGEMENTS

On behalf of the ICSA 2010 organizing committee, we warmly welcome you to Indianapolis, the Crossroads of America, for the 19th annual ICSA Applied Statistics Symposium. We are very proud and excited to host the 2010 symposium in Indianapolis for the first time.

Indianapolis is the 14th largest city in the U.S., home to the Indianapolis Colts and Indiana Pacers, and the site of the Indianapolis 500-mile race. Known as The Amateur Sports Capital of the World, Indianapolis has hosted the Final Four of the NCAA basketball championship several times, including the memorable 2010 Final Four, where we witnessed a miraculous run by our hometown favorite – the Butler Bulldogs. Central Indiana is also home to over 400 well-educated and experienced statisticians, employed mainly in the life science industry and major research institutes.

To those of you who have passionately nurtured, served, and empowered the ICSA for years, this is another memorable milestone; to those who are attending the symposium for the first time and have just joined the ICSA, this will be the beginning of a rewarding journey in your professional career. We hope you enjoy your experience at this year's symposium in all aspects: an excellent scientific program, great opportunities for learning and networking, and warm hospitality from the Hoosier state. Have a wonderful time in Central Indiana.

Please allow us to take this opportunity to thank all participants, including keynote speakers, short course instructors, session organizers and chairs, invited and contributed speakers, and student award participants and winners.

We would also like to thank our corporate sponsors and collaborators for their generous support. In particular, we would like to thank Eli Lilly, Amgen, Celgene, InVentiv, the American Statistical Association (ASA) Biopharmaceutical Section, Sanofi-Aventis, ASA Central Indiana Chapter, Pfizer, BioPIER, and all the student volunteers and many other individuals for their contributions and support.

Finally we would like to thank our fellow organizing committee members and their families. We are profoundly proud of the superb work done by our committee members in preparation for this symposium. Their dedication and hard work have once again proved that together we can make this organization one of the best in the professional world.

Yours sincerely,

Yongming Qu & Wei Shen

Co-Chairs of the ICSA 2010 Organizing Committee

ORGANIZING COMMITTEES

Executive Committee

Name	Responsibility	Affiliation
Wei Shen	Co-Chair	Eli Lilly and Company
Yongming Qu	Co-Chair	Eli Lilly and Company
Jun Xie	Co-Chair (Program), Fund Raising	Purdue University
Menggang Yu	Co-Chair (Program)	Indiana University
Rick Chappell	Student Paper Award, Co-Chair	University of Wisconsin
Jiashun Jin	Student Paper Award, Co-Chair	Carnegie Mellon University
Ming-Dauh Wang	Treasurer	Eli Lilly and Company
Grace Li	Chair, Local Events; Fund Raising	Eli Lilly and Company
Ning Li	Program Liaison	FDA
Haoda Fu	Webmaster	Eli Lilly and Company

Scientific Program Committee

Name	Affiliation	Email
Jun Xie	Purdue University	junxie@stat.purdue.edu
Menggang Yu	Indiana University	meyu@iupui.edu
Chuanhai Liu	Purdue University	chuanhai@stat.purdue.edu
Hao Zhang	Purdue University	zhanghao@stat.purdue.edu
Michael Zhu	Purdue University	yuzhu@stat.purdue.edu
Jiashun Jin	Carnegie Mellon University	jiashun@stat.cmu.edu
Ji Zhu	University of Michigan	jizhu@umich.edu
Annie Qu	UIUC	anniequ@illinois.edu
Bin Nan	University of Michigan	bnan@umich.edu
Spencer Huang	Northwestern University	huangcc@northwestern.edu
Xiaochun Li	IUPUI	Xiaochun@iupui.edu
Dawei Xie	University of Pennsylvania	dxie@mail.med.upenn.edu
Yuan Ji	MD Anderson	yuanji@mdanderson.org
Qi Long	Emory University	qlong@sph.emory.edu
Liqiang Ni	Central Florida University	lni@mail.ucf.edu
Ilya Lipkovich	Eli Lilly and Company	lipkovichia@lilly.com
Yanping Wang	Eli Lilly and Company	wang_yanping@lilly.com

Shengyan Hong	MedImmune, Inc	hongs@medimmune.com
Chengxing Lu	Merck Co., Inc	chengxing_lu@merck.com
Yuefeng Lu	Eli Lilly and Company	lu_yuefeng@lilly.com
Lei Shen	Eli Lilly and Company	shenle@lilly.com

Local Activity Committee

Name	Affiliation	Email
Grace Li	Eli Lilly and Company	li_ying_grace@lilly.com
Michael Case	Eli Lilly and Company	mgcase@lilly.com
Hank Wei	Eli Lilly and Company	hank_wei@lilly.com
Xiaohai Wan	Eli Lilly and Company	wanxi@lilly.com
Lei Xu	Eli Lilly and Company	xule@lilly.com
Xianming Zheng	Eli Lilly and Company	zhengxi@lilly.com
Rong Liu	Eli Lilly and Company	liurongrl@lilly.com
Jian Chen	Eli Lilly and Company	chenjianx1@lilly.com

Volunteering Assistants

Madan Gopal Kundu	IUPUI	Yanzhu Lin	Purdue University
Lili Yang	IUPUI	Kean Ming Tan	Purdue University
Jingwei Wu	IUPUI	Xin Lu Tan	Purdue University
Ritabrata Dutta	Purdue University	Libo Wang	Purdue University
Youran Fan	Purdue University	Han Wu	Purdue University
Xiang Han	Purdue University	Jin Xia	Purdue University
Shuang He	Purdue University	Tianhong He	Purdue University
Xian He	Purdue University	Vitara Pungpapong	Purdue University
Meihua QIao	Eli Lilly & Company	Zhuangzhuang Xi	Purdue University
Chang Xu	Eli Lilly & Company	Chunmei Zhou	Eli Lilly & Company
Yanni Zhu	Eli Lilly & Company	Cathy Bolling	Eli Lilly & Company
Cathy Bolling	Eli Lilly & Company		

SPONSORS

ASA Biopharmaceutical Section

ASA Central Indiana Chapter

EXHIBITORS

ICSA

K-Force

Precision Therapeutics Inc

Springer

Taylor & Francis Group

PROGRAM OVERVIEW

	Sunday, June 20, 2010	
	Session	**Rooms**
7:15 a.m. – 9:00 a.m.	Breakfast (provided)	Regency E&F
9:00 a.m.–12 a.m.	Full-Day Short courses #1 and #2	Network, Studio Five
8:00 a.m.–12 a.m.	Half-Day Short Courses #3 and #4;	Concept AB, Concept CD
12:00 p.m.–1:00 p.m.	Lunch (provided)	Regency E&F
1:00 p.m.–4:30 p.m.	Full-Day Short course #1 and #2	Network, Studio Five
1:00 p.m.–5:00 p.m.	Half-Day Short course #5 and #6	Concept AB, Concept CD
5:00 p.m.–8:00 p.m.	ICSA Board Meeting (Closed)	Vision
6:30 p.m.–9:00 p.m.	Mixer	Regency AB

	Monday, June 21, 2010	
	Session	**Rooms**
7:00 a.m. – 8:00 a.m.	Breakfast (provided)	Cosmopolitan B
8:00 a.m.–10:00 a.m.	Keynote Address #1	Regency AB
10:20 a.m.–12:10 p.m.	Parallel sessions	See program schedules
12:10 p.m.–1:30 p.m.	Lunch (on your own)	
12:20 p.m.–1:20 p.m.	Round Table Discussion	Vision
12:20 p.m.–1:20 p.m.	Lunch with keynote speakers (by invitation)	Studio Six
1:30 p.m.–3:20 p.m.	Parallel sessions	See program schedules
3:40 p.m.–5:30 p.m.	Parallel sessions	See program schedules

	Tuesday, June 22, 2010	
	Session	**Rooms**
7:00 a.m. – 8:00 a.m.	Breakfast (provided)	Cosmopolitan B
8:00 a.m.–10:00 a.m.	Keynote Address #2	Regency AB
10:20 a.m. –12:10 p.m.	Parallel sessions	See program schedules
12:10 p.m.–1:30 p.m.	Lunch (on your own)	
12:20 p.m.–1:20 p.m.	Round Table Discussion	Vision
1:30 p.m.–3:20 p.m.	Parallel sessions	See program schedules
3:40 p.m.–5:30 p.m.	Parallel sessions	See program schedules
6:30 p.m.–9:00 p.m.	Banquet	On Time Seafood Restaurant

	Wednesday, June 23, 2010	
	Session	**Rooms**
7:00 a.m. – 8:00 a.m.	Breakfast (provided)	Cosmopolitan B
8:00 a.m.–9:50 a.m.	Parallel sessions	See program schedules
10:10 a.m.–12:00 p.m.	Parallel sessions	See program schedules

PROGRAM SCHEDULE–SHORT COURSES

Sunday, June 20

Short Course 1: **Recent Developments in Practical Bayesian Methods for Clinical Trials**

Instructor(s): Dr. Peter F. Thall, M.D. Anderson Cancer Center

Location and Time: Network Room, Sunday, June 20, 2010, 8:00 a.m.– 5:30 p.m.

Short Course 2: **Modern Techniques in Data Mining**

Instructor(s): Dr. David Banks, Professor in the Department of Statistical Science at Duke University

Location and Time: Studio Five Room, Sunday, June 20, 2010, 8:00 a.m.– 5:30 p.m.

Short Course 3: **An Introduction to Propensity Score Methods in Observational Research**

Instructor(s): Dr. Peter Austin, Sr. Scientist at the Institute for Clinical Evaluative Sciences (ICES) in Toronto, Canada

Location and Time: Concept AB Room, Sunday, June 20, 2010, 8:00 a.m.–12:00 p.m.

Short Course 4: **Post-Marketing Drug Safety Evaluation**

Instructor(s): Dr. A. Lawrence Gould, Merck Research Laboratories

Location and Time: Concept CD Room, Sunday, June 20, 2010, 8:00 a.m.–12:00 p.m.

Short Course 5: **Principles and Techniques of Multiple testing and Multiple Comparisons**

Instructor(s): Dr. Jason C. Hsu, The Ohio State University

Location and Time: Concept AB Room, Sunday, June 20, 2010, 1:30 p.m.–5:30 p.m.

Short Course 6: **Pharmacogenomics Clinical Trials: Genomic Biomarker Associated Design and Analysis Issues**

Instructor(s): Dr. Sue-Jane Wang, U.S. Food and Drug Administration

Location and Time: Concept CD Room, Sunday, June 20, 2010, 1:30 p.m.–5:30 p.m.

PROGRAM SCHEDULE–KEYNOTE SPEECHES

Monday , June 21

K01-A: **FDA's Critical Path Initiative: Updates, Opportunities, and New Challenges**

Speaker: Dr. ShaAvhrée Buckman, Food and Drug Administration

Organizer: Jun Xie, Purdue University
Chair: Steve Ruberg, Eli Lilly and Company
Location and Time: Regency AB, Monday, June 21, 8:00 a.m.– 10:00 a.m.

K01-B: **Statisticians in the Pharmaceutical Industry: The 21st Century**

Speaker: Dr. Christy Chuang-Stein, Pfizer, Inc.

Organizer: Jun Xie, Purdue University
Chair: Steve Ruberg, Eli Lilly and Company
Location and Time: Regency AB, Monday, June 21, 8:00 a.m.– 10:00 a.m.

K01-C: **The future of statistics and the future of statisticians - an idiosyncratic assessment**

Speaker: Dr. Donald Rubin, Harvard University

Organizer: Jun Xie, Purdue University
Chair: Steve Ruberg, Eli Lilly and Company
Location and Time: Regency AB, Monday, June 21, 8:00 a.m. – 10:00 a.m.

Tuesday, June 22

K02-A: **Statistical Issues and Challenges in Analyzing High-throughput 'Omics Data in Population-Based Studies**

Speaker: Dr. Xihong Lin, Harvard University

Organizer: Yanping Wang, Eli Lilly and Company
Chair: Walt Offen, Eli Lilly and Company
Location and Time: Regency AB, Tuesday, June 22, 8:00 a.m.– 10:00 a.m.

K02-B: **Statistical Challenges in Personalized Medicine**

Speaker: Dr. Gregory Campbell, Food and Drug Administration

Organizer: Yanping Wang, Eli Lilly and Company
Chair: Walt Offen, Eli Lilly and Company
Location and Time: Regency AB, Tuesday, June 22, 8:00 a.m.– 10:00 a.m.

Banquet Keynote Speech: **Statistics: a critical pharma--academic interface**

Speaker: Dr. William W. Chin, Harvard Medical School

Organizer: Grace Li, Eli Lilly and Company
Chair: Wei Shen, Eli Lilly and Company
Location and Time: On Time Seafood Restaurant, Tuesday, June 22, 6:30 p.m.–9:00 p.m.

PROGRAM SCHEDULE—ROUND TABLES

Monday, June 21

Round Table L1: **Implementation Issues of Adaptive Design Clinical Trials**
Discussion Leader(s): Dr. Cyrus Mehta, Dr. Jason Connor
Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Round Table L2: **Opportunities and Challenges to Build Your Own Consulting Firm**
Discussion Leader(s): Dr. Tai Xie
Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Round Table L3: **Monitoring Safety During Drug Development**
Discussion Leader(s): Dr. Brenda Crowe
Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Round Table L4: **Biopharmeutical Statistics and Biostatistics Development in China**
Discussion Leader(s): Professor Jun Shao, Ms. Helen Yin
Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Round Table L5: **Effective Communications Between Industry and FDA**
Discussion Leader(s): Dr. Suktae Choi, Dr. James Hung
Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Tuesday, June 22

Round Table L6: **Statistical Outsourcing: Issues and Future**
Discussion Leader(s): Dr. Todd Sanger
Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Round Table L7: **How to Get Research Funding from Various Sources: Industry, Government and Academia**
Discussion Leader(s): Dr. Jeremy M G Taylor
Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Round Table L8: **Career in Industry: Technical or Administrative?**
Discussion Leader(s): Dr. Gordon Lan, Dr. Pandurang Kulkarni
Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Round Table L9: **Careers for New Statisticians in Industry**
Discussion Leader(s): Dr. Christy Chuang-Stein
Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

PROGRAM SCHEDULE–PARALLEL SESSIONS

Monday, June 21

Monday, June 21 10:20 a.m.–12:10 p.m.	Concept AB
Session M11	**Advanced Bayesian Adaptive Designs for Clinical Trials—Invited**
Organizer(s): Yuan Ji (The University of Texas M. D. Anderson), **Chair(s)**: Yuan Ji (The University of Texas M. D. Anderson)	
Pseudo-Data Augmented Bayesian Weighted Likelihood Bootstrap for Early Phase Oncology Trial Designs—Neby Bekele (M.D. Anderson Cancer Center) **A Bayesian Geometric Phase II-III Select-and-Test Design Based On Treatment Failure Time and Toxicity—**Peter Thall (M.D. Anderson Cancer Center) **Continual Reassessment Method for Late-Onset Toxicities Using Data Augmentation—**Ying Yuan (Univ. of Texas MD Anderson Cancer Center)	
Monday, June 21 10:20 a.m.–12:10 p.m.	**Concept CD**
Session M12	**Joint modeling of Survival and Longitudinal data—Invited**
Organizer(s): Grace Y. Yi (University of Waterloo), **Chair(s)**: Grace Y. Yi (University of Waterloo)	
Modeling Left-truncated and Right Censored Survival data with longitudinal covariates—Jane-Ling Wang (Univ. of California, Davis) **Using joint longitudinal and survival models for individual prediction.—**Jeremy Taylor (University of Michigan) **Joint Analysis of Longitudinal Growth and Interval Censored Mortality Data—**Charmaine Dean (Simon Fraser University) **Joint modelling longitudinal and survival data with missing values and measurement errors—**Lang Wu (University of British Columbia)	
Monday, June 21 10:20 a.m.–12:10 p.m.	**Network**
Session M13	**Special topic: Statistical research collaboration between Eli Lilly and University of Wisconsin-Madison—Invited**
Organizer(s): Yun-Fei Chen (Eli Lilly and Company), **Chair(s)**: Scott Clark (Eli Lilly and Company)	
From Mad as a Hatter to Madison and Indy: The Lilly/UW-Madison Collaboration—Scott Clark (Eli Lilly and Company) **A multiple testing procedure for combination drugs with two study endpoints—**Jun Shao (University of Wisconsin-Madison) **Subgroups defined by two or more genetic markers in genome-wide association studies—**Xu He (University of Wisconsin-Madison) **Defining Responder in Alcohol Dependence Patients—**Xiwen Ma (University of Wisconsin-Madison)	
Monday, June 21 10:20 a.m.–12:10 p.m.	**Cosmopolitan C**
Session M14	**Semiparametric Models—Invited**
Organizer(s): Lu Wang (The University of Michigan), **Chair(s)**: Lu Wang (The University of Michigan)	
Generalized Functional Latent Feature Models with Single-Index Interactions—Raymond J. Carroll (Texas A&M University) **Estimation for two - phase designs: semiparametric models and Z-theorems—**Jon A. Wellner (University of	

Washington, Statistics)

Estimating the Average Effect on the Survival Function of a Time-Dependent Treatment—Douglas E. Schaubel (University of Michigan)

How many iterations are sufficient for semiparametric estimation?—Guang Cheng (Dept of Statistics, Purdue University)

Monday, June 21 10:20 a.m.–12:10 p.m.	**Cosmopolitan D**

Session M15 **Interface of statistical learning and nonparametric statistics—Invited**

Organizer(s): Yufeng Liu (University of North Carolina), **Chair(s)**: Yufeng Liu (University of North Carolina)

Adaptive sampling procedures for estimating function thresholds—George Michailidis (The University of Michigan)

Does Modelling Lead to More Accurate Classification?—Yoonkyung Lee (Ohio State University)

Boosting for nonparametric high-dimensional models—Lifeng Wang (Michigan State University)

Non-Concave Penalized Likelihood with NP-Dimensionality—Jinchi Lv (University of Southern California)

Monday, June 21 10:20 a.m.–12:10 p.m.	**Regency A**

Session M16 **Statististics in BioSciences Editors's Invited Session—Invited**

Organizer(s): Xihong Lin (Harvard University), **Chair(s)**: Xihong Lin (Harvard University)

A Novel Matching Approach to Correct for Population Stratification in Case-Control Association Studies—Mike Epstein (Emory University)

Bayesian Analysis of Proteomics Data with Nonrandom Missingness—Hongyu Zhao (Yale University)

The null distributions of test statistics in genomewide association studies—Xiang Chen (Yale Univ)

Monday, June 21 10:20 a.m.–12:10 p.m.	**Regency B**

Session M17 **Informative But Not Invasive—Invited**

Organizer(s): Xiao-Li Meng (Harvard University), **Chair(s)**: Xiao-Li Meng (Harvard University)

Longitudinal Data Systems: New Frontier in Statistics—Jeremy Wu (US Census)

Measurement Error in Large-scale Administrative Record Statistical Systems—John Abowd (Cornell University)

Adapting Multiple Imputation to Protect Confidential Data—Jerry Reiter (Duke University)

Discussion—Alex Blocker (Harvard University)

Monday, June 21 10:20 a.m.–12:10 p.m.	**Studio One**

Session M18 **Statistics in Translational Research—Contributed**

Organizer(s): Shengyan Hong (MedImmune, Inc.), **Chair(s)**: Ling Chen (FDA)

An Efficient Approach for Genome-Wide SNP Selection—Min Zhang (Purdue University)

A Non-Parametric Bayesian AR Model - Application to DNA-sequencing—Maria Anna Di Lucca (M.D. Anderson Cancer Center)

Bayesian graphical model for histone modifications—Riten Mitra (UT MD Anderson Cancer Centre)

The temporal profile analysis of gene expresison for female rainbow trout under normal and compressed cycle—Yushi Liu (the Ohio State University)

Fence Methods for gene mapping under the Backcross Experiments—Thuan Nguyen (Oregon Health & Science University)

Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Network**
Session M21	**Design of clinical trials for targeted treatments guided by pharmacogenomic and other predictive biomarkers—Invited**
Organizer(s): Lei Shen (Eli Lilly and Company), **Chair(s)**: Hong Liu-Seifert (Eli Lilly and Company)	
A Bayesian Covariate-Adjusted Response-Adaptive Design with Biomarkers for Targeted Therapies in Cancer—KyungMann Kim (Univerrsity of Wisconsin-Madison) **Flexible Dose Study Designs: Tailoring the Dose to the Patient—**Walt Offen (Eli Lilly and Company) **Clinical Trial Designs for Predictive Biomarker Validation—**Sumithra Mandrekar (Mayo Clinic) **Statistical Considerations for Trials Incorporating Tailored Subgroups—**Brian A. Millen (Eli Lilly and Company)	
Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Regency A**
Session M22	**Tree-based Methods and Predictive Biomarker Identification for Tailored Therapeutics—Invited**
Organizer(s): Ilya Lipkovich, Yuefeng Lu (Eli Lilly and Company), **Chair(s)**: Stephen J. Ruberg (Eli Lilly and Company)	
Modern Statistical Methods for Subgroup Identification in Clinical Trials—Lei Shen (Eli Lilly and Company) **Decision Trees for Identifying Predictors of Treatment Effectiveness in Clinical Trials—**Heping Zhang (Yale University) **Finding and validating subgroups in clinical trials—**Jeremy Taylor (University of Michigan) **Regulatory Perspectives on Genomic Biomarker classifier in Therapeutic Development—**Sue-Jane Wang (FDA)	
Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Concept AB**
Session M23	**Jiann-Ping Hsu Invited Biopharmaceutical & Regulatory Sciences Session—Invited**
Organizer(s): Karl Peace & Lili Yu (Georgia Southern University), **Chair(s)**: Lili Yu (Georgia Southern University)	
Extended quasi-likelihood in the frame of AFT model—Lili Yu (Georgia Southern University) **The Ultimate Professional Statisticians: Thinkers, Practitioners, and Leaders—**Kao-Tai Tsai (Frontier Informatics Service) **Advanced and Adaptive Designs of thorough QT studies—**Yi Tsong (FDA) **Semiparametric Transformation Models for Panel Count Data with Dependent Observation Process—**Ni Li (University of Missouri)	
Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Concept CD**
Session M24	**Recent Adcances in Semiparametric Inferences—Invited**
Organizer(s): Guang Cheng (Purdue University), **Chair(s)**: Guang Cheng (Purdue University)	
BASELINE ZONE DETECTION VIA P-VALUES.—Mouli Banerjee (University of Michigan, Ann Arbor) **Semiparametric Inference for Partly Single-Index Hazards Model—**Donglin Zeng (University of North Carolina) **Semiparametric Regression with Missing Outcomes Using Weighted Kernel-Profile Estimating Equations—**Lu Wang (University of Michigan) **Adaptive Rank Penalized Estimators in Multivariate Regression—**Flori Bunea (Florida State University)	
Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Studio One**
Session M25	**Medical Imaging and Statistics—Invited**

Organizer(s): Christopher Tong (USDA), **Chair(s)**: Christopher Tong (USDA)

DCE-MRI Quantification and Modeling for Angiogenesis Tumor Animal Model Development—Yuefeng Lu (Eli Lilly and Company)

Functional Mixed-Effects Model for Analyzing fMRI Pain Data—Ragnheidur Haraldsdottir (Columbia University)

The Generalized Shrinkage Estimator for Spectral Analysis of Multivariate Time Series—Mark Fiecas (Brown University)

Longitudinal Image Analysis of Tumor/Healthy Brain Change in Contrast Uptake Induced by Radiation—Xiaoxi (Shelsea) Zhang (Pfizer Inc.)

Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Cosmopolitan C**

Session M26 **Advanced Monte Carlo Methods and Applications—Invited**

Organizer(s): Feng Liang (University of Illinois at Urbana-Champai), **Chair(s)**: Yuguo Chen (University of Illinois at Urbana-Champai)

Estimation of Large Families of Bayes Factors from Markov Chain Output—Hani Doss (University of Florida)

On the computation of normalizing constants—Christophe Andrieu (University of Bristol)

Regional Adaptation for Adaptive MCMC—Radu Craiu (University of Toronto)

Variable at-a-time Markov Chain Monte Carlo—Galin Jones (University of Minnesota)

Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Regency B**

Session M27 **High Dimensional data analysis: sifting a few useful feature from many useless ones—Invited**

Organizer(s): Yingying Fan (University of Southern California), **Chair(s)**: Yichao Wu (North Carolina State University)

FIRST: Combining forward iterative selection and shrinkage in high dimensional linear models—Hao Helen Zhang (North Carolina State University)

A unified algorithm for computing ultra high dimensional penalized models—Hui Zou (University of Minnesota)

Optimal Screening for Sparse Signals—Wenguang Sun (North Carolina State University)

Variable Selection in Linear Mixed Effects Models—Yingying Fan (University of Southern California)

Monday, June 21 **1:30 p.m.–3:20 p.m.**	**Cosmopolitan D**

Session M28 **Causal Inference in Observational Research—Invited**

Organizer(s): Doug Faries (Eli Lilly and Company), **Chair(s)**: Gerhardt Pohl (Eli Lilly and Company)

For Objective Causal Inference, Design Trumps Analysis—Donald Rubin (Harvard University)

Optimal estimation of risk differences using propensity-score matching—Peter Austin (Institute for Clinical Evaluative Sciences)

Making Fair and Efficient Comparisons—Bob Obenchain (Risk Benefit Statistics LLC)

Monday, June 21 **3:40 p.m.–5:30 p.m.**	**Concept AB**
Session M31	**Challenging issues for correlated data—Invited**
Organizer(s): Annie Qu (University of Illinois at Urbana Champai), **Chair(s)**: Annie Qu (University of Illinois at Urbana Champai)	
A Pairwise Likelihood Method for Clustered Binary Data under Generalized Partially Linear Single-Index Models—Grace Yi (University of Waterloo) **Analysis on Censored Quantile Residual Life Model via Spline Smoothing—**Ying Wei (Columbia University) **Mixed-effects model selection using conditional AIC—**Ronghui Xu (University of California, San Diego) **Causal Inference for Continuous Time Longitudinal Data When Covariates Are Observed Only at Discrete Times—**Dylan Small (University of Pennsylvania)	
Monday, June 21 **3:40 p.m.–5:30 p.m.**	**Regency A**
Session M32	**Bayesian methods for tailored therapeutics—Invited**
Organizer(s): Haoda Fu, Ph.D. (Eli Lilly and Company), **Chair(s)**: Stacy Lindborg, Ph.D (Eli Lilly and Company)	
Personlized Medicine and Tailored Therapeutics—William Macias (Eli Lilly and Company) **Personalized Medicine - Using Biomarker Signatures to Predict Response to New Therapies—**J. Kyle Wathen (M. D. Anderson Cancer Center) **BVS with Joint Modeling of Categorical and Survival Outcomes: An Application to Individualizing Chemotherapy—**Wei Chen (Wayne State University) **Discussion—**Jack Lee (Univ. of Texas M.D. Anderson Cancer Center)	
Monday, June 21 **3:40 p.m.–5:30 p.m.**	**Network**
Session M33	**Industry-academic collaboration session: Adaptive Designs in Clinical Trials—Invited**
Organizer(s): Hui Quan (Sanofi-aventis), **Chair(s)**: Jun Xie (Purdue University)	
Comparisons of Procedures for Two-Stage Adaptive Designs in Clinical Trials—Mingyu Li (Celgene Corp.) **Interim Treatment Selection with a Flexible Selection Margin in Clinical Trials—**Yujun Wu (Sanofi-Aventis) **Considerations for Design and Data Analysis of Adaptive Superiority/Non-Inferiority Cardiovascular Trials—**Meehyung Cho (Sanofi-aventis) **Discussion—**Abdus S. Wahed (University of Pittsburgh)	
Monday, June 21 **3:40 p.m.–5:30 p.m.**	**Concept CD**
Session M34	**Testing/Clustering nonlinear time series data —Invited**
Organizer(s): Jing Wang (University of Illinois at Chicago), **Chair(s)**: Lan Xue (Oregon State University)	
Global Property of Error Density Estimation in Nonlinear Autoregressive Time Series Models—Fuxia Cheng (Illinois State University) **Comparing two nonparametric regression curves in the presence of long memory in covariates and errors—**Fang Li (IUPUI) **Bispectral-based methods for clustering nonlinear time series—**Jane Harvill (Baylor University) **Modeling time series via spline confidence bands and block bootstrap—**Jing Wang (University of Illinois at Chicago)	
Monday, June 21 **3:40 p.m.–5:30 p.m.**	**Regency B**
Session M35	**Statistical Challenges in Next Generation Sequencing Data Analysis—Invited**

Organizer(s): Hongmei Jiang, Lingling An (Northwestern University, The University), **Chair(s)**: Lingling An (The University of Arizona)

Discovering cis-regulatory regions from differential nucleosome occupancy patterns—Xiaole Shirley Liu (Dana-Farber Cancer Institute / Harvard)

STAT1 regulates microRNA transcription in interferon γ – stimulated HeLa cells—Lang Li (Indiana University)

Statistical issues for analysis of metagenomics sequencing data—Hongmei Jiang (Northwestern University)

Monday, June 21 3:40 p.m.–5:30 p.m. | **Cosmopolitan C**

Session M36 **Population/Sample Markov chain Monte Carlo Methods—Invited**

Organizer(s): Faming Liang (Texas A&M University), **Chair(s)**: Chuanhai Liu (Purdue University)

Efficient P-value Evaluation for Resampling-based Tests—Kai Yu (National Cancer Institute)

Decomposition of Multimodal Distributions via Multi-Domain Sampling—Qing Zhou (UCLA Department of Statistics)

Bayesian Analysis of Geostatistical Models with an Auxiliary Lattice—Jincheol Park (Texas A&M University)

A Monte Carlo Metropolis-Hastings Algorithm for Sampling from Distributions with Intractable Normalizing Constants—Faming Liang (Texas A&M University)

Monday, June 21 3:40 p.m.–5:30 p.m. | **Cosmopolitan D**

Session M37 **Non/semi-parametric Methods for High Dimensional Data and Variable Selection—Invited**

Organizer(s): Lily Wang (University of Georgia), **Chair(s)**: Xiangrong Yin (University of Georgia)

Parametricness, Model Identifiability and Adaptation—Yuhong Yang (University of Minnesota)

Dynamical Multiple-Input, Single-Output Model of Neural Spike—Haonan Wang (Colorado State University)

Consistent Model Selection for Marginal Generalized Additive Model for Correlated Data—Lan Xue (Oregon State University)

Estimation and variable selection for generalized additive partial linear models—Lily Wang (University of Georgia)

Monday, June 21 3:40 p.m.–5:30 p.m. | **Studio One**

Session M38 **Statistical issues in clinical trials and healthcare—Contributed**

Organizer(s): Shengyan Hong (MedImmune, Inc.), **Chair(s)**: Min Zhang (Purdue University)

A Longitudinal Model for Medical Benefit-Risk Analysis, with Case Study—Jonathan D Norton (FDA/CDER)

Cutoff Based Designs for Community Intervention Studies—Michael Pennell (The Ohio State University)

An Equivalence Test for the Comparison between a Test Drug and Placebo in Human Abuse Potential Studies—Ling Chen (FDA)

BioPIER Clinical Workbench—Ji-yong Wang (BioPIER)

Comparing proportions of extremely rare events of uncertain status with applications to vaccine safety studies—Hongyuan Cao (UNC)

Tuesday, June 22

Tuesday, June 22 10:20 a.m.–12:10 p.m.	Concept AB
Session T11	**Statistical Consideration for the Design of Multi-Regional Clinical Trials—Invited**
Organizer(s): Yoko Tanaka (Eli Lilly and Company), **Chair(s)**: Yoko Tanaka (Eli Lilly and Company)	
Simultaneous Global Clinical Trials-Statistical Practice on the Evaluation of Drug Profile in Asian Studies—Masanori Takeuchi (Kitasato University) **An Optimal Adaptive Design to Address Local Regulations in Global Clinical Trials**—Xiaolong Luo (Celgene Corporation) **Points to Consider when Determining the Sample Size of a Given Region in an Oncology Multi-Regional Clinical Trial**—Sachio Ogawa (Eli Lilly Japan K.K.)	
Tuesday, June 22 **10:20 a.m.–12:10 p.m.**	**Concept CD**
Session T12	**Challenges in the development and application of Causal Inference methods—Invited**
Organizer(s): Lingling Li (Harvard University), **Chair(s)**: Xiaochun Li (IUPUI)	
Estimation of Optimal treatment Strategies from Longitudinal Data—James Robins (Harvard School of Public Health) **Double-robust adjustment for confounding in cohort and case-control studies**—Eric Joel Tchetgen (Harvard University) **Sensitivity analysis for causal inference using inverse probability weighting**—Changyu Shen (Indiana University) **Propensity-score based sensitivity analysis method for uncontrolled confounding**—Lingling Li (Harvard Pilgrim Health Care Institute)	
Tuesday, June 22 **10:20 a.m.–12:10 p.m.**	**Network**
Session T13	**Design and Analysis of Thorough QTc Clinical Studies—Invited**
Organizer(s): Yi Tsong (FDA), **Chair(s)**: Yi Tsong (FDA)	
Statistical Consideration in Testing for Assay Sensitivity in a 'thorough' QT Study—Venkat Sethuraman (Novartis Oncology) **Exposure-Response Modeling of QT Effect in 'Thorough' QT/QTc Studies**—Balakrishna Hosmane (Northern Illinois University) **Discussion**—Jie Chen (Abbott Laboratories)	
Tuesday, June 22 **10:20 a.m.–12:10 p.m.**	**Cosmopolitan C**
Session T14	**New Developments in Functional Data Analysis—Invited**
Organizer(s): Yehua Li (University of Georgia), **Chair(s)**: Yehua Li (University of Georgia)	
Beyond Functional Linear Regression—Hans G. Müller (UC Davis) **Longitudinal Functional Principal Component Analysis**—Ciprian M. Crainiceanu (Johns Hopkins University Biostatistics) **Functional regression for general exponential families**—Harrison H. Zhou (Yale University) **Reduced rank models for spatially correlated functional data**—Lan Zhou (Texas A&M University)	
Tuesday, June 22 **10:20 a.m.–12:10 p.m.**	**Regency A**
Session T15	**Recent advances and statistical challenges in genetical genomics**

analysis—Invited

Organizer(s): Yuehua Cui (Michigan State University), **Chair(s)**: Yuehua Cui (Michigan State University)

Studying Co-regulation and Inter-regulation of Genes via eQTL Mapping—Tian Zheng (Columbia University)
Statistical methods for RNA-seq data analysis—Wei Sun (University of North Carolina)
expression QTL mapping with sparse partial least squares—Hyonho Chun (Yale University)
Indentifying novel pathway regulation in eQTL mapping—Shaoyu Li (Michigan State University)

Tuesday, June 22 10:20 a.m.–12:10 p.m. | **Studio One**

Session T16 **Advances in Bayesian Methods applied to survey data—Invited**

Organizer(s): Dawei Xie (University of Pennsylvania), **Chair(s)**: Rong Liu (University of Toledo)

Small area estimation by combining two surveys with special consideration of cell only households—Dawei Xie (University of Pennsylvania)
Assessing Geographical Variations in Hospital Processes of Care Using Multilevel Item Response Models—Yulei He (Harvard Medical School)
Indentifying implausible gestational ages in preterm babies with Bayesian mixture models—Guangyu Zhang (University of Maryland, College Park)
Novel Bayesian multivariate hierarchical models of random coefficients of regression models for survey data—James O'Malley (Harvard Medical School)

Tuesday, June 22 10:20 a.m.–12:10 p.m. | **Regency B**

Session T17 **Large-scale multiple testing: new data, new opportunity, and new challenges—Invited**

Organizer(s): Wenguang Sun (North Carolina State Univ), **Chair(s)**: Wenguang Sun (North Carolina State Univ)

Testing for Heterosis in Gene Expression—Dan Nettleton (Iowa State University)
Classes of Multiple Decision Functions Strongly Controlling FWER and FDR—Edsel Pena (University of South Carolina)
Make No Mistake: Stepwise Multivariate Permutation Tests May Not Control Multiple Testing Error Rates—Eloise Kaizar (Ohio State University)
Adaptive Multiple Testing Procedures under Dependence—Wenge Guo (New Jersey Institute of Technology)

Tuesday, June 22 10:20 a.m.–12:10 p.m. | **Cosmopolitan D**

Session T18 **Recent advances in statistical machine learning—Invited**

Organizer(s): Junhui Wang (University of Illinois at Chicago), **Chair(s)**: Junhui Wang (University of Illinois at Chicago)

Adaptive Lasso for High-Dimensional Models Under A Class of Convex Loss Functions—Yuan Jiang (Yale University)
Multicategory Composite Least Squares Classifiers—Yufeng Liu (University of North Carolina)
An ordinary differential equation based solution path algorithm—Yichao Wu (NCSU)

Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Regency A**
Session T21	**Adaptive Clinical Trial Designs: Methods and Applications—Invited**
Organizer(s): Yanping Wang (Eli Lilly and Company), **Chair(s)**: Luping Zhao (Eli Lilly and Company)	
A Sample of Adaptive Dose-Finding Case Studies—Brenda Gaydos (Eli Lilly and Company) **Adaptive Sample Size Re-estimation in Randomized Clinical Trials—**Cyrus Mehta (Cytel Inc) **Bayesian Adaptive Trials in Practice—**Jason Connor (Berry Consultants) **Roles of Adaptive Design in Clinical Program—**Jim Hung (FDA)	
Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Concept AB**
Session T22	**Recent Developments in Dynamic Treatment Regime—Invited**
Organizer(s): Qi Long,Peng Zhang (Emory University), **Chair(s)**: Peng Zhang (University of Michigan)	
Statistical Inference In Dynamic Treatment Regimes via Q-Learning—Eric Laber (University of Michigan) **Statistical inference for treatment strategies from two-stage randomization designs when second randomization is delayed—**Abdus Wahed (University of Pittsburgh) **Support Vector Quantile Regression—**Yair Goldberg (University of North Carolina at Chap) **Discussion—**James Robins (Harvard University)	
Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Regency B**
Session T23	**Bayesian Models and Trial Simulations in Translational Pharmacometrics—Invited**
Organizer(s): Menggang Yu (Indiana University), **Chair(s)**: Lang Li (Indiana University)	
Predicting Drug-Drug Interaction—Menggang Yu (Indiana University) **A Novel Bayesian Markov Chain Monte Carlo (MCMC) Method for Nonlinear Pharmacokinetics Models—**Seongho Kim (University of Louisville) **Population pharmacokinetics as a means of capturing exposure magnitude and consistency—**Robert Bies (Indiana University School of Medicine) **A Personalized Drug Exposure Predictive Model Framework—**Lang Li (Indiana University School of Medicine)	
Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Concept CD**
Session T24	**Kernel density estimation, Wavelet, and some other Nonparametric estimation in time series, semimartingales, and Ultra-High Dimensional Additive Models.—Invited**
Organizer(s): Fuxia Cheng (Illinois State University), **Chair(s)**: Fuxia Cheng (Illinois State University)	
On the jump activity index for semimartingales—Bingyi Jing (Hong Kong Univ. of Science & Technology) **Rate-Optimal Wavelet Estimation of Mean Regression with Long Memory Infinite Moving Average Errors—**Linyuan Li (University of New Hampshire) **Asymptotic Properties for Lp-Norms of Error Density Estimators in Nonlinear Autoregressive Time Series Models—**Shuxia Sun (Wright State University)	
Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Network**
Session T25	**Statistical issues in analyzing emerging genomic data—Invited**
Organizer(s): Spencer Huang (Northwestern Univ), **Chair(s)**: Pan Du (Northwestern Univ)	
The power of imputation using variants identified by the 1000 Genomes Project—Liming Liang (Harvard University)	

Model-based methods for analyzing ChIP sequencing data—Ming Hu (University of Michigan)
Benefit from Sequencing Data in the Public Domain: In Silico CNP Genotyping from SNP Genotypes—Yun Li (University of North Carolina)
Methylation detection call for whole genome methylation data—Chiang-Ching Huang (Northwestern University)

Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Cosmopolitan C**

Session T26 **Foundations of Statistical Inference Reconsidered—Invited**

Organizer(s): Chuanhai Liu (Purdue University), **Chair(s)**: Faming Liang (Texas A&M University)

On likelihood and probabilistic inference without priors—Ryan Martin (IUPUI)
A Weak Belief Approach to One Sample Inference—Jianchuan Zhang (Purdue University)
Large-Scale Multinomial Inference and Its Applications in Genome-Wide Association Studies—Jun Xie (Purdue University)
Statistical Inference: Reconsideration for High-Dimensional Problems—Chuanhai Liu (Purdue University)

Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Cosmopolitan D**

Session T27 **New developments of variable selection methods --- from grouping to large p small n problems—Invited**

Organizer(s): Yichao Wu (North Carolina State University), **Chair(s)**: Yingying Fan (University of Southern California)

High-Dimensional Variable Selection in Meta Analysis for Censored Data—Fei Liu (IBM)
Sparse Regularization for High Dimensional Additive Models—Ming Yuan (Georgia Tech)
Penalized orthogonal-components regression for large p small n data—Dabao Zhang (Purdue University)
Sparse multivariate regression with covariance estimation—Adam Rothman (University of Michigan)

Tuesday, June 22 **1:30 p.m.–3:20 p.m.**	**Studio One**

Session T28 **Regularization Methods in Machine Learning—Invited**

Organizer(s): Ji Zhu (University of Michigan), **Chair(s)**: Yanni Zhu (Eli Lilly and Company)

Bayesian Regularization in Quantile Regression—Nan Lin (Washington University in St. Louis)
Model selection consistency for l1-penalized M-estimators—Guilherme Rocha (Indiana University)
Regularized REML for Estimation and Selection of Fixed and Random Effects in Linear Mixed-Effects Models—SijianWang (University of Wisconsin, Madison)
Sufficient dimension reduction based on an ensemble of minimum average variance estimators—Xiangrong Yin (University of Georgia)

Tuesday, June 22 3:40 p.m.–5:30 p.m.	Concept AB
Session T31	**Current Research Topics in Small Area Estimation—Invited**
Organizer(s): Jiming Jiang (UC Davis), **Chair(s)**: Jingyi Liu (Eli Lilly and Company)	
Small Area Estimation by Shrinking Means and Variances—Taps Maiti (Michigan State University) **Rapid Response Health Surveillance and the Utility of Small Area Estimates: Methods for Estimating County-level Outcome—**Haomiao Jia (Columbia University) **Fence Method for Nonparametric Small Area Estimation—**Thuan Nguyen (Oregon Health & Science University) **Discussion—**Jiming Jiang (UC Davis)	
Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Regency A**
Session T32	**Improved analyses for drug safety: Bayesian, meta-analytical, and biomarker-based methods—Invited**
Organizer(s): Lei Shen (Eli Lilly and Company), **Chair(s)**: Wei Wang (Eli Lilly and Company)	
Bayesian approaches to handle pharmacoepidemiological data with an unmeasured confounder and response misclassification—James Stamey (Baylor University) **Truncated Robust Distance for Clinical Laboratory Safety Data Monitoring and Assessment—**Xiwu Lin (GlaxoSmithKline) **Meta-analysis for Rare Adverse Event Data from Clinical Trials—**Brenda Crowe (Eli Lilly and Company) **Discussion—**Karen L. Price (Eli Lilly and Company)	
Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Network**
Session T33	**Recent development in data collection, analysis and inference—Contributed**
Organizer(s): Shengyan Hong (MedImmune, Inc.), **Chair(s)**: Haiyan Su (Montclair State University)	
Some Computational Foundations for Online Correlated Data Analysis—Edward C. Chao (Data Numerica Institute) **Confidence Intervals for Difference of Means of Zero-Clustered Data—**Karen Rosales (Western Michigan University) **Nonparametric Derivative Estimation and Posterior Probabilities for Nanoparticle Characteristics—**Richard Charnigo (University of Kentucky) **Introduction of a New Generalized Similarity Index And Its Distribution—**Hanzhe Zheng (Merck.& Co. Inc) **The Dynamic ECME Algorithm for Accelerating the EM Algorithm—**Yunxiao He (Yale University)	
Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Concept CD**
Session T34	**Recent Advances in Functional Data Analysis—Invited**
Organizer(s): Jaroslaw Harezlak (Indiana University School of Medicine), **Chair(s)**: Jaroslaw Harezlak (Indiana University School of Medicine)	
Functional Additive Regression—Gareth James (University of Southern California) **Extensions of function-on-scalar linear regression—**Philip Reiss (New York University & Nathan Kline Inst.) **Structured penalties for functional linear models---partially empirical eigenvectors for regression—**Timothy Randolph (Fred Hutchinson Cancer Research Center) **Adolescent Blood Pressure Development: A Functional Data Analysis Perspective—**Wanzhu Tu (Indiana University School of Medicine)	

Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Regency B**
Session T35	**Statistical issues for analysis of copy number variation—Invited**
Organizer(s): Lingsong Zhang (Purdue University), **Chair(s)**: Lingsong Zhang (Purdue University)	
Copy Number Profiling Using Next-Generation DNA Sequencing with Change-Point Methods—Jeremy Shen (Stanford University) **Optimal Sparse Segment Identification with Application in Copy Number Variation Analysis—**Jessie Jeng (University of Pennsylvania) **Genomic meta-analysis for dimension reduction and gene clustering—**Goerge Tseng (Dept of Biostatistics, U of Pittsburgh) **Statistical Methods for Copy Number Alternation—**Lingsong Zhang (Purdue University)	
Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Cosmopolitan C**
Session T36	**Spatial Statistics and Computation—Invited**
Organizer(s): Hao Zhang (Purdue University), **Chair(s)**: Bo Li (Purdue University)	
On Variable Selection in Spatial Linear Regression—Jun Zhu (Colorado State University) **Statistical inference under spatial preferential sampling—**Zhengyuan Zhu (Iowa State University) **On a class of space-time intrinsic random functions—**Michael Stein (University of Chicago) **An approach to modeling asymmetric multivariate spatial covariance structure—**Bo Li (Purdue University)	
Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Cosmopolitan D**
Session T37	**New Advances in Bayesian Modeling for Bioinformatics Data—Invited**
Organizer(s): Yuan Ji (The University of Texas M. D. Anderson), **Chair(s)**: Ming-Hui Chen (University of Connecticut)	
Bayesian Inference for Genome-Wide Association Studies—Luis Carvalho (Boston University) **Bayesian base calling for Solexa sequencing data—**Yuan Ji (M. D. Anderson Cancer Center) **Bayesian Modeling of MPSS Data: Gene Expression Analysis of Bovine Salmonella Infection—**Bani Mallick (Texas A&M) **Modeling Dependent Gene Expression—**Peter Muller (M.D. Anderson Cancer Center)	
Tuesday, June 22 **3:40 p.m.–5:30 p.m.**	**Studio One**
Session T38	**Applications of Non- and Semiparametric Methods to Time Series and Longitudinal Data—Invited**
Organizer(s): Lijian Yang (Michigan State University), **Chair(s)**: Fuxia Cheng (Illinois State University)	
Mixing density estimation using the nonparametric penalized likelihood maximization—Michael Levine (Purdue University) **Spline estimation of a semiparametric GARCH model—**Rong Liu (University of Toledo) **A simultaneous confidence band for sparse longitudinal regression—**Shujie Ma (Michigan State University) **Discussion—**Lily Wang (University of Georgia)	

Wednesday, June 23

Wednesday, June 23 8:00 a.m.–9:50 a.m.	Regency A
Session W11	**Methodological and Software Development for Bayesian Clinical Trials Designs—Invited**
Organizer(s): Ming-Hui Chen, J. Jake Lee (University of Connecticut), **Chair(s)**: Peter Lam (Boston Scientific Corporation)	
Predictive Probability and Adaptive Randomization in Phase II Clinical Trials—J.Jack Lee (Univ. of Texas M. D. Anderson Cancer Ctr)	
Bayesian Design of Non-Inferiority Trials for Medical Devices Using Historical Data—Ming-Hui Chen (University of Connecticut)	
Aspects of Development of Statistical Software in the Design and Execution of Adaptive Bayesian Clin—Ashish Sanil (Berry Consultants)	
Case Study of a Bayesian Clinical Study with a Hierarchical Prior in a Medical Device Trial—Scott Wehrenberg (Boston Scientific Corporation)	
Wednesday, June 23 8:00 a.m.–9:50 a.m.	**Regency B**
Session W12	**Multiplicity issues in clinical trials with multiple endpoints—Invited**
Organizer(s): Alex Dmitrienko (Eli Lilly and Company), **Chair(s)**: Yan D. Zhao (UT Southwestern Medical Center)	
Multiple testing problems with general logical restrictions in clinical trials—Alex Dmitrienko (Eli Lilly and Company)	
Statistical Give and Take: Power Implications on Strong Control of the Type I Error Rate in a Clinical Trial Setting—Chris Holland (MacroGenics)	
Hommel-based gatekeeping multiple comparison procedure: an example—Jane Xu (Sepracor)	
A general approach for testing a prespecifeid subgroup in clinical trials—Mohamed Alosh (FDA)	
Wednesday, June 23 8:00 a.m.–9:50 a.m.	**Concept AB**
Session W13	**Model based analysis—Contributed**
Organizer(s): Shengyan Hong (MedImmune, Inc.), **Chair(s)**: Jie Yang (University of Illinois at Chicago)	
Modeling HIV viral dynamic model with longitudinal data—Jianwei Chen (San Diego Stae University)	
Adaptive Fitting of Linear Mixed-Effects Models with Correlated Random-effects—Guangxiang Zhang (Stony Brook University)	
Sieve Estimation in Nonlinear Ordinary Differential Equation Models—Hongqi Xue (University of Rochester)	
Combination of Confidence Distributions and an Efficient Approach for Meta-Analysis of Heterogeneous Studies—Dungang Liu (Rutgers University)	
Wednesday, June 23 8:00 a.m.–9:50 a.m.	**Concept CD**
Session W14	**Design of Experiment—Contributed**
Organizer(s): Shengyan Hong (MedImmune, Inc.), **Chair(s)**: Eloise Kaizar (The Ohio State University)	
Randomization Inference for the Trimmed Mean of Effects Attributable to Treatment—Yang Feng (UIUC)	
Sample size re-estimation in two-stage design using p-value combination tests—Shanhong Guan (Merck & Co., Corp.)	
Optimal Design for Two-Level Factorial Experiments with Binary Response—Jie Yang (University of Illinois at Chicago)	

Wednesday, June 23 8:00 a.m.–9:50 a.m.	Cosmopolitan C
Session W15	**Spatial Statistics for Environmental Sciences—Invited**
Organizer(s): Bo Li (Purdue University), **Chair(s)**: Bo Li (Purdue University)	
Fixed-domain Asymptotic Properties of Tapered Maximum Likelihood Estimators—Juan Du (Kansas State University) **Nonparametric variogram estimation on the sphere—**Chunfeng Huang (Indiana University) **Parameter Estimation for Large Gridded Spatial Datasets—**Jonathan Stroud (George Washington University) **Reconstructing Land Cover Change Trajectories from Time Series Satellite Observations—**Desheng Liu (The Ohio State University)	
Wednesday, June 23 **8:00 a.m.–9:50 a.m.**	**Network**
Session W16	**High-dimensional data analysis in Cancer Research—Invited**
Organizer(s): Xiaochun Li and Ronghui Xu (IUPUI), **Chair(s)**: Xiaochun Li (IUPUI)	
Secondary Analysis of Case-control Data and Application in Genetic Association Studies—Huilin Li (NCI) **Interactions in Association Models—**Mike LeBlanc (Fred Hutchinson Cancer Research Center) **Assessing and Accounting for Dependence in Gene Expression Summaries—**John Stevens (Utah State University) **Oligoarray data analysis incorporating probe sequence thermodynamics model—**Wenjiang Fu (Michigan State University)	
Wednesday, June 23 **8:00 a.m.–9:50 a.m.**	**Cosmopolitan D**
Session W17	**Current Topics in Multivariate Outcomes with Survival Endpoints—Invited**
Organizer(s): Ronghui (Lily) Xu (Univ of California, San Diego), **Chair(s)**: Ronghui (Lily) Xu (Univ of California, San Diego)	
Analysis of Survival Data With High Dimensional Predictors—Yi Li (Harvard School of Public Health) **Bivariate Analysis of Competing Risks Data—**Jong-Hyeon Jeong (University of Pittsburgh) **The relative efficiency of time-to-event and longitudinal modeling—**Michael Donohue (University of California, San Diego) **Estimation of Random Effects in Semiparametric Frailty Models—**Il Do Ha (Daegu Haany University)	
Wednesday, June 23 **8:00 a.m.–9:50 a.m.**	**Studio One**
Session W18	**Censored or Missing Data Analysis—Contributed**
Organizer(s): Shengyan Hong (MedImmune, Inc), **Chair(s)**: Richard Charmigo (University of Kentucky)	
A simple diagnostic for proportional hazards and the logrank test—Joshua Naranjo (Western Michigan University) **Semiparametric hybrid empirical likelihood inference for two sample comparison with censored data—**Haiyan Su (Montclair State University) **A proof for the underestimation of the Greenwood's type of estimation—**Jiantian Wang (Kean University) **Mann-Whitney Test for Two Comparing Waiting Time Distributions when Transition Times are Right Censored—**Jie Fan (University of Louisville)	

Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Regenecy A**
Session W21	**Student Award Session—Invited**
Organizer(s): Rick Chappell (University of Wisconsin-Madison), **Chair(s)**: Rick Chappell (University of Wisconsin-Madison)	
Pairwise variable selection for classification—Xingye Qiao (UNC Chapel Hill) **Principled sure independence screening for Cox models with ultra-high-dimensional covariates—**Sihai Dave Zhao (Harvard Univ. Dept. of Biostatistics) **Statistical Inference in Factor Analysis for High-Dimensional, Low-Sample Size Data—**Miguel Marino (Harvard Univ. Dept. of Biostatistics)	
Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Concept AB**
Session W22	**Phase 2 Trial Design: Breaking the alpha=0.05, two-sided mind-set—Invited**
Organizer(s): Ouhong Wang (Amgen), **Chair(s)**: Brian Smith (Amgen)	
Biomarker-based adaptive dose-finding trials: A case study in Phase II Oncology—Chyi-Hung Hsu (Novartis Pharmaceuticals) **Optimal Cost-effective Designs for Phase II Proof of Concept Trials—**Yang Song (Merck Research Labs) **A Seamless 2/3 Design Incorporating a Clinical Utility Index—**Zachary Skrivanek (Eli Lilly and Company) **Modeling analysis of longitudinal time-couse data in clinical trials can significantly improve efficiency—**Quan Hong (Eli Lilly and Company)	
Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Regency B**
Session W23	**Special topic: Placebo Response - Recent Trend in Psychiatric Clinical Trials, Design and Analysis—Invited**
Organizer(s): Lu Zhang (Eli Lilly and Company), **Chair(s)**: Lu Zhang (Eli Lilly and Company)	
Observational and Causal Components of Placebo Response—Craig Mallinckrodt (Eli Lilly and Company) **Optimal Partitioning for Linear Mixed Effects Models: Applications to Identifying Placebo Responders—**Thaddeus Tarpey (Wright State University) **Placebo Response and the Sequential Parallel Design - Statistical Considerations—**Roy Tamura (Eli Lilly and Company) **Placebo Response in Major Depression and Schizophrenia Trials: Statistical Consideration and Design Strategies—**Yeh-Fong Chen (FDA)	
Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Network**
Session W24	**Bridging genetics and medicine: statistical challenges—Invited**
Organizer(s): Zhaoling Meng (Sanofi-aventis), **Chair(s)**: Yujun Wu (Sanofi-aventis)	
Risk modeling in personalized medicine—Hongyu Zhao (Yale University) **A statistical model for mapping morphological shapes—**Rongling Wu (Pennsylvania State University) **Tailoring Therapies for Complex Diseases based on Multiple Genotypic Markers—**Yanni Zhu (Eli Lilly and Company) **Discussion—**Yujun Wu (Sanofi-aventis)	
Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Concept CD**
Session W25	**New Frontiers in Empirical Likelihood—Invited**
Organizer(s): Yichuan Zhao (Georgia State University), **Chair(s)**: Hanxiang Peng (Indiana University-Purdue University at Indianapolis)	

Adjusted empirical likelihood—Jiahua Chen (Department of Statistics, UBC, Canada)

Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood—Yanqing Sun (The University of North Carolina at Char)

Pseudo Empirical Likelihood Inference for Multiple Frame Surveys—Changbao Wu (University of Waterloo)

On the Use of Empirical Likelihood for the Analysis of Longitudinal Data—Nicole Lazar (University of Georgia)

Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Studio One**
Session W26	**Statistics in Environmental, Financial and Social Sciences—Contributed**

Organizer(s): Shengyan Hong (MedImmune, Inc.), **Chair(s)**: Guangxiang Zhang (Stony Brook University)

Developing A New BIC for Detecting Change-points—Shen, Gang (North Dakota State University)

Is there evidence for the high frequency data being purely discontinuous?—Jing Bing-Yi, Kong (HKUST)

A Joint Statistical Model of Social Behavior and Social Network—Liping Tong (Loyola University Chicago)

Seasonality Analysis of Time Series in Partial Linear Models—Qin Shao (The University of Toledo)

Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Cosmopolitan C**
Session W27	**Incorporating/combining indirect evidence to statistical analysis—Invited**

Organizer(s): Minge Xie (Rutgers University), **Chair(s)**: Haoda Fu (Eli Lilly and Company)

Penalized Maximum Likelihood Estimation for Stationary Time Series—Xiaodong Lin (Rutgers University)

Variable Selection with Prior Information via the pLasso Method—Yunxiao He (Yale University)

Network Meta Analysis Design Adapted to a Historically Controlled Clinical Trial—A James O'malley (Harvard Medical School)

Wednesday, June 23 10:10 a.m.–12:00 p.m.	**Cosmopolitan D**
Session W28	**Dimension Reduction and Variable Selection under High Dimensional Semiparametric Models—Invited**

Organizer(s): Michael Zhu (Purdue University), **Chair(s)**: Guang Cheng (Purdue University)

Sufficient Dimension Reduction, Prediction and Variable Screening in High Dimensional Regressions.—Dennis Cook (University of Minnesota)

Groupwise Dimension Reduction—Bing Li (Pennsylvania State University)

SIM-Lasso for estimation and variable selection in single-index models—Peng Zeng (Auburn University)

ANNOUNCEMENTS

J.P. Hsu Student Paper Award Winner

- **Ni Li**, Department of Statistics, University of Missouri, *Semiparametric Transformation Models for Panel Count Data with Dependent Observation Process*

Student Paper Awards Winners

(by last name)

- **Xu He**, Univeristy of Wisconsin-Madison, *Subgroups defined by two or more genetic markers in genome-wide association studies*
- **Miguel Marino**, Harvard School of Public Health, *Statistical Inference in Factor Analysis for High-Dimensional, Low-Sample Size Data: Understanding the Change Patterns of the US Cancer Mortality Rates*
- **Xingye Qiao**, Dept. of Statistics and Operations Research, University of North Carolina, Chapel Hill, *Pairwise Variable Selection for Classification*
- **Sihai Dave Zhao**, Dept. of Biostatistics, Harvard School of Public Health, *Principled sure independence screening for Cox models with ultra-high-dimensional covariates*

SOCIAL EVENTS

Mixer

Mixer and Welcome Reception will be held at Hyatt Regency from 6:30 p.m. to 9:00 p.m. on Sunday, June 20, 2010. This is a free event to welcome all symposium participants.

ICSA 2010 SYMPOSIUM BANQUET

(Fee Event: $40/person)

Evening Program

Time: 6:30 p.m. – 9:00 p.m.

Date: Tuesday, June 22, 2010

Place: On Time Seafood Restaurant

3623 Commerical Dr. Indianapolis, IN 46222, (317)293-8888

Please take the shuttle bus at 5:35 in front of the hotel

- Authentic Chinese Dinner
- Meals will be served at 6:30 p.m.
- Banquet Keynote Address at 7:00 p.m.
  - Introduction of the Speaker by Wei Shen, Research Advisor, Eli Lilly and Company
  - Speaker: Dr. William W. Chin, Executive Dean, Harvard Medical School, Harvard University

    Title: Statistics: a critical pharma--academic interface
- Acknowledgement: Yongming Qu, Principle Research Scientist of Eli Lilly and Company
- Raffle and Entertainment at 8:30 p.m. (Raffle: Dozens of Fun Gifts).

SHORT COURSES

Short Course #1: **Recent Developments in Practical Bayesian Methods for Clinical Trials**

Instructor: Dr. Peter F. Thall, M.D. Anderson Cancer Center

Location and Time: Network Room, Sunday, June 20, 2010, 9:00 a.m.– 4:30 p.m.

Abstract: This one-day short course will cover a variety of practical Bayesian methods for clinical trial design and conduct. Some frequentist methods also will be discussed. The course will include numerous illustrations from actual clinical trials. Emphasis will be on newer methods that I have developed with colleagues in recent years. As time permits, the topics will include (1) designs that deal with multiplicities and heterogeneity in early phase trials, including individualized dose-finding in phase I/II, (2) using elicited utilities as a basis for trial design and conduct, (3) monitoring multiple outcomes, monitoring possibly right-censored event times and accounting for patient heterogeneity in phase II trials, (4) a method for computing the effective sample size of a parametric prior, including both standard Bayesian models and conditionally independent hierarchical models, (5) a phase III group sequential design that uses Bayesian model selection and controls overall frequentist error probabilities (6) designs to evaluate and compare multi-stage dynamic treatment regimes, rather than individual treatments, (7) geometric methods for settings with two-dimensional parameters, including a cord blood cell transplantation trial and a phase II-III trial of chemotherapies for pediatric brain tumors, (8) a phase II-III design of prophylactic agents for atrial fibrillation following lung surgery that accounts for patient heterogeneity and (9) adaptive randomization. Many of the illustrative applications will include examples of prior elicitation and calibration, incorporating historical data, and using computer simulation to establish a design's frequentist properties. Attendees should have at least a Masters degree in statistics, or equivalent experience, and an understanding of clinical trials and elementary Bayesian concepts.

About the Instructor: **Dr. Peter F. Thall** is the Anise J. Sorrell Professor in the Department of Biostatistics at M.D. Anderson Cancer Center. Dr. Thall is an author of over 160 papers and book chapters in the statistical and medical literature, with research interests including Bayesian statistics, medical statistics and clinical trials. He has presented 21 short courses on statistical methods for clinical trials and over 130 invited talks. He has served as an associate editor of Journal of the National Cancer Institute, Statistics in Medicine and

Biometrics, and currently is an associate editor of Clinical Trials and Statistics in Biosciences, serves on several external advisory boards and grant review panels, and is an American Statistical Association Media Expert.

Short Course #2: **Modern Techniques in Data Mining**

Instructor: Dr. David Banks, Professor in the Department of Statistical Science at Duke University

Location and Time: Studio Five Room, Sunday, June 20, 2010, 9:00 a.m.– 4:30 p.m.

Abstract: This one-day short course surveys the development of an important new subfield at the intersection of statistics and computer science. It has important applications in many areas, especially bioinformatics and information technology. The course starts from the perspective of nonparametric regression, addressing the problems of variable selection, local fitting, model assessment and uncertainty, all in the context of the Curse of Dimensionality. Then the course moves to consider comparable issues in the case of classification, cluster analysis, and multidimensional scaling. Finally, the course gives a short summary of three areas of emerging theory that provide insight on when it is possible to make inference in high dimensions, and when there is no really no hope. Key topics include support vector machines, neural networks, boosting and bagging, random forests, overcompleteness, wavelets, sparsity, VC classes, and the LASSO and related methods.

About the Instructor: **Dr. David Banks** is a professor in the Department of Statistical Science at Duke University. Before that, he held positions in the Food and Drug Administration, the Department of Transportation, the National Institute of Standards and Technology, and Carnegie Mellon University. He was editor of the Applications and Case Studies section of the Journal of the American Statistical Association (2007-2009), serves on the Board of the American Statistical Association, and has chaired the ASA Sections on Risk Analysis and Statistics in Defense and National Security. His research focuses on computer-intensive methods, complex data sets, dynamic network models, metabolomics, public policy, and adversarial risk analysis. He has published more than 60 papers, edited six books, and co-authored a monograph.

Short Course #3: **An Introduction to Propensity Score Methods in Observational Research**

Instructor: Dr. Peter Austin, Sr. Scientist at the Institute for Clinical Evaluative Sciences (ICES) in Toronto, Canada

Location and Time: Concept AB Room, Sunday, June 20, 2010, 8:00 a.m.–12:00 p.m.

Abstract: This half-day workshop examines the use of propensity score methods in observational research. Confounding frequently occurs in observational studies of the effects of treatments and exposures on health outcomes. This workshop will address several statistical issues in estimating treatment effects in the presence of confounding. First, a theoretical framework for confounding will be developed and causal diagrams will be introduced. Second, the design of randomized controlled trials (RCTs) will be briefly reviewed. Third, issues in the design of observational studies will be highlighted. We will discuss designing observational studies so that their design mimics some of the characteristics of that of RCTs. Fourth, we will describe design-based and analysis-based methods for removing confounding when estimating treatment effects using observational data. Analysis-based methods include regression-based approaches. Design-based approaches include stratification and propensity score-based methods. Fifth, attendees will be introduced to the concept of the propensity score and how it can be used to remove confounding in observational studies. Issues in propensity score analyses will be covered in more depth. The workshop will cover the following issues in propensity score analyses: specifying the propensity score model; balance diagnostics for assessing the adequacy of the specification of the propensity score model; different propensity score methods (matching, stratification, weighting, and covariate adjustment); and sensitivity analyses for propensity score analyses.

About the Instructor: **Dr. Peter Austin** is a Senior Scientist at the Institute for Clinical Evaluative Sciences (ICES) in Toronto, Canada and an Associate Professor in the Department of Health Policy, Management and Evaluation at the University of Toronto. His research interests include propensity score methods for causal inference, predictive models for cardiovascular outcomes, statistical methods for provider profiling, and applied Bayesian methods in health services research. He has published actively on propensity score methods and frequently collaborates with investigators in health services research, pharmacoepidemiology, and clinical epidemiology.

Short Course #4: **Post-Marketing Drug Safety Evaluation**

Instructor: Dr. A. Lawrence Gould, Merck Research Laboratories

Location and Time: Concept CD Room, Sunday, June 20, 2010, 8:00 a.m.–12:00 p.m.

Abstract: Surveillance of drug products in the marketplace continues after regulatory approval for a variety of reasons, e.g., to identify rare potential toxicities that are unlikely to have

been observed in the clinical trials carried out before approval. Conventional statistically based postmarketing surveillance traditionally has focused on the large number of spontaneous reports of adverse events in spontaneous report databases. There currently is considerable interest and effort being directed towards using large observational databases containing insurance claim information or routinely maintained electronic health records. Determining which drug-event associations, of which there may be many tens of thousands, are real reporting associations and which random noise presents a substantial problem of multiplicity because the resources available for medical and epidemiologic followup are limited. This presentation will address some of the issues associated with using these sources for postmarketing safety surveillance, and will describe and compare methods for identifying potential 'signals' from spontaneous reporting databases.

About the Instructor: **Dr. A. Lawrence Gould** received his PhD degree in Biometry from Case Western Reserve University (1967). He is currently a Senior Director at Merck Research Laboratory. He was elected as Fellow of the American Statistical Association in 1988. He has served as Fellows Committee Chair and Publications Officer of the Biopharmaceutical Section. Member of Biometric Society ENAR, served as Secretary/Treasurer 1982-1986. Served as Editor of Journal of Biopharmaceutical Statistics 2001-2002. His areas of research interest include use of Bayesian methods to improve effectiveness of the drug development process, adaptive trial design (including group sequential methods), evaluation of safety data from clinical trials, application of data mining and Bayesian methods to pharmacovigilance, use of data mining to identify relationships that can be used to design future trials, meta-analysis, modeling and simulation techniques to reduce cost and unnecessary patient exposure in drug development, and application of decision science methods to drug development strategy.

Short Course #5: **Principles and Techniques of Multiple testing and Multiple Comparisons**

Instructor: Dr. Jason C. Hsu, The Ohio State University

Location and Time: Concept AB Room, Sunday, June 20, 2010, 1:30 p.m.–5:30 p.m.

Abstract: This half-day short course is about fundamental concepts and techniques of multiple testing, in clinical trials and bioinformatics. We will discuss Familywise Error Rate (FWER), generalized Familywise Error Rate (gFWER), and various versions of False Discovery Rate (FDR, Fdr). Issues to consider in error rate control include

- True, average, or worst case scenario (to sup or not to sup)

- Number or proportion of incorrect rejections
- Conditional or unconditional
- Tail probability or expectation.

We will describe the multiple test construction techniques of closed testing and partition testing. Familiar methods such as Holm's and Hochberg's step-wise tests turn out to be special cases of partition testing. Using clinical trial with Multiple Endpoints as an illustration, we will show Partition Testing requires drastically fewer tests than Gatekeeping, and is more powerful. The rush to meet the challenge of Bioinformatics seems to have occasionally overlooked fundamental principles of multiple testing. Using testing for association between biomarkers and drug response or adverse events as an example, we will show that common permutation tests do not control multiple testing error rate (unless assumptions are made on the joint distributions of gene expression levels or SNP alleles between phenotypes). We will illustrate the application of fundamental principle and techniques in multiple primary-secondary endpoints efficacy testing and genome-wide association studies (GWAS). The objective of this course is to help the participant decide on an error rate to logically control, and to confidently control it.

About the Instructor: **Dr. Jason C. Hsu** is a professor in the Department of Statistics at the Ohio State University. He works in the area of multiple testing and multiple comparisons. Since 1998, the approach he has emphasized is to connect methodological development with emerging biomedical issues. Besides fundamental concepts and techniques, his current interests include analysis of multiple endpoints data, and pharmacogenomics.

Short Course #6: **Pharmacogenomics Clinical Trials: Genomic Biomarker Associated Design and Analysis Issues**

Instructor: Dr. Sue-Jane Wang, U.S. Food and Drug Administration

Location and Time: Concept CD Room, Sunday, June 20, 2010, 1:30 p.m.–5:30 p.m.

Abstract: In recent years, early phase clinical studies have begun to incorporate mRNA microarrays or whole genome DNA scanning high throughput biotechnologies as a means to explore the potential genomic associations between high dimensional genomic data and clinical outcome. There are high expectations on the use of such biotechnologies to develop and validate genomic composite biomarker with prognostic, diagnostic screening, and predictive potential for drug treatment in complex disease area, such as psychiatric, cardio-renal diseases or life-threatening diseases, such as AIDS and oncology. This half-day short course will provide an overview of many considerations and challenges in incorporating pharmacogenomics in a

clinical drug development program ranging from development of a genomic composite biomarker to implementation of personalized medicine. The topics will include statistical concepts in pharmacogenomics exploratory studies and evaluation of confirmatory pharmacogenomics clinical trials. The course outlines include

- Development of a genomic composite biomarker
- Multiplicity issues on false discovery
- Establishment and validation of a genomic predictive model
- Genomewide association study
- Clinical utility of genomic biomarker and biomarker qualification
- Diagnostics performance characteristics assessment: companion or co-developed
- Convenience samples and confounding issues
- Evaluation of study designs and analysis methods in confirmatory PG trials
- Probability of imbalance and design issues with biomarker negative patient subset
- Reproducibility issues.

The course will include clinical trial examples mimicking new drug application submissions. Literature overview and the instructor's collaborative research will be introduced. Attendees should have at least a Masters degree in statistics, or equivalent experience, and an understanding of clinical trials.

About the Instructor: **Dr. Sue-Jane Wang** is Office Associate Director for Pharmacogenomics and Adaptive Design in the Office of Biostatistics under Office of Translational Sciences, Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration. She is an elected member of the International Statistical Institute. She has served as an Editor-in-Chief of Pharmaceutical Statistics, and currently is an Associate Editor of Statistics in Medicine and Statistics BioSciences. Based on her collaborative research publications, Dr.Wang had given short courses on adaptive design, pharmacogenomics, multi-regional clinical trials, non-inferiority and bioinformatics, and has served numerous (co)chair, discussant, and keynote roles in addition to invited talks at the professional meetings.

KEYNOTE SPEECHES

Speech K01-A: **FDA's Critical Path Initiative: Updates, Opportunities, and New Challenges**

Speaker: Dr. ShaAvhrée Buckman, Food and Drug Administration

Location and Time: Regency AB, Monday, June 21, 8:00 a.m.– 10:00 a.m.

Abstract: In March 2004, the FDA released a report entitled "Innovation or Stagnation, Challenge and Opportunities on the Critical Path to New Medical Products ". This report highlighted the rising cost of drug development coupled with the decline in new drug submissions to the FDA. It emphasized the urgent need to modernize the medical product development process to keep pace with scientific innovation. In 2006, FDA published 76 specific scientific activities in a 2nd report, entitled "Critical Path Opportunities Report and List ". The Critical Path Initiative (CPI) is FDA's effort to stimulate and facilitate a national effort to modernize the scientific process through which a potential human drug, biological product, or medical device is transformed from discovery or "proof of concept" into a medical product. The goal of the CPI is to bring new scientific discoveries to bear on product development, and to improve the accuracy of the tests we use to predict the safety and efficacy of investigational medical products. CPI research projects focus on six key areas:

- Better Evaluation Tools -- Biomarkers and Disease Models
- Streamlining Clinical Trials
- Harnessing Bioinformatics
- Moving Manufacturing into the 21st Century
- Products to Address Urgent Public Health Needs
- At-Risk Populations.

This talk is focused on approaches which promote automation of data collection, genomic data collection, enhanced study design techniques including adaptive designs, incorporation of pharmacogenomic information in trials and labeling, partnering of industry and academia, educational programs to teach and train future investigators, development of disease models and biomarkers in clinical trials, and use of patient populations in early clinical trials.

All of these areas are critical to modernize medical product development.

About the Speaker: **Dr. ShaAvhrée Buckman** is the Director at the Center for Drug Evaluation and Research, Food & Drug Administration. She received her Ph.D. in Molecular Cell Biology from

Washington University in 1999. Her current reserach interset is in clinical trial design strategies and novel approaches (pharmacogenomics, biomarker development, disease modeling, bioinformatics) to stimulate efficient drug development.

Speech K01-B: **Statisticians in the Pharmaceutical Industry: The 21st Century**

Speaker: Dr. Christy Chuang-Stein, Pfizer, Inc.

Location and Time: Regency AB, Monday, June 21, 8:00 a.m.– 10:00 a.m.

Abstract: The 21st century brings significant challenges and along with them many opportunities for statisticians working in the pharmaceutical industry. The public is expecting the industry to develop breakthrough life-saving medicines, safe and affordable. Cutting edge technology has helped fuel information explosion, yet much of the promise from new technology remains un-fulfilled. Business model has been evolving constantly since the 90's, creating an operational structure that often decentralizes the statistical function within an organization. Changes brought on by mergers and acquisitions have led to unplanned career changes for some in our industry. Yet, in the midst of the energy and opportunities created by this dynamic environment also lie unprecedented opportunities for many. The pursuit of innovations has resulted in closer collaborations among statisticians across companies. Common business objectives have led the formation of alliances for shared adventures. In this talk, we will look at the core competencies of statisticians in the 21st century. We will examine how statisticians can help further shape a highly quantitative drug development process, better articulation of the benefit and risk balance, closer industry-academia-government collaborations, and a more agile business model. Our worth in this competitive world comes from the value we bring to the table. We need to promote statistical excellence and champion statistical influence in our industry. Equally important, we need to advocate transparent decisions to help bring integrity and trust back to an industry that had made tremendous contributions to the medical evolutions in the 20th century.

About the Speaker: **Dr. Christy Chuang-Stein** , a Fellow of the American Statistical Association since 1998, is Head of the Statistical Research and Consulting Center at Pfizer. Christy has 25 years of pharmaceutical industry experience. She serves on the editorial boards of Pharmaceutical Statistics and Drug Information Journal (DIJ). Christy is a Vice President of the American Statistical Association and is a member of many industry-wide collaborative working groups. Christy has over 115 publications in peer-reviewed journals and book chapters.

Speech K01-C: **The future of statistics and the future of statisticians - an idiosyncratic assessment**

Speaker: Dr. Donald Rubin, Harvard University

Location and Time: Regency AB, Monday, June 21, 8:00 a.m. – 10:00 a.m.

Abstract: In my experience, when asking relatively successful people in any field, "What is the future of that field and the future of workers in it?" the almost inevitable reply is: "The future is very bright, especially if it continues to train more people like me!" In other words, "From my perspective, to be more successful, you should be more like me!!" Of course, in many ways this response is simply a statement about what has appeared to have created success for the one individual, but with enough such comments from enough senior individuals, younger individuals may be able to create a sensible perspective. In this short presentation, I will not deviate from this idiosyncratic approach, and will rely on the combination with other presentations to create a more objective perspective.

About the Speaker: **Dr. Donald Rubin** is John L. Loeb Professor of Statistics, Department of Statistics, Harvard University, where he has served as chairman for 13 of his 25 years there. He has over 350 publications (including several books) on a variety of topics, including computational methods, causal inference in experiments and observational studies, survey methods, techniques for handling missing data, Bayesian methods, multiple imputation, matched sampling, and applications in many areas of social and biomedical science. Professor Rubin is a Fellow of the American Statistical Association, the Institute for Mathematical Statistics, the International Statistical Institute, the Woodrow Wilson Society, the John Simon Guggenheim Society, the New York Academy of Sciences, the American Association for the Advancement of Sciences, the American Academy of Arts and Sciences, and the Alexander von Humboldt Foundation. He is also the recipient of four of the most prestigious awards available to statisticians: the Samuel S. Wilks Medal of the American Statistical Association, the Parzen Prize for Statistical Innovation, the Fisher Lectureship, and George W. Snedecor Award of the Committee of Presidents of Statistical Societies. Professor Rubin has lectured extensively throughout The Americas, Europe, and Asia. For many years, he has been one of the most highly cited writers in mathematics in the world, according to ISI Science Watch.

Speech K02-A: **Statistical Issues and Challenges in Analyzing High-throughput 'Omics Data in Population-Based Studies**

Speaker: Dr. Xihong Lin, Harvard University

Location and Time: Regency AB, Tuesday, June 22, 8:00 a.m.– 10:00 a.m.

Abstract: With the advance of biotechnology, massive "omics" data, such as genomic and proteomic data, become rapidly available in population based studies to study interplay of genes and environment in causing human diseases. An increasing challenge is how to design such studies, managing

the data, analyze such high-throughput "omics" data, interpret the results, make the findings reproducible. We discuss several statistical issues in analysis of high-dimensional "omics" data in population based "omics" studies. We present statistical methods for analysis of several types of "omics" data, including incorporation of biological structures in analysis of data from genome-wide association studies, next generation sequencing data for rare variants, and epigenetic analysis of genome-wide DNA methylation data for genes and environment. Data examples are presented to illustrate the methods. Strategies for interdisciplinary training in statistical genetics, computational biology and genetic epidemiology will also be discussed.

About the Speaker: **Dr. Xihong Lin** is Professor of Biostatistics in the Department of Biostatistics in Harvard School of Public Health. She currently serves as the coordinating director of the Program of Quantitative Genomics of Harvard School of Public Heath. Dr. Lin is a Fellow of the American Statistical Association, the Institute for Mathematical Statistics, and the International Statistical Institute. She is also the recipient of the COPSS Presidents' Award for the outstanding statistician in 2006, the Mortimer Spiegelman Award for the outstanding biostatistician by American Public Health Association in 2002, and the Noether Young Researcher Scholar Award by American Statistical Association in 2002. Her statistical research is currently funded by the MERIT award from the National Cancer Institute. Dr. Lin has over 125 publications and has received over 30 research grants.

Speech K02-B: ***Statistical Challenges in Personalized Medicine***

Speaker: Dr. Gregory Campbell, Food and Drug Administration

Location and Time: Regency AB, Tuesday, June 22, 8:00 a.m.– 10:00 a.m.

Abstract: Since the announcement in June of 2000 of the successful sequencing of the human genome, the public has looked to science for practical breakthroughs in medicine directly from this achievement. One promise has been that of personalized medicine; namely, that medical products and procedures can be tailored to the individual patient. With the evolution of evermore sophisticated microarrays (chips with thousands of genetic segments) and the potential to bring tests based on microarrays and on single nucleotide polymorphisms (SNPs) to market, an explosion in the amount of genetic and genomic data is now taking place at an ever increasing pace. One statistical challenge is that of construction: how to discover that a particular medical product (a drug or a device) is safer or more efficacious for one individual based on the particular result of a diagnostic (probably genomic) test. Various types of diagnostic tests or biomarkers are introduced and illustrated. One instructive project that can illustrate some of the many statistical challenges involving classifier construction is the FDA-led Microarray Quality Control Project (MAQC), now in its second phase. A FDA Draft Guidance on In

Vitro Diagnostic Multivariate Index Assays (IVDMIAs) is also discussed. A second statistical challenge is that of design: how can diagnostic tests and therapeutic products be co-developed. Statistical designs that allow for adaptation or for the use of retrospective data in a scientifically valid manner are discussed. Bayesian designs are also mentioned. A third statistical challenge is the analysis of data from such trials and particularly the issues of multiplicity and of missing data. Several drug-diagnostic examples are reviewed and a number of clinical trials are discussed. The implications for the future of individualized medicine are enormous and it is clear that an interdisciplinary effort involving statistics, bioinformatics and biology will be crucial in unlocking the promise of personalized medicine.

About the Speaker: **Dr. Gregory Campbell** is the Director of the Division of Biostatistics in the Office of Surveillance and Biometrics (OSB) of Center for Devices and Radiological Health (CDRH) of the Food and Drug Administration (FDA) since he came to FDA in 1995. With a B.S. in Mathematics from the University of Dayton and M.S. in Mathematics from Michigan State University, he received a Ph.D. in Mathematical Statistics from Florida State University. After serving on the faculty in the Department of Statistics at Purdue University, he joined the National Institutes of Health, where he became a tenured intramural research scientist, Acting Chief for the Laboratory of Statistical and Mathematical Methodology, and Chief of the Analytical Biometrics Section. Dr. Campbell currently leads a group of about 55 statisticians at the FDA that provides statistical support to CDRH as a whole and, in particular, the statistical reviews of FDA's pre-market device submissions. With the help of statisticians in his Division, he pioneered the implementation in a regulatory environment of Bayesian statistics (and more recently causal inference). His current research interests include the evaluation of diagnostic tests especially using Receiver Operating Characteristic methodology as well as statistical issues in clinical trials, with a more recent focus on microarrays. He serves as Associate Editor for Statistics in Pharmaceutical Research and Journal of Biopharmaceutical Statistics. He has been the recipient of the FDA's Commendable Service Award, Award of Merit and Outstanding Service Award as well as the CDRH Outstanding Scientific Award for Excellence in Analytical Science. He is a Fellow of the American Statistical Association. He has been a member since 1998 of the Senior Biomedical Research Service in the Department of Health and Human Services. He has served in leadership positions for the Eastern North American Region of the International Biometric Society and on the Board of Directors of the Society for Clinical Trials.

Banquet Keynote Speech: ***Statistics: a critical pharma--academic interface***

Speaker: Dr. William W. Chin, Harvard Medical School

Location and Time: On Time Seafood Restaurant, Tuesday, June 22, 6:30 p.m.–9:00 p.m.

About the Speaker: **Dr. William W. Chin** was born in New York, and received his A.B. in chemistry summa cum laude from Columbia College and his medical degree from Harvard Medical School. He then completed a residency in internal medicine at the Beth Israel Hospital and a fellowship in endocrinology and metabolism at the Massachusetts General Hospital, Boston, Massachusetts. Dr. Chin served on the faculty of Harvard Medical School for 25 years and was professor of medicine, and professor of obstetrics, gynecology and reproductive biology at Harvard Medical School; investigator of the Howard Hughes Medical Institute; and chief of the division of genetics and senior physician at the Brigham and Women's Hospital, Boston. Dr. Chin then joined Eli Lilly and Company in 1999, and served most recently as senior vice president of discovery research and clinical investigation, with overall responsibilities for the therapeutic areas, chemistry, toxicology, ADME and early clinical development. He serves as a member of the company's senior management council. He previously was vice president of discovery biology research and clinical investigation. In 2010, he returned to academia as the Executive Dean for Research and Professor of Medicine at Harvard Medical School. Dr. Chin is a world-renowned molecular endocrinologist who has pioneered the understanding of the mechanisms of nuclear receptor action with a focus on thyroid, estrogen, and other hormones, as well as various aspects of hormonal regulation of pituitary hormone gene expression. As the leader of the genetics program at the Brigham and Women's Hospital, he was responsible for the establishment of genomics, bioinformatics, and clinical genetics in the practice of medicine. He is author or co-author of more than 270 original papers, invited chapters or books, and he has served on numerous editorial boards and private and governmental review panels. He has received many accolades, including the Robert H. Williams Distinguished Leadership Award from the Endocrine Society, the Sidney H. Ingbar and Van Meter Awards from the American Thyroid Association, the Bowditch Award from the American Physiological Society, and the AFCR Young Investigator Award, and election to the American Society for Clinical Investigation and the American Association of Physicians. Dr. Chin has been a member of the board of scientific counselors of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH), and chair of the NIH endocrinology study section. He has also provided leadership on the council of several key professional societies, including service as president of the American Thyroid Association and the Interurban Clinical Club. Dr. Chin is a member of the Board of Directors at the Indianapolis Museum of Art and the Indianapolis Prize [the largest monetary prize for wildlife conservation in the world] Jury. He has also been an overseer of the New England Conservatory of Music and a co-chair of Project Success, a science program for underprivileged secondary students in Boston.

ROUND TABLE DISCUSSIONS

Round Table L1: **Implementation Issues of Adaptive Design Clinical Trials**

Discussion Leader(s): Dr. Cyrus Mehta, Dr. Jason Connor

Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Abstract: In this roundtable we discuss benefits and challenges of adaptive designs. Regulatory agencies are increasingly amenable to adaptive designs and Berry Consultants and Cytel both have extensive experience in working with pharmaceutical and medical device companies in designing adaptive clinical trials within the regulatory setting. This entails creating flexible, prospectively adaptive trials that also maintain statistical integrity. Depending upon the interests of participants we will discuss our experiences with adaptive designs, their benefits, the additional infrastructure they require, and our interactions with regulators (e.g. FDA, EMEA) regarding adaptive designs.

About Discussion Leader(s): **Dr. Cyrus Mehta** was born in Bombay, India. He studied engineering at the Indian Institute of Technology, Bombay and obtained his Ph.D. from MIT in 1973. Dr. Mehta is President and co-founder of Cytel Corporation and Adjunct Professor of Biostatistics, Harvard University. Cytel is a leading provider of software and services for the design, interim monitoring and implementation of adaptive clinical trials. Dr. Mehta consults extensively with the biopharmaceutical industry on group sequential and adaptive design, offers workshops on these topics, and sits on several data monitoring committees for these types of clinical trials. He has led the development of the StatXact, LogXact and East software packages that are widely used in the biopharmaceutical industry and at academic research centers. He publishes his methodological research results in leading statistics journals and is a past co-winner of the George W. Snedecor Award from the American Statistical Association for the best paper in biometry. He was elected a Fellow of the American Statistical Association in 1995 and named the Mosteller Statistician of the Year by the Massachusetts Chapter of the American Statistical Association in 2000. In 2002, Dr. Mehta was named Outstanding Zoroastrian Entrepreneur by the World Zoroastrian Chamber of Commerce. Dr. Mehta is the Zoroastrian Chaplain at Harvard and MIT.
Contact information: Cytel Inc. 675 Massachusetts Avenue, Cambridge, MA 02139, mehta@cytel.com, 617-661-2011, www.cytel.com.

Dr. Jason Connor received his PhD in Statistics & Public Policy from Carnegie Mellon University and his BS in Biomedical Engineering from Texas A&M University. He is uniquely trained as a biomedical engineer and a Bayesian biostatistician. Dr. Connor has worked for Berry Consultants since 2006 designing Bayesian adaptive trials for clients from a wide range of pharmaceutical and medical device

companies. He is a frequent speaker on Bayesian adaptive designs and the benefits they provide. He has coauthored over 50 papers in clinical journals and just finished a 6-year term as Associate Editor of The American Journal of Gastroenterology. He is also an Assistant Professor of Medical Education at the University of Central Florida.

Contact information: Berry Consultants, LLC 9757 Cypress Pine St Orlando FL 32827, jason@berryconsultants.com, 317-877-1084,www.berryconsultants.com.

Round Table L2: **Opportunities and Challenges to Build Your Own Consulting Firm**

Discussion Leader(s): Dr. Tai Xie

Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Abstract: In this economic down turn, some of us may have already felt unsecure of our current job. Some of us may have thought about being a consultant or establishing a consulting firm. What are the opportunities and what are the challenges? Is "being a boss of your own company" suitable for you? In this discussion, I will discuss the pro and con of this career option. I will share my own experience when I made the decision 5 years ago. I will share some of my stories about running my own CRO, which mixed with success and struggles.

About Discussion Leader(s): **Dr. Tai Xie** obtained his Ph.D. in Statistics in 1993 from the University of Arizona. Right after his Ph.D., he joined the Arizona Cancer Center as a Sr. Research Statistician. In 1996, Tai Xie joined Wyeth as a Sr. Biostatistician (1996-1999) and then rehired as an Assistant Director (2000-2003) managing several staff members and clinical projects in analgesics and man's health. Between 1999-2000, Tai Xie joined biostatistical group at the Pharmaceutical Research Institute (PRI) of Johnson & Johnson as a Lead Statistician for number of therapeutic areas. From May 2003-October 2004, Dr. Tai Xie joined ImClone System Incorporated, an oncology company, as an Associated Director responsible for statistics and data management aspects. In October 2004, Dr. Tai resigned his permanent job and founded a CRO, Brightech International. Since then, his company has been providing services in clinical data management and biostatistics to more than 60 clinical projects for 20+ pharmaceutical companies in US. The company currently has 20 employees and keeps a rapid growth and expansion. Brightech has established subsidiaries in Hong Kang, Shanghai and Chengdu for extending its business and services globally. In 2008, Tai Xie received an Excellent Business Achievement Award from New Jersey Chinese Chamber of Commerce. Besides the CRO business, Dr. Xie is serving as the CEO for BioPharm Solutions Inc, an innovative drug delivery company. He is also a partner of SUMMA Life Sciences LLC, a specialty biotech company focus on developing oncology drug candidates. Dr. Tai Xie has a number of research papers published in Biostatistical and Clinical Trial

journals. He has extensive experience in NDA/BLA submissions and FDA interactions, with rich knowledge in new drug application process and regulatory procedures. He is a member of American Statistical Association, Drug Information Association, ICSA and SAPA.

Contact information: Brightech International, tomx@brightech-intl.com, 908-790-8888, www.brightech-intl.com.

Round Table L3: **Monitoring Safety During Drug Development**

Discussion Leader(s): Dr. Brenda Crowe

Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Abstract: The Safety Planning Evaluation and Reporting Team (SPERT) formed in 2006 by the Pharmaceutical Research and Manufacturers of America includes industry biostatisticians, epidemiologists, and safety physicians, as well as representatives from the FDA. Our goal was to recommend an industry standard for safety planning, data collection, evaluation and reporting. The scope included new product development programs, from first-in-human studies through planning of post-approval period. In 2009 SPERT published a manuscript with several important ideas. A chief recommendation was that sponsors should create a Program Safety Analysis Plan (PSAP) early in development. We also gave recommendations for the planning of repeated, cumulative meta-analyses of the safety data obtained from the studies conducted in the development program. These included clear definitions of adverse events of special interest and standardization of many aspects of data collection and study design. In this lunch session we will discuss the SPERT recommendations and issues such as ways that companies can prepare for a potential regulatory expectation that PSAPs will be performed, timing of PSAPs, analyses that can impact preparedness for a regulatory filing/submission and any other related topics that arise.

About Discussion Leader(s): **Dr. Brenda Crowe** is a Research Advisor at Eli Lilly and Company, where she has worked since obtaining a PhD in Statistics from the University of Toronto in 1997. She has designed and analyzed many clinical trials and observational studies. She currently provides statistical expertise to Lilly's Global Patient Safety division. She is a co-chair of the Safety Planning, Evaluation and Reporting Team (SPERT), a cross-company group whose goal is recommend an industry standard for safety planning, data collection, evaluation and reporting.

Contact information: Eli Lilly And Company, Lilly Corporate Center, Indianapolis, IN 46285, bjcrowe@lilly.com, 317-276-3529.

Round Table L4: **Biopharmeutical Statistics and Biostatistics Development in China**

Discussion Leader(s): Professor Jun Shao, Ms. Helen Yin

Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Abstract: With growing presence of global pharmaceutical and CRO companies in China, alignment of China SFD A requirements with ICH/ GCP, rapid growth for outsourcing of statistical services to China , the number of biostatisticians and the demand for the biostatistics expertise are quickly increasing. In this lunch table discussion, we would like to share and exchange ideas on the development of biostatistics in China. Our discussion will cover: (1) the current state of biostatistics workforce in China, their educational background and experience, work scope exposure; (2) the increasing trend of junior or middle level staff with international education or working experience; and (3) the on-job training model, including advanced degrees at local post-graduate university programs, short-term working abroad opportunities, domestic conferences or workshops, e.g.: by SBF (Shanghai Biostatistics Forum, coordinated by a group of professionals from various institutions, professors at universities, and key opinion leaders in statistics in China). The focus of this discussion is to speed up the capability and capacity build-up in China biostatistics and to keep China biostatistics to be competitive enough to out-risk the potential trend of cost enhancement in China, especially in big cities.

About Discussion Leader(s): **Professor Jun Shao** is Professor of the Department of Statistics, University of Wisconsin-Madison. He is Fellow of the Institute of Mathematical Statistics and Fellow of the American Statistical Association. He was the president of the International Chinese Statistical Association in 2007. He was or is an associate editor for many statistics journals. Jun Shao received the B.S. degree in 1982 in mathematics from East China Normal University and the PhD degree in 1987 in statistics from the University of Wisconsin-Madison. He is an author or coauthor of nearly 150 research articles and 6 books. He has produced 19 PhD students in statistics since 1996.

Contact information: Room 1235A, MSC, 1300 University Ave., Madison, WI 53706-1685, shao@stat.wisc.edu, 608-262-7938, http://www.stat.wisc.edu/~shao/.

Ms. Helen Yin is the CEO of MacroStat China, a CRO providing statistics and data management services in clinical research. Helen worked at Parexel US for 2 years, and then at AstraZeneca for 6 years in the US and 4 years in China. Prior to returning to China in 2003, Helen was Associate Director in the Biometrics Department at AstraZenica US and the US therapeutic lead for oncology products. Helen is the co-founder and CEO of MacroStat China since December of 2005. She has made significant contributions to the growth of MacroStat China, the leading statistics CRO in China, which has grown to 50 employees as of December 2009.

Contact information: MacroStat China, Room 11301-11304, Block 11, 498 Guo Shou Jing Road Shanghai, 201203, China, helen.yin@macrostat.com, 86 21 5027-2925.

Round Table L5: **Effective Communications Between Industry and FDA**

Discussion Leader(s): Dr. Suktae Choi, Dr. James Hung

Location and Time: Vision, Monday, June 21, 2010, 12:20 p.m.–1:20 p.m.

Abstract: Effective communication between pharmaceutical companies and regulatory agencies, such as the Food and Drug Administration (FDA), is a key factor in the process of drug development and also for regulatory reviewers. Communication is done by many ways including briefing documents, face-to-face or telecom meetings, and written communications (e.g. fax and email). Recently, the role of statistical methodology in clinical trials has increased greatly. Thus, communication between statisticians in the pharmaceutical industry and those at the FDA has become more important. Many statisticians in the pharmaceutical industry are eager to exchange ideas with statisticians at the FDA. However, appropriate ways of communication are not obvious. During this lunch meeting, each participant will share examples and perspectives on communication between the pharmaceutical industry and the FDA.

About Discussion Leader(s): **Dr. Suktae Choi** has 12 years experience in the clinical trial area beginning with his employment as a project statistician of Organon, Inc. In 2000, he moved into Center for Drug Evaluation and Research, US FDA as Mathematical Statistician. During 7+ years with the FDA, he reviewed many clinical trials in analgesic, arthritis, ophthalmic, and anti-viral area submitted for approval of drug indication. He joined Eli-Lilly and Company in 2007 as Sr. Research Scientist, and currently leader of Byetta statistics group. The research areas that he is interested in are multivariate survival analyses, multiple comparisons, and missing data imputations. He published statistical articles in statistics and medical journals and presented in public. He received Master's degree of Statistics from Michigan State University, and also received Ph.D. degree of Statistics from Rutgers University in 1998. *Contact information*: Eli Lilly and Company, Lilly Corporate Center Indianapolis, IN 46285, CHOI_SUKTAE@LILLY.COM, 317-655-9104.

Dr. James Hung is presently Director of Division of Biometrics I, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, US FDA. The division provides services for Division of Cardiovascular and Renal Products, Division of Neurology Products and Division of Psychiatry Products. During his 20+ year tenure with FDA, he reviewed many large mortality/morbidity clinical trials in cardiovascular and renal disease areas. He published articles in Biometrics, Statistics in Medicine, Controlled Clinical Trials, Biometrical Journal, Journal of Biopharmaceutical Statistics. His research areas include factorial design clinical trials, utility of p-value distribution, adaptive design/analysis in clinical trials, and non-inferiority trials, multi-regional clinical trials. Dr. Hung received two FDA/CDER Scientific Achievement Awards and many other awards for the recognition of his scientific contributions to the US FDA. He gave many invited talks, lectures or short

courses in United States, France, Germany, United Kingdom, Japan, China, and Taiwan. Currently, he serves as an Editor-in-Chief for Journal of Pharmaceutical Statistics and an Associate Editor for Statistics in Medicine and Journal of Biopharmaceutical Statistics.

Contact information: FDA, 10903 New Hampshire Avenue, Building 21, Room 4616, Silver Spring, MD 20993, hsienming.hung@fda.hhs.gov, 301-796-1092.

Round Table L6: ***Statistical Outsourcing: Issues and Future***

Discussion Leader(s): Dr. Todd Sanger

Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Abstract: Across the industry the growing trend is to outsource statistical programming as well as other biostatistical activities to outsourcing companies. This outsourcing is separate from using a CRO to run a clinical trial. Many of these outsourcing companies are offshore. This increase in outsourcing necessitates highly efficient communication of requirements, issues managemet and final oversight in order to get a quality final product. We will talk about the issues with such outsourcing and attempt to predict future developments.

About Discussion Leader(s): **Dr. Todd Sanger** is currently a Senior Research Advisor in Global Statistical Sciences supporting the Auto Immune and Critical Care Business Unit. He advises on technical issues across the business unit as well as throughout GSS. From 2002 to 2009, Todd was the Senior Director of Global Statistical Sciences. He was responsible for all areas of statistics within Global Brand Development as well as for Global Statistical Operations at Lilly. From 1999-2002, Dr. Sanger was the manager of operations for the Symbyax Product Team where he was responsible for all aspects of the development and registration of the combination product. Prior to that, he was the principal statistician for the Zyprexa Product Team leading the team through numerous submissions and approvals for the treatment of psychosis and bipolar disorder. These experiences have offered Dr. Sanger numerous opportunities to interact with regulatory agencies throughout the world.

Dr. Sanger is an author of over 25 manuscripts and 100 presentations, many in the field of psychiatry. His current interests include clinical trial design, linear models, causality analysis, statistical computing, and statistical organization and administration.

Dr. Sanger is a native of Michigan. He obtained his bachelor's degree in mathematics summa cum laude in 1987 from Michigan Technological University. He went on to obtain his doctorate in statistics at Iowa State University in 1992.

Dr. Sanger is a member of the American Statistical Association, including the Biopharmaceutical Section and the local Central Indiana Chapter, the Biometric Society, and the Drug Information Association.

Contact information: Eli Lilly and Company, Lilly Corporate Center Indianapolis, IN 46285, sanger@lilly.com, 317-276-0279.

Round Table L7: **How to Get Research Funding from Various Sources: Industry, Government and Academia**

Discussion Leader(s): Dr. Jeremy M G Taylor

Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Abstract: Academic statisticians have a number of avenues for pursuing funding for their own methods research. Academic institutions typically have seed money and pilot grant programs that are targeted at junior faculty and provide a relatively small amount of support for 1 or 2 years. Amongst government sources NSF is particularly well suited for those in Statistics Departments who are seeking summer research funding. The NIH is a major source of funding for those whose research interests are linked with biomedical research. Biostatisticians may seek funding as collaborators or co-investigators working with others on a defined biomedical research problem or they make seek funding as a Principal Investigator to develop their own research. A grant application to NIH would be reviewed by an appropriate study section, the most likely study section for statistical methodology development is the Biostatistical Methods and Research Design (BMRD) study section. Review committees at NSF and NIH will review the quality and likely impact of the proposed research. NIH and NSF accept both unsolicited grant applications and applications that respond to a particular need. Obtaining funding from NSF and NIH for methodology development is competitive and can be challenging with typically between 10% and 30% of grants being funded. Obtaining research funding from industry is also possible. The nature of the funding is more likely to be driven by the needs of the company, and personal contacts can be important for securing funding.

About Discussion Leader(s): **Dr. Jeremy M G Taylor** obtained his PhD from UC Berkeley, Department of Statistics in 1983. He was on the faculty of the Department of Biostatistics at UCLA from 1983 to 1998. He is currently the Pharmacia Professor of Biostatistics at the University of Michigan. He is a former winner of the Mortimer Spiegelman Award from APHA and the Michael Fry Award from the Radiation Research Society. He is a Fellow of the ASA. He is a former chair of the Biometrics Section of ASA and is a member of the IBC Council. He is an Associate Editor of Biometrics. He is currently a member of the BMRD study section. His research interests are in Box-Cox transformation, longitudinal and survival analysis, cure models, biomarkers, joint longitudinal/survival models, bioinformatics, clinical trial design and applications in cancer and radiation oncology.

Contact information: Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI, 48109, jmgt@umich.edu, 734-936-3287.

Round Table L8: **Career in Industry: Technical or Administrative?**

Discussion Leader(s): Dr. Gordon Lan, Dr. Pandurang Kulkarni

Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Abstract: Drs Pandurang Kulkarni (Eli Lilly) and Gordon Lan (Johnson & Johnson) were professors before they joined the pharmaceutical industry. They will share with you their experience on career development and lead discussion on topics include:

(1) Leadership role statisticians can play in the pharmaceutical industry

(2) Managing across boundaries and leading strategies in the pharmaceutical industry

(3) Role and responsibility of a statistical consultant in the pharmaceutical industry

(4) Role and responsibility of a research statistician in the pharmaceutical industry.

About Discussion Leader(s): **Dr. Gordon Lan** received his Ph.D. in Mathematical Statistics from Columbia University in 1974. Before joining Johnson & Johnson in 2005 as Senior Director of Statistical Science, he held positions as Mathematical Statistician at the National Heart, Lung and Blood Institute (NHLBI/NIH), Professor of Statistics at George Washington University, Distinguished Scientist at Pfizer and Statistics Fellow at Sanofi-Aventis. Dr. Gordon is responsible for providing scientific and technical consultation to Biostatistics & Programming colleagues, clinical teams and management. He participates in methodological research projects, introduce new statistical approaches in support of organizational goals and represent programs both internal and externally (e.g., medical and statistical community forums). Dr. Gordon has published more than 50 research papers in professional journals and has given more than 200 invited talks at universities and professional meetings. He is interested in statistical methods for clinical trial design and data analysis. Gordon was elected Fellow of the American Statistical Association (ASA) in 1992 and Fellow of the Society for Clinical Trials in 2009. Currently, he serves as Co-Editor of the journal Statistics in Biosciences. He also serves on the Board of Directors, Society for Clinical Trials.

Contact information: Johnson & Johnson, glan@its.jnj.com, 908-704-5001.

Dr. Pandurang Kulkarni is the Senior Director of Global Statistical Sciences, Oncology. He obtained his Ph.D. in Statistics at the LaTrobe University, Melbourne, Australia. He taught and did research in statistics at the University of South Alabama for 10 years where he attained tenure and full professorship.

He has published more than 50 articles in statistics and medical areas. He received several grants from the Air Force for the development of techniques to evaluate and study the effects of environmental contaminants such as dioxin on humans. He also received grants from NASA to study the effects of microgravity to see the effectiveness of exercise countermeasures in minimizing the effects of microgravity on astronauts. Dr. Pandu Kulkarni joined Eli Lilly in 2000. He has worked in several therapeutic areas including Osteoporosis, Oncology and Neuroscience. He has designed clinical trials and has numerous scientific publications in these areas. He is regularly invited to provide training, CME, and workshops on use of statistics in medical research. He has managed large groups consisting of Medical Liaisons, Medical Information, Scientific Communications and Statistics within the US Medical Division of Eli Lilly and Company. He currently manages Statisticians supporting Oncology portfolio across phases and geographical regions.

Contact information: Eli Lilly and Company, Lilly Corporate Center Indianapolis, IN 46285, kulkarni@lilly.com, 317-433-1486.

Round Table L9: **Careers for New Statisticians in Industry**

Discussion Leader(s): Dr. Christy Chuang-Stein

Location and Time: Vision, Tuesday, June 22, 2010, 12:20 p.m.–1:20 p.m.

Abstract: Many statisticians choose a career in industry, lured by the prospect of making contributions to real-life problems. Unfortunately, the need to take care of urgent and higher priority project needs on a daily basis means that there is often little time left to attend to individual's professional development. Since scholarly activities do not necessarily carry much weight for job performance reviews, it is easy for industry statisticians to trade off professional development needs for project needs. At this round table, we will discuss ways to balance these different needs for industry statisticians. Part of the solutions is to convince management and ourselves that professional development is essential to effectively delivering on our jobs. As a professional, we need to take time to regularly sharpen our saws so we could efficiently execute on our jobs. In addition, taking advantage of professional networks could help us connect with others with whom we can grow and learn together. Please join me at this luncheon table to exchange ideas, both from the perspective of sharing your experience and that of learning from others' experience.

About Discussion Leader(s): **Dr. Christy Chuang-Stein** , a Fellow of the American Statistical Association since 1998, is Head of the Statistical Research and Consulting Center at Pfizer. Christy has 25 years of pharmaceutical industry experience. She serves on the editorial boards of Pharmaceutical Statistics and Drug Information Journal (DIJ). Christy is a Vice President of the American

Statistical Association and is a member of many industry-wide collaborative working groups. Christy has over 115 publications in peer-reviewed journals and book chapters.

Contact information: Pfizer Inc. 50 Pequot Avenue, New London, CT 06320, christy.j.chuang-stein@pfizer.com, 269-567-0263.

ABSTRACTS BY PRESENTER'S LAST NAME

Presenter: *Abowd, John* **Affiliation:** *Cornell University*
Abstract Title: ***Measurement Error in Large-scale Administrative Record Statistical Systems***
Author(s): *John M. Abowd**
Abstract: Building frames, preparing summary tabulations, and sourcing some survey information from administrative data has a long history in statistics. Constructing entire statistical systems from integrated, multi-source administrative data is a much more recent phenomenon. Population coverage, source data quality, linkage management, edit procedures, imputation, and privacy protection all interact in such systems. The paper provides examples of the importance of each of these sources of potential error in the context of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics program infrastructure file system. This longitudinally integrated employer-employee database is the source for multiple detailed public-use data products (Quarterly Workforce Indicators, OnTheMap), survey data integration (Survey of Income and Program Participation), and restricted access studies based on the confidential micro-data.

Session: M17
Regency B
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Alosh, Mohamed* **Affiliation:** *FDA*
Abstract Title: ***A general approach for testing a prespecifeid subgroup in clinical trials***
Author(s): *Mohamed Alosh* and Mohammad Huque*
Abstract: Subgroup analyses are commonly used in clinical trials with the objective of learning about differential treatment effect across subgroups. Due to power consideration among other factors, clinical trials are seldom considered for establishing an efficacy claim for a subgroup in case the trial fails to establish an efficacy claim for the total population. However, through proper study design and analysis the clinical trial can be designed to establish efficacy claim for the total population as well as for the subgroup, thus increasing the chance of a positive trial. The concern that clinical trials are underpowered for subgroups can be relaxed somewhat through enrichment of the patient population for a priori identified subgroup and by using statistical testing strategies which spend the overall Type I error rate more efficiently than many traditional methods. In this presentation we consider a multiple testing strategy for the total population and the subgroup with the following features: (i) ensuring consistency of efficacy findings of the total population and that of the subgroup so that the results of the study overall are interpretable and (ii) allowing the significance level for testing the subgroup to adapt to the efficacy findings of the total population in a general form. We consider application of the proposed methodology to clinical trial data.

Session: W12
Regency B
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Andrieu, Christophe* **Affiliation:** *University of Bristol*
Abstract Title: ***On the computation of normalizing constants***
Author(s): *C. Andrieu*, joint with N. Whiteley*
Abstract: The estimation of normalizing constants is central to numerous inference procedure. We review some techniques, provide novel theoretical results as well as new methods which aim to achieve optimality in a sense to be made precise during the presentation.

Session: M26
Cosmopolitan C
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Austin, Peter* **Affiliation:** *Institute for Clinical Evaluative Sciences*
Abstract Title: ***Optimal estimation of risk differences using propensity-score matching***
Author(s): *Peter Austin**
Abstract: Propensity-score matching allows one to create a sample of treated and untreated subjects

Session: M28
Cosmopolitan D
Monday, June 21
1:30 p.m.–3:20 p.m.

who, on average, will have a similar distribution of measured baseline covariates. In this talk, I will discuss two issues related to estimating risk differences (or absolute risk reductions) using propensity-score matching. First, I examine the relationship between the caliper width used for matching on the propensity score and the bias and precision of the estimated risk difference. I determine caliper widths that optimize the variance-bias tradeoff. Second, balance in measured baseline covariates is a large-sample property: treated and untreated subjects in a propensity-score matched sample will, on average, have a similar distribution of measured baseline covariates. However, just as in RCTs, it is possible for there to be residual differences in baseline characteristics between treated and untreated subjects after matching on the propensity score. I discuss methods to estimate risk differences in propensity-score matched samples that allow for one to account for residual imbalance in measured baseline covariates. These methods allow one to retain the risk difference as the measure of treatment effect, rather than adopting the adjusted odds ratio.

Presenter: *Banerjee, Moulinath* **Affiliation:** *University of Michigan, Ann Arbor*

Abstract Title: ***BASELINE ZONE DETECTION VIA P-VALUES.***

Author(s): *Moulinath Banerjee, Atul Mallik, Bodhi Sen, George Michailidis*

Abstract:

Session: M24
Concept CD
Monday, June 21
1:30 p.m.–3:20 p.m.

We introduce the problem of detecting a region where a regression function achieves an extremal value, i.e. a minimum or a maximum (which may be unknown) in nonparametric and semiparametric settings. The identification of such regions is important in many applications like dose-response studies, astronomy, `omics' experiments in biology, image-detection, signal-processing and spatial problems. We discuss the dose-response problem in detail, using p--values as a discrepancy criterion for identifying the threshold value at which the response takes off from its baseline value. We study the problem in a controlled sampling setting, where multiple responses, discrete or continuous, can be obtained at a number of different dose-levels. Our procedure involves testing the hypothesis that the regression function is at its baseline at each dose using the sampled responses at that value and then computing the p--value of the test. An estimate of the threshold is provided by fitting a stump, i.e., a piecewise constant function with a single jump discontinuity, to the observed p--values, since the corresponding p--values behave in markedly different ways on different sides of the threshold. The estimate is shown to be consistent, as both the number of doses and the number of responses sampled at each dose become large,and its finite sample properties are studied. Our approach is computationally simple and highly competitive as comparisons to standard procedures in the dose-response literature demonstrate.

Presenter: *Bekele, B. Nebiyou* **Affiliation:** *M.D. Anderson Cancer Center*

Abstract Title: ***Pseudo-Data Augmented Bayesian Weighted Likelihood Bootstrap for Early Phase Oncology Trial Designs***

Author(s): *B. Nebiyou Bekele*

Abstract:

Session: M11
Concept AB
Monday, June 21
10:20 a.m.–12:10 p.m.

In this talk I discuss a general approach to designing Bayesian early phase clinical trials using standard generalized linear models procedures and functions in R. I propose an extension of the Bayesian Bootstrap (Rubin, 1981) and Bayesian Weighted Likelihood estimation (Newton and Raftery, 1994). This extension, which I call Pseudo-data Augmented Bayesian Weighted Likelihood Bootstrap, can be applied to a variety of designs which are difficult to model from scratch but can easily be handled via standard generalized linear models. I will provide several examples and will discuss some of the advantages of this approach over models developed from scratch or using WinBUGS.

Presenter: *Bies, Robert* **Affiliation:** *Indiana University School of Medicine*

Abstract Title: ***Population pharmacokinetics as a means of capturing exposure magnitude and consistency***

Author(s): *Robert R Bies**

Abstract: The application of population pharmacokinetics to clinical studies of psychotropics encompassing large observational studies such as the Clinical Antipsychotic Trials of Intervention Effectiveness, the SPECTRUM and Maintenance Therapy in Late Life Depression studies will be presented along with a rationale for the potential utility of magnitude and consistency of exposure when considering response patterns in schizophrenia, Alzheimer's disease and depression.

Session: T23
Regency B
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Bunea, Florentina* **Affiliation:** *Florida State University*

Abstract Title: ***Adaptive Rank Penalized Estimators in Multivariate Regression***

Author(s): *Florentina Bunea*

Abstract: We introduce a new criterion, the Rank Selection Criterion (RSC), for selecting the optimal reduced rank estimator of the coefficient matrix in multivariate response regression models. The corresponding RSC estimator minimizes the Frobenius norm of the fit plus a regularization term proportional to the number of parameters in the reduced rank model. The rank of the RSC estimator provides a consistent estimator of the rank of the coefficient matrix. The consistency results are valid not only in the classic asymptotic regime, when the number of responses n and predictors p stays bounded, and the number of observations m grows, but also when either, or both, n and p grow, possibly much faster than m. Our finite sample prediction and estimation performance bounds show that the RSC estimator achieves the optimal balance between the approximation error and the penalty term. Furthermore, our procedure has very low computational complexity, linear in the number of candidate models, making it particularly appealing for large scale problems. We contrast our estimator with the nuclear norm penalized least squares estimator (NNP). We show that NNP has estimation and prediction properties similar to those of RSC, albeit under stronger conditions. However, it is not as parsimonious as RSC. We offer a simple correction of the NNP estimator which leads to consistent rank estimation. An application to neuro-imaging demonstrates the usage of our methods in practice.

Session: M24
Concept CD
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Cao, Hongyuan* **Affiliation:** *UNC-Chapel Hill*

Abstract Title: ***Comparing proportions of extremely rare events of uncertain status with applications to vaccine safety studies***

Author(s): *Hongyuan Cao, Lisa M. LaVange, Joseph F. Heyse, Michael R. Kosorok*

Abstract: We generalize the logrank statistic to the extremely rare setting where only a small number of events are observed and the events have uncertain status. In this setting, the limiting distribution is not Guassian but, rather, a ratio of Poisson random variables. We borrow a Bayesian idea to develop a frequentist valid testing procedure for comparing the intensity difference between two populations. Both numerical and asymptotic properties of the procedure are studied. This testing procedure is applied to a post-market safety study of a vaccine.

Session: M38
Studio One
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Carroll, Raymond* **Affiliation:** *Texas A&M University*

Abstract Title: ***Generalized Functional Latent Feature Models with Single-Index Interactions***

Author(s): *Yuhua Li, Naisyin Wang and Raymond J. Carroll**

Abstract: We introduce a new class of functional generalized linear models, where the response is a scalar and some of the covariates are functional. We assume that the response depends on multiple covariates, a finite number of latent features in the functional predictor, and interaction between the two. To achieve parsimony, the interaction between the multiple covariates and the functional predictor is modeled semiparametrically with a single-index

Session: M14
Cosmopolitan C
Monday, June 21

10:20 a.m.–12:10 p.m.

structure. We propose a two step estimation procedure based on local estimating equations, and investigate two situations: (a) when the basis functions are pre-determined, e.g., Fourier or wavelet basis functions and the functional features of interest are known; and (b) when the basis functions are data driven, such as with functional principal components. Asymptotic properties are developed. Our methods are illustrated with a simulation study and applied to an empirical data set, where a previously unknown interaction is detected.

Presenter: *Carvalho, Luis* **Affiliation:** *Boston University*

Abstract Title: ***Bayesian Inference for Genome-Wide Association Studies***

Author(s): *Luis Carvalho*

Abstract:

Session: T37
Cosmopolitan D
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Genome-Wide Association Studies (GWAS) attempt to identify a (possibly small) subset of single nucleotide polymorphisms (SNPs) from a large number of measured candidates that are associated with a specific observable trait. Identification of associated genetic variants is particularly hard due to very large genotype sizes in comparison to case-control group sizes. We formally frame this problem as Bayesian variable selection in a logistic regression model with spike-and-slab priors in the coefficients. To help overcome the curse of dimensionality, we further explore a co-dependency structure between SNPs by setting an Ising hyper-prior on the space of possible SNP associations to trait. We introduce and derive centroid and graph centroid estimators and contrast them to the classical MAP estimators in this setup. We illustrate this approach with a toy example and a larger simulated dataset based on HapMap. Finally, we offer a few concluding remarks and directions for future work.

Presenter: *Chao, Edward* **Affiliation:** *Data Numerica Institute*

Abstract Title: ***Some Computational Foundations for Online Correlated Data Analysis***

Author(s): *Edward C. Chao*

Abstract:

Session: T33
Network
Tuesday, June 22
3:40 p.m.–5:30 p.m.

An integrated online data analysis platform requires client-server computation, statistical computation engine, dynamic graphical methods, and web technologies, etc. Data analysis tools such as R are useful in constructing advanced statistical methods, which might not be available to internet applications. We are developing several tools for online correlated data analysis. A web environment incorporating these tools allows data analysts to apply advanced methods online. A graphical library can be used for developing dynamically linked graphical windows and animation. Statistical packages based on LME, GLME, GAMM and GEE kernels are available. These tools are useful for analyzing multilevel correlated data and non-ignorable dropout data in model fitting and sensitivity analysis. We can visualize the impacts from the departure of missing not at random vs. misisng at random. The environment aims at users for sharing resources online, e.g. publishing and sharing data, analyses, graphics and methods. In addition to running data analysis, people can search, learn and make comments on case studies.

Presenter: *Charnigo, Richard* **Affiliation:** *University of Kentucky*

Abstract Title: ***Nonparametric Derivative Estimation and Posterior Probabilities for Nanoparticle Characteristics***

Author(s): *Richard Charnigo*, Mathieu Francoeur, Patrick Kenkel, M Pinar Menguc, Benjamin Hall, Cidambi Srinivasan*

Abstract:

Session: T33
Network
Tuesday, June 22
3:40 p.m.–5:30 p.m.

The characterization of nanoparticles from surface wave scattering data is of great interest in applied engineering because of its potential to advance nanoparticle-based manufacturing concepts. Meanwhile, a recent development in methodology for the nonparametric estimation of a mean response function and its derivatives has provided a valuable tool for nanoparticle characterization: namely, a mechanism to identify the most plausible configuration for a collection of nanoparticles given the estimated derivatives of surface wave scattering profiles from those nanoparticles. In this talk, after briefly reviewing the preceding work, we propose an extension that additionally furnishes posterior probabilities for the various possible

configurations of nanoparticles. An empirical study is included as a demonstration.

Presenter: *Chen, Ling* **Affiliation:** *FDA*

Abstract Title: ***An Equivalence Test for the Comparison between a Test Drug and Placebo in Human Abuse Potential Studies***

Author(s): *Ling Chen*, Katherine Bonson*

Abstract: A randomized double-blind, double-dummy, placebo- and positive-controlled crossover designed human abuse potential study has at least four treatments: placebo, at least two doses of a test drug, and at least one dose of a positive control drug (a drug with known abuse potential). Study subjects are healthy volunteers with histories of recreational drug use. In such studies, placebo responses are compared with responses from the positive control drug in order to validate the study, and are also compared with responses from the test drug to assess whether the test drug has abuse potential. In the past, the U.S. Food and Drug Administration (FDA) has concluded that a test drug has abuse potential if positive subjective responses to the test drug are not significantly lower than double the mean response of placebo, regardless of drug class. This method of using twice the mean placebo response, rather than the single mean, in the comparison has been challenged by some Sponsors. This presentation proposes a new methodology for assessing whether a test drug has abuse potential when compared to placebo.

Session: M38
Studio One
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Chen, Xiang* **Affiliation:** *Yale Univ*

Abstract Title: ***The null distributions of test statistics in genomewide association studies***

Author(s): *Xiang Chen*, Heping Zhang*

Abstract: Genomewide association (GWA) studies assay hundreds of thousands of SNPs simultaneously across the entire genome and associate them with diseases, other biological or clinical traits. The association analysis usually tests each SNP as an independent entity and ignores the biological information such as linkage disequilibrium. Although the Bonferroni correction and other approaches have been proposed to address the issue of multiple comparisons as a result of testing many SNPs, there is a lack of understanding of the distribution of an association test statistic when an entire genome is considered together. In other words, there are extensive efforts in hypothesis testing, and almost no attempt in estimating the density under the null hypothesis. By estimating the true null distribution, we can apply the result directly to hypothesis testing; better assess the existing approaches of multiple comparisons; and evaluate the impact of linkage disequilibrium on the GWA studies. To this end, we estimate the empirical null distribution of an association test statistic in GWA studies using simulated population data. We further propose a convenient and accurate method based on adaptive spline to estimate the empirical value in GWA studies and validate our findings using a real data set. Our method enables us to fully characterize the null distribution of an association test that not only can be used to test the null hypothesis of no association, but also provide important information about the impact of density of the genetic markers on the significance of the tests. Our method does not require users to perform computationally intensive permutations, and hence provides a timely solution to an important and difficult problem in GWA studies.

Session: M16
Regency A
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Chen, Wei* **Affiliation:** *Wayne State University*

Abstract Title: ***BVS with Joint Modeling of Categorical and Survival Outcomes: An Application to Individualizing Chemotherapy***

Author(s): *Wei Chen∗, Debashis Ghosh, Trivellore E. Raghunathan, and Daniel J. Sargent*

Abstract: Colorectal cancer is the second leading cause of cancer related deaths in the United States, with more than 130,000 new cases of colorectal cancer diagnosed each year. Clinical studies have shown that genetic alterations lead to different responses to the same treatment, despite the morphologic similarities of tumors. A molecular test prior to treatment could help in

Session: M32
Regency A
Monday, June 21

3:40 p.m.–5:30 p.m.

determining an optimal treatment for a patient with regard to both toxicity and efficacy. This article introduces a statistical method appropriate for predicting and comparing multiple endpoints given different treatment options and molecular profiles of an individual. A latent variable-based multivariate regression model with structured variance covariance matrix is considered here. The latent variables account for the correlated nature of multiple endpoints and accommodate the fact that some clinical endpoints are categorical variables and others are censored variables. The mixture normal hierarchical structure admits a natural variable selection rule. Inference was conducted using the posterior distribution sampling Markov chain Monte Carlo method. We analyzed the finite-sample properties of the proposed method using simulation studies. The application to the advanced colorectal cancer study revealed associations between multiple endpoints and particular biomarkers, demonstrating the potential of individualizing treatment based on genetic profiles.

Presenter: *Chen, Jianwei* **Affiliation:** *San Diego Stae University*

Abstract Title: ***Modeling HIV viral dynamic model with longitudinal data***

Author(s): *Jianwei Chen* and Suzanne Papp*

Abstract:

Session: W13
Concept AB
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Deterministic dynamic equations have proved very important in describing the response of HIV-1 to treatment in vivo. This thesis takes as a starting point a linear differential equation that models the rate of HIV-1 RNA concentration in the patient's plasma. In [3] a method was developed for local estimation of the time-varying coefficients in this equation. A two-step method is used. In the first step the viral load and the time derivative of the viral load is estimated. The second step uses the results of step one in estimating the time-varying coefficients. This paper extends the method from a single patient to longitudinal analysis of multiple patients using the local polynomial mixed-effect model developed in [28]. The first step uses local polynomial mixed-effects methods to estimate the viral load and the time derivative of the viral load. There are two models used for the second step. In the first, mixed-effects methods are used to estimate constant coefficients. In the second, local polynomal mixed-effects methods are used to estimate time-varying coefficients. We use a real data application and also numerical simulations to illustrate the two estimation methods.

Presenter: *Chen, Yeh-Fong* **Affiliation:** *FDA*

Abstract Title: ***Placebo Response in Major Depression and Schizophrenia Trials: Statistical Consideration and Design Strategies***

Author(s): *Yeh-Fong Chen* and Yang Yang*

Abstract:

Session: W23
Regency B
Wednesday, June 23
10:10 a.m.–12:00 p.m.

High placebo response is a common reason to blame for a clinical trial that failed to demonstrate a drug's efficacy. There has been a notion that placebo response in psychiatric trials is increasingly strong so that it becomes more challenging to conduct a positive clinical trial successfully. In this talk, we will present the placebo response over time observed from the major depression and schizophrenia trials compiled in the FDA database in the hope for promoting extensive discussions. Some proposed designs aiming to reduce the impact of placebo response on psychiatric trials will be reviewed and examined. Chief among them is a novel design named sequential parallel design (Fava et al. 2003), which intends to reduce both the placebo response rate and the sample size requirement. Finally, detailed results of our proposed analysis and simulations concerning the applicability of this design will be illustrated.

Presenter: *Chen, Ming-Hui* **Affiliation:** *University of Connecticut*

Abstract Title: ***Bayesian Design of Non-Inferiority Trials for Medical Devices Using Historical Data***

Author(s): *Ming-Hui Chen*, Joseph G. Ibrahim, Peter Lam, Alan Yu, and Yuanye Zhang*

Abstract:

Session: W11

We develop a new Bayesian approach of sample size determination (SSD) for the design of non-inferiority clinical trials. We extend the fitting and sampling priors of Wang and Gelfand (2002) to Bayesian SSD with a focus on controlling the type I error and power. The historical

Regency A
Wednesday, June 23
8:00 a.m.–9:50 a.m.

data are incorporated via a hierarchical modeling approach as well as the power prior approach of Ibrahim and Chen (2000). Various properties of the proposed Bayesian SSD methodology are examined and a simulation-based computational algorithm is developed. The proposed methodology is applied to the design of a non-inferiority medical device clinical trial with historical data from previous trials.

Presenter: *Chen, John J.* **Affiliation:** *Stony Brook University*

Abstract Title: ***Adaptive Fitting of Linear Mixed-Effects Models with Correlated Random-effects***

Author(s): *Guangxiang Zhang*, John J. Chen*

Abstract:

Session: W13
Concept AB
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Mixed-effect models (MEMs) have been widely used in longitudinal data analysis as they allow for correlations among repeated measurements from the same unit. How to best model random effects of MEMs is still an important and unresolved issue in practice. For example, the non-convergence rate is about 20 percent on average in simulation studied by current R algorithm. We propose a data-driven algorithm to adaptively fit the MEMs that reduces the correlation among random effects in transformed parameter space. Simulations show that the proposed algorithm significantly improves the non-convergence rate (down to 1 percent), the reduction of correlation among random effects and the fitted log-likelihoods. Two real data sets are used to illustrate the application of this algorithm.

Presenter: *Chen, jiahua* **Affiliation:** *Department of Statistics, UBC, Canada*

Abstract Title: ***Adjusted empirical likelihood***

Author(s): *jiahua Chen*, Yukun Liu*

Abstract:

Session: W25
Concept CD
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Empirical likelihood enjoys many nice statistical properties. Yet when the sample size is small, or the dimension of the accompanying estimating function is high, its application is hindered by low precision of the chisquare approximation and by non-existence of solutions to the estimating equations. In this talk, we show that the adjusted empirical likelihood is effective at addressing both problems. At a specific level of adjustment, the adjusted empirical likelihood achieves the high-order precision of the Bartlett correction, in addition to the advantage of a guaranteed solution to the estimating equations. Simulation results indicate that the confidence regions constructed by the adjusted empirical likelihood have coverage probabilities comparable to or substantially more accurate than the original empirical likelihood enhanced by the Bartlett correction.

Presenter: *Cheng, Fuxia* **Affiliation:** *Illinois State University*

Abstract Title: ***Global Property of Error Density Estimation in Nonlinear Autoregressive Time Series Models***

Author(s): *Fuxia Cheng*

Abstract:

Session: M34
Concept CD
Monday, June 21
3:40 p.m.–5:30 p.m.

In this talk, we consider estimation of the error density function in nonlinear autoregressive stationary time series regression model. The asymptotic distribution of the maximum of a suitably normalized deviation of the density estimator from the expectation of the kernel error density (based on the true error) is obtained to be the same as in the case of the one sample set up, which is given in Bickel and Rosenblatt (1973).

Presenter: *Cheng, Guang* **Affiliation:** *Dept of Statistics, Purdue University*

Abstract Title: ***How many iterations are sufficient for semiparametric estimation?***

Author(s): *Guang Cheng*

Abstract:

Session: M14
Cosmopolitan C
Monday, June 21
10:20 a.m.–12:10 p.m.

In semiparametric models, a common practice in obtaining an efficient estimator for the Euclidean parameter is through iteratively optimizing some objective function w.r.t. its Euclidean parameter and functional parameter via some numerical algorithm. For example, the semiparametric MLE can be obtained by maximizing its profile likelihood via the Newton-Raphson algorithm. The main purpose of this talk is to propose a general approach in constructing such numerical outcome (without knowing the form of the efficient influence

function) and, more importantly, calculate the minimal number of iterations needed to obtain a semiparametric efficient estimator from a theoretical point of view. In addition, we also answer the below two questions: (a) what factors determine the higher order asymptotic efficiency of our estimator; (b) how to find a consistent initial estimate for the above iterative procedure.

Presenter: *Cho, Meehyung* **Affiliation:** *Sanofi-aventis*

Abstract Title: ***Considerations for Design and Data Analysis of Adaptive Superiority/Non-Inferiority Cardiovascular Trials***

Author(s): *Meehyung Cho*, Hui Quan, Mingyu Li, Peng-Liang Zhao, Ji Zhang and Yujun Wu*

Abstract: The FDA guidance for evaluating CV risk in new antidiabetic therapies recommends sponsors to conduct appropriate data analysis to rule out CV safety concern for drugs treating type 2 diabetes. We derive closed form formulas for the required number of events and total exposure as functions of enrollment rate, enrollment period, delayed treatment effect and other parameters. We propose an adaptive superiority/non-inferiority trial design which is applicable to long term and large scale CV trials of antidiabetic drugs or drugs of other indications for chronic conditions. In addition, we explicate consideration on power calculation for the combined analysis of CV data from the Phase II/III program and the CV trial to rule out an excessive CV risk (risk ratio <1.8) for the initial antidiabetic New Drug Application submission. An example is used to illustrate the application of the methods.

Session: M33
Network
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Chun, Hyonho* **Affiliation:** *yale university*

Abstract Title: ***expression QTL mapping with sparse partial least squares***

Author(s): *hyonho chun, sunduz keles*

Abstract: Expression quantitative trait loci (eQTL) mapping concerns finding genomic variation to elucidate variation of expression traits. This problem poses significant challenges due to high dimensionality of both the gene expression and genomic marker data. We propose a multivariate response regression approach with simultaneous variable selection and dimension reduction for the eQTL mapping problem. Transcripts with similar expression are clustered into groups, and their expression profiles are viewed as a multivariate response. Then, we employ our recently developed sparse partial least squares regression methodology to select markers associated with each cluster of genes. We demonstrate with extensive simulations that our eQTL mapping with multivariate response sparse partial least squares regression (M-SPLS eQTL) method overcomes the issue of multiple transcript- or marker-specific analyses, thereby avoids potential elevation of Type-I error. Additionally, joint analysis of multiple transcripts by multivariate response regression increases power for detecting weak linkages.

Session: T15
Regency A
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Clark, W. Scott* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***From Mad as a Hatter to Madison and Indy: The Lilly/UW-Madison Collaboration***

Author(s): *W. Scott Clark, Ph.D.*

Abstract: The relationship between academia and industry has historically been one laden with potential, but all too often this potential goes unfulfilled because of mistrust or misaligned priorities. In the summer of 2001, the Statistical and Mathematical Sciences component of Eli Lilly and Company set out overcome this mistrust and misalignments by forging new bonds with some of the country's most esteemed statistics programs. Lilly management invited distinguished faculty from premier graduate statistics programs to a dinner during the Joint Statistical Meetings in Atlanta, Georgia. The goal of this dinner was to thank the faculty for their partnership in recruiting for Lilly and to begin discussion on how to build stronger relationships in solving statistical problems of mutual interest. However, when one distinguished faculty member spoke out against very idea of the dinner he was attending, the dreams of collaborative relationships disintegrated quickly. "The money used for this dinner

Session: M13
Network
Monday, June 21
10:20 a.m.–12:10 p.m.

should be funding my students instead," he boisterously suggested. Across the next few years, creating the environment to collaboratively work with UW-Madison has taken great effort to overcome the comments from a faculty member of the same institution. This presentation will focus on this story of what started as mistrust and has burgeoned into a model of academe/industry collaboration.

Presenter: *Connor, Jason* **Affiliation:** *Berry Consultants*

Abstract Title: ***Bayesian Adaptive Trials in Practice***

Author(s): *Jason Connor**

Abstract: This talk briefly describes Bayesian adaptive trials in general then showcases a Bayesian adaptive trial to demonstrate cardiac safety of a diabetes drug in line with FDA's 2008 Guidance. After the heart safety of rosiglitazone was questioned in 2007, FDA asked drug makers to demonstrate cardiac safety for new anti-diabetes drugs. These are complex trials whose power is proportional to event rates which are often low and unknown. If event rates are low, huge samples may be necessary to achieve the definition of safety. Therefore a classical trial with a conservative event rate assumption requires a very large number of patients. We illustrate an adaptive design that addresses the FDA guidance and uses frequent interim analyses incorporating Bayesian predictive probabilities based upon accumulating data to know when to stop accrual with just the right sample size.

Session: T21
Regency A
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Cook, R. Dennis* **Affiliation:** *University of Minnesota*

Abstract Title: ***Sufficient Dimension Reduction, Prediction and Variable Screening in High Dimensional Regressions.***

Author(s): *R. Dennis Cook* and Liliana Forzani*

Abstract: Sufficient dimension reduction methods for regression have a history of successful application over the past 15 years, including relatively recent methods for variable selection. Nearly all of these methods were designed for traditional settings in which the sample size is large relative to the number of predictors, and so may not be directly applicable in contemporary high dimensional regressions. With prediction as the ultimate goal, we will consider a class of dimension reduction methods that may be applicable in high dimensional settings. The role of variable screening and the importance of distinguishing between sparse and abundant regressions will be discussed.

Session: W28
Cosmopolitan D
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Crainiceanu, Ciprian* **Affiliation:** *Johns Hopkins University Biostatistics*

Abstract Title: ***Longitudinal Functional Principal Component Analysis***

Author(s): *Ciprian M. Crainiceanu**

Abstract: We introduce models for the analysis of functional data observed at multiple time points. The dynamic behavior of functional data is decomposed into a time-dependent population average, baseline (or static) subject-specific variability, longitudinal (or dynamic) subject-specific variability, subject-visit-specific variability and measurement error. The model can be viewed as the functional analog of the classical mixed effects model where random effects are replaced by random processes. Methods have wide applicability and are computationally feasible for moderate and large data sets. Computational feasibility is assured by using principal component bases for the functional processes. The methodology is motivated by and applied to a diffusion tensor imaging (DTI) study designed to analyze differences and changes in brain connectivity in healthy volunteers and multiple sclerosis (MS) patients.

Session: T14
Cosmopolitan C
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Craiu, Radu* **Affiliation:** *University of Toronto*

Abstract Title: ***Regional Adaptation for Adaptive MCMC***

Author(s): *Radu V. Craiu*

Abstract: Adaptive MCMC (AMCMC) algorithms allow the automatic tuning of the parameters while the simulation is in progress. A multimodal target distribution may call for regional adaptation of Metropolis-Hastings samplers in which the proposal distribution varies across

Session: M26

Cosmopolitan C
Monday, June 21
1:30 p.m.–3:20 p.m.

regions. Recent developments for adaptive regional MCMC sampling will be discussed. The methods will be illustrated using simulated as well as a real data example.

Presenter: *Crowe, Brenda* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Meta-analysis for Rare Adverse Event Data from Clinical Trials***

Author(s): *Brenda Crowe**

Abstract: An integral part of the development of a drug is the assessment of adverse events through the evaluation of multiple clinical trials via meta-analysis. There are a variety of effect measures that can be used in the analysis, for example, risk difference, risk ratio and odds ratio. Rare adverse events are particularly problematic as there is a high chance of zero events in some treatment groups or entire studies. For trials with no events, neither the risk ratio nor the odds ratio is defined. In this presentation I will discuss various statistical aspects of meta-analysis for rare adverse events. These include choice of effect measure, continuity correction and analytical method, including a discussion of fixed vs. random effects, existing methods and new advances. A direct application is the assessment of cardiovascular risk in antidiabetic medications.

Session: T32
Regency A
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Dean, Charmaine* **Affiliation:** *Simon Fraser University*

Abstract Title: ***Joint Analysis of Longitudinal Growth and Interval Censored Mortality Data***

Author(s): *Charmaine Dean*, Terry Lee, Leilei Zeng, Jason Nielsen*

Abstract: Joint analysis of longitudinal and survival data has received considerable attention in the recent literature. This talk will review methods developed for such joint analysis and develop a joint model for the analysis of longitudinal data monitoring the growth and survival of trees subject to various interventions in a designed experiment. Of interest is the development of methods to handle features of the data which are not common in considerations of joint analyses. A main feature is interval censoring of the survival response and there is also the need to account for spatial effects jointly in the survival and longitudinal outcomes. We adopt linkages in random effects over multiple outcomes to develop a joint modelling framework, and handle interval censoring in the joint longitudinal-survival context using imputation methods based on local likelihood density estimation. Properties of the EM algorithmic scheme for estimation are considered. We also discuss the conditions under which there are efficiency gains in joint analyses with regard determination of treatment effects.

Session: M12
Concept CD
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Di Lucca, Maria Anna* **Affiliation:** *M.D. Anderson Cancer Center*

Abstract Title: ***A Non-Parametric Bayesian AR Model - Application to DNA-sequencing***

Author(s): *Maria Anna Di Lucca*, Peter Mueller, Yuan Ji*

Abstract: We propose a nonparametric Bayesian autoregression for a se- quence {y_t }. We assume yt ~ Fy_{t−1} for a family of random probability measures F = {Fy ; y ∈ X }. We define a prior probability model for F using a dependent Dirichlet process (DDP) model. Specifically, we use common weights for the Fy and define the point masses as a function of y. We refer to the model as DDP-AR(1). We illustrate the model and posterior computation using base calling for Solexa se- quencing data. The application uses a mixture of four DDP-AR(1) models, one for each base, A,C,G or T. This new model allows to resolve three different sources of noise in the data: fading, phasing and cross-talk for channels which are known to affect the accuracy of the base calling. Keywords DDP, mixture model, time series, DNA sequencing, finite DP.

Session: M18
Studio One
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Dmitrienko, Alex* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Multiple testing problems with general logical restrictions in clinical trials***

Author(s): *Alex Dmitrienko**

Abstract: In this presentation we discuss multiple testing problems with general logical restrictions

Session: W12
Regency B
Wednesday, June 23
8:00 a.m.–9:50 a.m.

arising in clinical trials. The mixture-based framework can be used to construct gatekeeping procedures that account for complex logical relationships among multiple null hypotheses. These gatekeeping procedures control the familywise error rate in the strong sense and enable trial sponsors to set up powerful procedures based on p-value-based tests (eg, Hochberg-type or Hommel-type tests) or parametric tests (eg, regular and step-down Dunnett tests). The general methodology is illustrated using clinical trials with multiple dose-placebo comparisons and patient populations as well as dose-placebo comparisons and noninferiority/superiority objectives.

Presenter: *Donohue, Michael* **Affiliation:** *University of California, San Diego*
Abstract Title: ***The relative efficiency of time-to-event and longitudinal modeling***
Author(s): *Michael Donohue*, Anthony Gamst, Ronald Thomas, Paul Aisen*
Abstract:

Session: W17
Cosmopolitan D
Wednesday, June 23
8:00 a.m.–9:50 a.m.

We will explore the relative efficiency of time-to-event models, in particular the Cox Proportional Hazards model, compared to linear mixed effects models. In cases where both endpoints are available, time-to-event is often preferred because it is easy to interpret, whereas intercept or slope estimates from a linear model may be more obscure. However, in some cases this ease of interpretation may come at the expense of precise estimation of effect. We attempt to provide a theoretical quantification of this trade-off. Examples and simulations from Alzheimer's disease clinical trials are also discussed.

Presenter: *Doss, Hani* **Affiliation:** *University of Florida*
Abstract Title: ***Estimation of Large Families of Bayes Factors from Markov Chain Output***
Author(s): *Hani Doss* and Eugenia Buta*
Abstract:

Session: M26
Cosmopolitan C
Monday, June 21
1:30 p.m.–3:20 p.m.

We consider situations in Bayesian analysis where the prior is indexed by a hyperparameter taking on a continuum of values. We distinguish some arbitrary value of the hyperparameter, and consider the problem of estimating the Bayes factor for the model indexed by the hyperparameter vs. the model specified by the distinguished point, as the hyperparameter varies. We assume that we have Markov chain output from the posterior for a finite number of the priors, and develop a method for efficiently computing estimates of the entire family of Bayes factors. As an application of the ideas, we consider a standard model for variable selection in Bayesian linear regression, in which a hierarchical prior involves a prior distribution on which variables go into the model and then a prior distribution on the regression parameters for the selected variables. The hyperparameters governing this prior have a big effect on subsequent inference, including variable selection, and we show how our methodology can be used to select these hyperparameters.

Presenter: *Du, Juan* **Affiliation:** *Kansas State University*
Abstract Title: ***Fixed-domain Asymptotic Properties of Tapered Maximum Likelihood Estimators***
Author(s): *Juan Du*, Hao Zhang, V. S. Mandrekar*
Abstract:

Session: W15
Cosmopolitan C
Wednesday, June 23
8:00 a.m.–9:50 a.m.

When the spatial sample size is extremely large, as in many environmental and ecological studies, operations on the large covariance matrix are a numerical challenge. Covariance tapering is a technique to alleviate the numerical challenges. We investigate how tapering affects asymptotic efficiency of the maximum likelihood estimator (MLE) and establish asymptotic properties, particularly asymptotic distribution of the exact MLE and tapered MLE under the fixed-domain asymptotic framework for Matérn model. We show that under some conditions on the tapering function, the tapered MLE is asymptotically as efficient as the true MLE for the microergodic parameter in the Matérn model. The computational gain and comparable estimation are illustrated by simulation study and an application to the US precipitation data .

Presenter: *Epstein, Michael* **Affiliation:** *Emory University*

Abstract Title: ***A Novel Matching Approach to Correct for Population Stratification in Case-Control Association Studies***

Author(s): *Michael P. Epstein*, Richard Duncan, K. Alaine Broadaway, Andrew S. Allen, Glen A. Satten*

Abstract: Confounding due to population stratification remains an important issue when interpreting results from case-control association studies of complex disease. Fine matching of cases and controls based on genetic ancestry is an increasingly popular strategy to correct for such confounding, particularly when a study ascertains control subjects from large public databases of healthy individuals. One issue with existing matching methods is that such procedures match on a scalar measure of genetic ancestry that collapses the effects of ancestry components that are true confounders with ancestry components that are not confounders. As we show, the inclusion of ancestry components that are not confounders within the matching measure can lead to poor matches and an improper correction for confounding. To resolve this issue, we propose a novel matching method that assigns cases and controls to matched strata based on a measure called the stratification score, which is the baseline odds of disease of a subject conditional on components of ancestry. As we show, our measure weights the contribution of ancestry components based on their correlation with disease risk thereby leading to more optimal matching. We illustrate our method with an application to a case-control genomewide association study of schizophrenia from an African-American population. Within the sample, we observe confounding due to stratification that can be resolved by our stratification-score matching approach but not by other existing matching procedures. We also use simulated data under both admixed and gradient models to show our novel matching approach can provide a more appropriate correction for population stratification than both existing matching approaches as well as traditional direct-adjustment methods using principal components.

Session: M16
Regency A
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Fan, Jie* **Affiliation:** *University of Louisville*

Abstract Title: ***Mann-Whitney Test for Two Comparing Waiting Time Distributions when Transition Times are Right Censored***

Author(s): *Jie Fan * and Somnath Datta*

Abstract: We consider the problem of comparing two waiting time distributions of a transient state in a general multistate system when the transition times right censored. Under this setup the censoring induced on the weight times is complex since both the state entry and exit are subjected to right censoring. Using the reweighting principle, a Mann-Whitney generalized two sample U-statistic is constructed that compares the uncensored state waiting times from the two distributions. A second generalized Mann-Whitney statistic is also constructed that allows for comparison when one of the two waiting times is either uncensored or singly censored. While both test statistics are asymptotically unbiased under the null hypothesis and reduce to the standard Mann-Whitney statistic when there is no censoring, the second statistic turns out to have slightly larger bias while smaller variance for small to moderate sample sizes. Asymptotic normality of these test statistics are established. An application to kidney disease patients data is considered.

Session: W18
Studio One
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Fan, Yingying* **Affiliation:** *University of Southern California*

Abstract Title: ***Variable Selection in Linear Mixed Effects Models***

Author(s): *Yingying Fan*; Runze Li*

Abstract: This paper is concerned with the selection and estimation of fixed and random effects in linear mixed effects models. We propose a class of nonconcave penalized profile likelihood methods for selecting and estimating significant fixed effects parameters simultaneously. We further demonstrate that the proposed procedure enjoys the oracle property where the dimension can grow exponentially with sample size. We then propose a group variable

Session: M27
Regency B
Monday, June 21

1:30 p.m.–3:20 p.m.

selection strategy to simultaneously select and estimate the significant random effects. The resulting random effects estimator is compared with the oracle-assisted Bayes estimator. We prove that, with probability tending to one, the proposed procedure identifies all true random effects, and furthermore, that the resulting estimates are close to the oracle-assisted Bayes estimates for the selected random effects, where the dimension of the random effects coefficient vector is allowed to increase exponentially with sample size. A Monte Carlo simulation study is conducted to examine the finite sample performances of the proposed procedures. We further illustrate the proposed procedure via a real data example.

Presenter: *Feng, Yang* **Affiliation:** *UIUC*

Abstract Title: ***Randomization Inference for the Trimmed Mean of Effects Attributable to Treatment***

Author(s): *Xingdong Feng, Yang Feng*, Yuguo Chen and Dylan S. Small*

Abstract:

Randomization is described by Fisher [1] as the reasoned basis for inference about the effectiveness of treatments. Fisher advocated both using randomization in designing experiments and using "randomization inference" to analyze experiments that have been randomized. Randomization inference is inference that assumes only the physical act of randomization for its validity. It provides exact, distribution free inferences in randomized experiments. In this paper, we expand the scope of randomization inference by developing randomization inference for the trimmed mean of effects attributable to treatment. Trimmed means of the effects attributable to treatment are interpretable summaries of the treatment effect that are robust to outliers. We connect the inference problem for trimmed means of effects attributable to treatment to a multiple choice knapsack problem, and use an efficient combinatorial optimization algorithm.

Session: W14
Concept CD
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Fiecas, Mark Joseph* **Affiliation:** *Brown University*

Abstract Title: ***The Generalized Shrinkage Estimator for Spectral Analysis of Multivariate Time Series***

Author(s): *Mark Fiecas*, Hernando Ombao*

Abstract:

We develop a new statistical method for spectral analysis of neurophysiological signals represented by a multivariate time series. To perform spectral analysis, one would first need an estimate of the spectral density matrix of the multivariate time series. Parametric estimators of the spectral density matrix provide good frequency resolution but could be sensitive when the parametric model is incorrectly specified. Smoothing-based nonparametric estimators are model-free and are consistent estimators but may have poor frequency resolution. In this work, we develop the generalized shrinkage estimator, which is a weighted average of a parametric estimator and a nonparametric estimator. The optimal weights are frequency-specific and derived under the quadratic risk criterion so that the estimator that does better at a particular frequency receives heavier weight. We validate the proposed estimator in a simulation study and apply it on electroencephalogram recordings from a visual-motor experiment.

Session: M25
Studio One
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Fu, Wenjiang* **Affiliation:** *Michigan State University*

Abstract Title: ***Oligoarray data analysis incorporating probe sequence thermodynamics model***

Author(s): *Yalu Wen, Ming Li, Wenjiang Fu**

Abstract:

The genome-wide studies have challenged statisticians in recent years. Although, a large number of statistical methods have been developed for microarray data analysis, including genome-wide association studies (GWAS), many of them use only partial data from the arrays, and few use the array mechanism in analyzing the probe intensity data. In oligoarray data analysis, many use only perfect match probe intensities and discard mismatch probe data because a large portion of mismatch probes yield a larger intensity than their perfect match probes. Such an approach not only wastes data source of the extremely expensive GWAS studies but also misses important biological signals, which prompts us to ask the following questions. How to fully utilize the data, particularly the mismatch data? Can we perform

Session: W16
Network
Wednesday, June 23
8:00 a.m.–9:50 a.m.

better by incorporating the mismatch data? In this presentation, I will review current frequently used procedures, and summarize a number of important findings in the literature that have been largely overlooked. I will further present our new approach using both perfect match and mismatch data through a thermodynamic model of the binding between the probe and target sequences. Our model incorporates the array mechanism and provides a fast algorithm to analyze probe intensity data based on single array apporach. We demonstrate that using this approach, SNP genotype calling in the GWAS studies can achieve much better accuracy than other methods, while not suffering from the batch effects and genomic wave artifacts. We also discuss the potential usage of this method in other types of oligoarray data analysis.

Presenter: *Gaydos, Brenda* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***A Sample of Adaptive Dose-Finding Case Studies***

Author(s): **Gaydos, Brenda*

Abstract: An increasing number of clinical trials are simulated prior to finalizing the design. In this presentation, phase 2 design archetypes routinely considered will be discussed. Case studies will be used to briefly illustrate the deciding factors in design selection.

Session: T21
Regency A
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Goldberg, Yair* **Affiliation:** *The University of North Carolina at Chap*

Abstract Title: ***Support Vector Quantile Regression***

Author(s): *Yair Goldberg* and Michael R. Kosorok*

Abstract: We propose a novel support vector approach for non-linear censored quantile regression. Our approach is based on inverse probability of censoring weighted average. The only assumption required to ensure validity of the proposed method is independence of the survival time and the censoring, conditional on the covariates. The regression estimator is found by minimizing a convex objective function. This minimization can be performed using quadratic programming. We prove consistency and finite sample upper bounds for the proposed estimator. The simplicity of the proposed approach, its efficient computation, and the relatively weak assumptions under which this approach is valid make it a valuable alternative to existing approaches for censored quantile regression.

Session: T22
Concept AB
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Guan, Shanhong* **Affiliation:** *Merck & Co., Corp.*

Abstract Title: ***Sample size re-estimation in two-stage design using p-value combination tests***

Author(s): *Shanhong Guan*

Abstract: One of the important challenges in planning clinical trials is determining the sample size at the designing stage. Despite great effort, incorrect assumptions associated with the design parameters, for example population variance or treatment effect size, could lead to underpowered study and fail to show an effect. Several different methods have been proposed to allow for re-estimation of sample size at the interim stage of the trial. These methods naturally lead the adaptive multi-stage clincial trial designs, in which case test statistic can be constructed based on combination of the stage-wise p-values, assuming the condition of p-clud property is satisfied. In this paper, we will investigate the sample size modification method based on conditional power using different combination functions of p-values in a two-stage design setting. Numerical studies are performed to evaluate the operating characteristics of these methods. An application to the trial design is illustrated and discussion on these methods is presented.

Session: W14
Concept CD
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Guo, Wenge* **Affiliation:** *New Jersey Institute of Technology*
Abstract Title: ***Adaptive Multiple Testing Procedures under Dependence***
Author(s): *Wenge Guo*, Sanat Sarkar, Helmut Finner*
Abstract: In the context of multiple hypotheses testing, the proportion of true null hypotheses among all nulls often plays an important role, although it is generally unknown a priori. In adaptive procedures this proportion is estimated and then used to derive more powerful multiple testing procedures. Hochberg and Benjamini (1990) first presented adaptive procedures for controlling familywise error rate (FWER). However, until now, no mathematical proof has been provided to demonstrate that these procedures control the FWER. In this talk, we present new adaptive multiple testing procedures with control of the FWER under various conditions of dependence. First, we introduce a class of adaptive Bonferroni and Holm procedures including a simplified version of Hochberg and Benjamini's. In a conditional dependence model we prove that the former procedures controls the FWER in finite samples while the latter control it approximately. Second, we present new adaptive Hochberg procedures and prove they can control the FWER under positive dependence. Finally, through some simulation studies, we illustrate that these adaptive procedures are more powerful than the corresponding conventional procedures.

Session: T17
Regency B
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Ha, Il Do* **Affiliation:** *Daegu Haany University*
Abstract Title: ***Estimation of Random Effects in Semiparametric Frailty Models***
Author(s): *Il Do Ha*Florin Vaida*Youngjo Lee*Maengseok Noh*
Abstract: Semiparametric frailty models, Cox's PH models with random effects, are widely used for the analysis of clustered or correlated survival data which are often collected from time-to-event by clusters (e.g. centers). Investigation of potential heterogeneity in outcomes among clusters is important in understanding and interpretation of the variability in the data. For this the estimation (or prediction) of random effects (frailties) is useful. In this talk we propose the use of the hierarchical-likelihood (h-likelihood) procedure for frailties. We show that the h-likelihood interval for frailties can be interpreted as a frequentist confidence interval and Bayesain credible interval under a uniform prior. We also propose an adjustment of the proposed interval to avoid null intervals. Simulation studies show that our proposed interval preserves the nominal confidence level. The procedure is illustrated using survival data from a multi-center lung cancer clinical trial.

Session: W17
Cosmopolitan D
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Haraldsdottir, Ragnheidur* **Affiliation:** *Columbia University*
Abstract Title: ***Functional Mixed-Effects Model for Analyzing fMRI Pain Data***
Author(s): *Haraldsdottir*, Ragnheidur and Lindquist, Martin*
Abstract: In this work we introduce a functional mixed-effects approach for analyzing fMRI data. For each individual the goal is to model the fMRI response curve using a series of basis functions corresponding to main effect, subject effect and condition effect. We derive an EM-algorithm to efficiently estimate the model parameters. We apply this method on data from a study of overt vs. covert administration of remifentanil, an opiate analgesic, on brain responses to noxious thermal stimulation. This method allows us to decompose the fMRI response into subject specific, drug specific and temperature specific components in a flexible manner without relying on a presumed hemodynamic response function.

Session: M25
Studio One
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Harvill, Jane* **Affiliation:** *Baylor University*
Abstract Title: ***Bispectral-based methods for clustering nonlinear time series***
Author(s): *Jane Harvill*, Bonnie Ray, Nalini Ravishankar*
Abstract: It is well-known that in general, second-order properties are insufficient for describing nonlinear time series. In particular, it is easily shown that the normalized bispectral density function is constant for a linear, Gaussian series, but typically not for nonlinear series. Furthermore different nonlinear time series models have different bispectral signatures.

Session: M34
Concept CD

Monday, June 21
3:40 p.m.–5:30 p.m.

Based on these properties, we propose the use of various distance measures computed on the square modulus of the estimated normalized bispectrum as a means for clustering nonlinear series. In this presentation, we empirically investigate misclassification rates of the various distance measures. We then apply the methods to a set of EEG data for a single patient collected at six different locations in the brain during an epileptic seizure.

Presenter: *He, Yulei* **Affiliation:** *Harvard Medical School*

Abstract Title: ***Assessing Geographical Variations in Hospital Processes of Care Using Multilevel Item Response Models***

Author(s): *Yulei He*, Robert E. Wolf, and Sharon-Lise T. Normand*

Abstract: National effort is directed toward developing and disseminating comparative information on some standardized processes of care for health care providers. We propose the use of Bayesian multilevel item response models to estimate hospital quality from multiple process measures and to assess their geographical variations. This approach fully incorporates the nesting structure of measures, patients, hospitals, and various levels of geographical units to provide a summary of the hospital quality. We apply the method to a national dataset of patients treated for a heart attack, heart failure, or pneumonia. We demonstrate considerable geographical differences in the quality of hospital care in these conditions. The variations across census regions and states accounted for slightly more than 10% of the total variation. Some states performed well for all three conditions (e.g., the respective posterior probability of having better than the national average performance was 1 or close to 1 in Iowa, New Jersey, South Dakota, and Wisconsin). In contrast, some states varied across conditions (e.g., the corresponding posterior probability was close to 1 in Massachusetts for the care of heart attack and heart failure, but reduced to less than 0.5 for the care of pneumonia). The study results provide a comprehensive picture of hospital comparison at both regional and national level, and might be informative for policy development.

Session: T16
Studio One
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *He, Yunxiao* **Affiliation:** *Yale University*

Abstract Title: ***The Dynamic ECME Algorithm for Accelerating the EM Algorithm***

Author(s): *Yunxiao He*, Chuanhai Liu*

Abstract: The ECME algorithm has proven to be an effective way of accelerating the EM algorithm for many problems. Recognizing the limitation of using prefixed acceleration subspaces in ECME, we propose a new Dynamic ECME (DECME) algorithm which allows the acceleration subspaces to be chosen dynamically. Our investigation of the classical Successive Overrelaxation (SOR) method, which can be considered as a special case of DECME, leads to an efficient, simple, stable, and widely applicable DECME implementation, called DECME_v1. The fast convergence of DECME_v1 is established by the theoretical result that, in a small neighborhood of the maximum likelihood estimate (MLE), DECME_v1 is equivalent to a conjugate direction method. Numerical results show that DECME_v1 and its two variants often converge faster than EM by a factor of one hundred in terms of number of iterations and a factor of thirty in terms of CPU time when EM is very slow.

Session: T33
Network
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *He, Yunxiao* **Affiliation:** *Yale University*

Abstract Title: ***Variable Selection with Prior Information via the pLasso Method***

Author(s): *Yunxiao He*, Yuan Jiang, Heping Zhang*

Abstract: The Lasso is a popular model selection tool which is often used in conjunction with generalized linear models. When the number of parameters is much larger than the sample size, as often is in many biomedical studies, the power of Lasso can be very limited. On the other hand, intensive biology and biomedical researches have provided large amount of plausible information about the possible significance of certain predictive variables. This paper proposes an extension of Lasso, called pLasso (prior Lasso), to incorporate the

Session: W27
Cosmopolitan C
Wednesday, June 23
10:10 a.m.–12:00 p.m.

aforementioned prior information in variable selection for generalized linear models. The goal is achieved by further penalizing the criterion function of Lasso with a measure of the discrepancy between the prior information and the information from data. Asymptotic theory and simulation result show that pLasso provides significant improvement over Lasso in terms of the ability to identify the true model when the prior information is relatively accurate. When the prior information is less reliable, pLasso shows great robustness from being distracted.

Presenter: *He, Xu* **Affiliation:** *University of Wisconsin-Madison*

Abstract Title: ***Subgroups defined by two or more genetic markers in genome-wide association studies***

Author(s): *Xu He*, Wei-Yin Loh and Michael Man*

Abstract: Correct identification of the important genetic markers in a genome-wide association study is a formidable task due to the high chance of false positives. This article considers the particular problem of identifying the subgroups and their associated markers that yield large absolute risk reductions in a placebo-controlled setting. Several promising methods are proposed and examined for detecting subgroups defined by two or more of such markers. The methods employ LOWESS smoothing, linear discriminant analysis, importance scoring from a classification tree algorithm, and random data perturbation. Results from simulation experiments demonstrating the effectiveness of the methods are reported.

Session: M13
Network
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Holland, Chris* **Affiliation:** *MacroGenics*

Abstract Title: ***Statistical Give and Take: Power Implications on Strong Control of the Type I Error Rate in a Clinical Trial Setting***

Author(s): *Chris Holland**

Abstract: Much progress has been made over the past decade with the development of gatekeeping statistical procedures that provide strong control of the Type I error rate in increasingly complex clinical trial study designs that require, for example, multiple dose groups and co-primary endpoints to undergo formal hypothesis testing. The advantage that such procedures provide is greater assurance that positive results will make it into an approved product's labeling and can therefore be used for marketing claims. The disadvantage, of course, is the need to increase the sample size in order to provide adequate power for meeting a study's key objectives. However, the true power implications between competing procedures or approaches are not always well understood. This is a particularly relevant problem in situations where strong Type I error control is more of a "nice to have" than a regulatory requirement, such as when dealing with key secondary endpoints. In this presentation we will look at one such procedure, the truncated Holm test, applied to a situation that involves multiple treatment groups and co-primary endpoints. The properties of this "separable" test will be examined with respect to varying truncation fractions and Type II error rates compared to a standard Holm procedure (which is not separable and therefore does not provide strong control of the Type I error).

Session: W12
Regency B
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Hong, Quan* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Modeling analysis of longitudinal time-couse data in clinical trials can significantly improve efficiency***

Author(s): *Quan Hong*

Abstract: In longitudinal clinical trials, response measurements are collected at multiple time points. In practice, these data are often analyzed using mixed-effect models and conclusions are drawn by comparing treatment responses at the endpoint. In some cases, the time courses of the responses clearly follow certain shape or pattern, such as a monotone exponential shape. Thus, the longitudinal data can be more efficiently analyzed through modeling of the longitudinal profiles of treatment responses. In doing so, data measured at one time point can effectively influence estimation of responses at other time points and result in more

Session: W22
Concept AB
Wednesday, June 23
10:10 a.m.–12:00 p.m.

efficient and robust estimators. Simulations are done based on clinical examples that demonstrate that such an approach can dramatically improve power over typical mixed-effect model analysis.

Presenter: *Hosmane, Balakrishna* **Affiliation:** *Northern Illinois University*

Abstract Title: ***Exposure-Response Modeling of QT Effect in 'Thorough' QT/QTc Studies***

Author(s): *Balakrishna Hosmane*, Charles Locke, Yi-Lin Chiu*

Abstract:

Session: T13
Network
Tuesday, June 22
10:20 a.m.–12:10 p.m.

A potential problem with some drugs is their ability to delay cardiac repolarization, a condition that may trigger a cardiac arrest. Over the past decade, these cardiac adverse events have resulted in a number of patient deaths and either the subsequent removal from the market of previously approved drugs (e.g., terfenadine and cisapride) or the placement of restrictions on their use (e.g., ziprasidone and ranolizine). The delay of cardiac repolarization is defined as the QT interval and is measured using the patient's electrocardiogram (ECG). Drugs are expected to receive a clinical ECG evaluation, as mandated by the International Conference on Harmonization (ICH E14) guidance. Such an evaluation of the effect of a drug on the QT interval is referred to as a 'thorough QT/QTc study' in the biopharmaceutical literature. We assess the QT effect using an exposure-response model in a 'thorough QT/QTc study' with a four-period crossover design in which the treatments are placebo, positive control, higher dose of investigational drug, and therapeutic dose of investigational drug.

Presenter: *Hsu, Chyi-Hung* **Affiliation:** *Novartis Pharmaceuticals*

Abstract Title: ***Biomarker-based adaptive dose-finding trials: A case study in Phase II Oncology***

Author(s): *Chyi-Hung Hsu*, Jose Pinheiro*

Abstract:

Session: W22
Concept AB
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Dose-regimen selection for confirmatory trials and characterization of dose-response relationship are arguably among the most important and difficult tasks in clinical drug development. By allowing midtrial adjustments as information accumulates, adaptive designs provide a flexible framework for improving the knowledge efficiency of clinical trials, being particularly useful in the context of dose-finding studies. This presentation will discuss some of the challenges of dose-finding in oncology and how biomarkers can enable the use of adaptive trials in this context. Some of the difficulties involved include the use of model-based approaches for interim decision making, for both adapting and possibly stopping for futility, under a priori model uncertainty. An extension of the MCP-Mod methodology for time-to-event data will be presented and illustrated via a case study from a real phase II oncology trial.

Presenter: *Hu, Ming* **Affiliation:** *University of Michigan*

Abstract Title: ***Model-based methods for analyzing ChIP sequencing data***

Author(s): *Ming Hu, Jindan Yu, Jeremy Taylor, Arul Chinnaiyan, Zhaohui Qin*

Abstract:

Session: T25
Network
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Protein-DNA interaction constitutes a basic mechanism for genetic regulation of target gene expression. Deciphering this mechanism is challenging due to the difficulty in characterizing protein-bound DNA on a genomic scale. The recent arrival of ultra-high throughput sequencing technologies has revolutionized this field by allowing quantitative sequencing analysis of target DNAs in a rapid and cost-effective way named ChIP-Seq. The rapid accumulation of ChIP-Seq data has created a daunting analysis challenge. Here we will discuss two analysis problems involving ChIP-Seq experiments. The first is on integration of data from both ChIP-Seq and ChIP-chip experiments; the second is on post-process ChIP-Seq data to detect and refine the transcription factor binding motif pattern.

Presenter: *Huang, Chiang-Ching* **Affiliation:** *Northwestern University*

Abstract Title: ***Methylation detection call for whole genome methylation data***

Author(s): *Chiang-Ching Huang*; Pan Du; Simon Lin*

Abstract: DNA high---throughput methylation profiling is crucial to understand epigenetic control of transcriptomics and genomic stability. Illumina methylation assay has been widely used in

Session: T25
Network
Tuesday, June 22
1:30 p.m.–3:20 p.m.

epigenomic study due to its quality, low cost, and small sample requirement. However, because the total amount of CpG methylation can differ significantly between samples, especially between tumors and normal controls, many assumptions used by mRNA microarray preprocessing are not valid in processing DNA methylation data. We will present the analytic challenges in identifying differential methylation levels and use a mixture of truncated Gamma distributions to model the methylation beta values to determine the within-sample methylation calls. A methylation titration data set will be used to demonstrate the sensitivity of our modeling detection calls

Presenter: *Huang, Chunfeng* **Affiliation:** *Indiana University*

Abstract Title: ***Nonparametric variogram estimation on the sphere***

Author(s): *Haimeng Zhang, Scott Robeson*

Abstract:

Session: W15
Cosmopolitan C
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Abstract: In this talk, we first investigate the validity of commonly used variogram functions on the sphere. In particular, we show that the spherical and exponential variogram functions, as well as power variogram with $0 < alpha le 1$, are valid on the sphere. However, two radon transforms of the exponential model, Cauchy model, the hole-effect model and power variogram with $1 < alpha le 2$ are not valid on the sphere. Then, a nonparametric method to estimate the variogram on the sphere is proposed. We apply our method to Microwave Sounding Unit (MSU) data. In addition, the range and sill estimation is discussed.

Presenter: *Hung, H.M. James* **Affiliation:** *FDA*

Abstract Title: ***Roles of Adaptive Design in Clinical Program***

Author(s): *H.M. James Hung, Sue-Jane Wang*

Abstract:

Session: T21
Regency A
Tuesday, June 22
1:30 p.m.–3:20 p.m.

The recent advances on clinical trial methodology provide an opportunity to take a fresh look of fixed designs, group sequential designs and a broader class of flexible designs. The FDA draft adaptive design guidance lays out some pivotal foundations for adaptive designs to be properly considered in clinical development programs. In this presentation, I shall revisit the thought process necessary to entertain usage of adaptive designs in the clinical development program, distinguishing between learning stage and confirmatory stage. Topics include statistical modeling with adaptation, sample size planning, sample size re-estimation to a mid-term change of statistical decision tree based on interim information accumulated from the internal data or external data, logistics issues needing resolution. Some regulatory experiences will be shared.

Presenter: *James, Gareth* **Affiliation:** *University of Southern California*

Abstract Title: ***Functional Additive Regression***

Author(s): *Gareth James* and Yingying Fan*

Abstract:

Session: T34
Concept CD
Tuesday, June 22
3:40 p.m.–5:30 p.m.

We suggest a new method, called "Functional Additive Regression", or FAR, for efficiently performing high dimensional functional regression. FAR extends the usual linear regression model involving a functional predictor, X(t), and a scalar response, Y, in two key respects. First, FAR uses a penalized least squares optimization approach to efficiently deal with high dimensional problems involving a large number of different functional predictors. Second, FAR extends beyond the standard linear regression setting to fit general non-linear additive models. We demonstrate that FAR can be implemented with a wide range of penalty functions using a highly efficient coordinate descent algorithm. Theoretical results are developed which provide motivation for the FAR optimization criterion.

Presenter: *Jeng, X Jessie* **Affiliation:** *University of Pennsylvania*

Abstract Title: ***Optimal Sparse Segment Identification with Application in Copy Number Variation Analysis***

Author(s): *X. Jessie Jeng*, T. Tony Cai and Hongzhe Li*

Abstract: Motivated by DNA copy number variation (CNV) analysis based on high-density single nucleotide polymorphism (SNP) data, we consider the problem of detecting and identifying

Session: T35
Regency B
Tuesday, June 22
3:40 p.m.–5:30 p.m.

sparse short segments in a long one-dimensional sequence of data with additive Gaussian white noise, where the number, length and location of the segments are unknown. We present a statistical characterization of the identifiable region of a segment where it is possible to reliably separate the segment from noise. An efficient likelihood ratio selection (LRS) procedure for identifying the segments is developed, and the asymptotic optimality of this method is presented in the sense that the LRS can separate the signal segments from the noise as long as the signal segments are in the identifiable regions. The proposed method is demonstrated with simulations and analysis of a real data set on identification of copy number variants based on high-density SNP data. The results show that the LRS procedure can yield greater gain in power for detecting the true segments than some standard signal identification methods.

Presenter: *Jeong, Jong-Hyeon* **Affiliation:** *University of Pittsburgh*

Abstract Title: ***Bivariate Analysis of Competing Risks Data***

Author(s): *Jong-Hyeon Jeong**

Abstract:

Session: W17
Cosmopolitan D
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Competing risks data have been analyzed by modeling a subdistribution of cause-specific events marginally, assuming only fixed correlation structures with a subdistribution of other competing events. In practice, however, the correlation structure could be arbitrary, so that it needs to be incorporated into the inference procedure. In this talk, we will discuss parametric and nonparametric inferences on a cause-specific subdistribution considering such correlation structures. We will also discuss an extension to bivariate survival data with competing risks.

Presenter: *Ji, Yuan* **Affiliation:** *M. D. Anderson Cancer Center*

Abstract Title: ***Bayesian base calling for Solexa sequencing data***

Author(s): *Yuan Ji*, Fernando Quintana, Riten Mitra, Alejandro Jara, Ping Liu, Yue Lu, and Shoudan Liang*

Abstract:

Session: T37
Cosmopolitan D
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Next-generation sequencing is a new revolutionary tool in modern biological research that allows researchers to acquire "digital" information about the sequence and expression of DNA and RNA. Base cal ling is a critical step in the Solexa next-generation sequencing procedure, which produces estimates of base calls A, C, G, or T for each genome location by comparing the location- specific intensity measurements that reflect the signal strength of four possible nucleotides at each base. The Bayesian method builds on a hierarchical model that accounts for three sources of noise in the data: fading, phasing, and cross-talk of channels, which are known to affect the accuracy in the base calling. We show that the new method improves the precision of base-calling compared with the Solexa calling embedded in their commercial software. Furthermore, the proposed method provides a probability score that measures the confidence of each base call. This probability score can be used to estimate the false discovery rate of the base calling or to rank the precision of the estimated DNA sequences, which can be useful for downstream analysis such as sequence alignment. We make available R and WinBUGS programs that can be downloaded freely.

Presenter: *Jia, Haomiao* **Affiliation:** *Columbia University*

Abstract Title: ***Rapid Response Health Surveillance and the Utility of Small Area Estimates: Methods for Estimating County-level Outcome***

Author(s): *Haomiao Jia*

Abstract:

Session: T31
Concept AB
Tuesday, June 22
3:40 p.m.–5:30 p.m.

In response to emerging public health emergencies, such as the 2009 H1NI influenza outbreak, the CDC had added questions to the ongoing state-based Behavioral Risk Factor Surveillance System (BRFSS) for tracking impacts on population health and evaluating the effective of recommendations to public health officials. It is critical to obtain rapid assessment of local responses with valid and reliable monthly (and weekly) county-level estimates for mounting an effective response, such as redistributing resources and developing

effective messages to inform the public. This study delineates how the BRFSS was used for rapid health surveillance and highlight the importance of small-area estimates (SAEs) in addressing emerging health emergencies. We examined the challenges and requirements to obtain "real-time" county-level estimates with limited or no data in some areas. We surveyed some SAE procedures used at the CDC and the State Health Departments. Specifically, we examined a two-step procedure suitable for fast SAE. In step 1, we used a linear or generalized linear mixed model for the initial estimations. In step 2, we use a nonparametric spatial smoothing method to further improve the reliability of estimates and to identify spatial clustering of outcomes. In the end, we summarize the methodology for the construction of monthly vaccination rates that could be updated steadily and investigation of the differences between counties during the 2004-08 influenza vaccine shortage. The method also is adaptable to spatial-temporal mapping using the GIS system, which can provide visual scanning for clusters of hotspots having low vaccination coverage.

Presenter: *Jiang, Hongmei* **Affiliation:** *Northwestern University*

Abstract Title: ***Statistical issues for analysis of metagenomics sequencing data***

Author(s): *Hongmei Jiang*, Lingling An*

Abstract: The next generation sequencing technology has enabled the rapid sequencing and analysis of mixed genomes sampled directly from the environment, which is recently emerged as metagenomics and applied in different fields including human health, environmental remediation, and agriculture. By direct sequencing, researchers can study organisms that are not easily cultured or even cannot be cultured at all in the laboratory. Based on the sequence data from a metagenomic sample, we would like to address the following basic questions "what species or genomes are there?", and "what are their relative abundance?" One approach to answer these questions is to use sequence homology by aligning sequence reads to known reference sequences using a sequence comparison program such as the basic local alignment search tool (BLAST) and assigning taxonomy based on the best match or multiple high-scoring BLAST hits. We propose one method to estimate the multiple genomes and their relative abundance within a metagenomics sample by modeling the results from BLAST. The current statistical and computational methods that are being developed to analyze the metagnoimcs data and the challenges will also be highlighted in the talk.

Session: M35
Regency B
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Jiang, Yuan* **Affiliation:** *Yale University*

Abstract Title: ***Adaptive Lasso for High-Dimensional Models Under A Class of Convex Loss Functions***

Author(s): *Yuan Jiang* and Chunming Zhang*

Abstract: The adaptive Lasso method uses adaptive weights for the L1 penalties to revise the traditional Lasso in the linear regression model under a quadratic loss. We investigate applications of the adaptive Lasso to high-dimensional models for regression and classification under a wide class of convex loss functions. We show that for the dimension growing nearly exponentially with the sample size, the resulting adaptive Lasso estimator possesses the oracle property for suitable weights. Moreover, we propose two methods, called componentwise regression (CR) and penalized componentwise regression (PCR), for estimating weights. Theoretical advantages of PCR over CR are analyzed. In addition, the adaptive Lasso classifier is shown to be consistent to the optimal Bayes rule. Simulation studies demonstrate the advantage of PCR over CR in both regression and classification. The effectiveness of the proposed method is illustrated using real data sets. (This is a joint work with Professor Chunming Zhang from University of Wisconsin, Madison.)

Session: T18
Cosmopolitan D
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Jing, Bing-Yi* **Affiliation:** *Hong Kong Univ. of Science & Technology*

Abstract Title: ***On the jump activity index for semimartingales***

Author(s): *Bing-Yi JING, Xingbing KONG, Zhi LIU*

Abstract: Empirical evidence of asset price discontinuities or jumps in financial markets has been well

Session: T24
Concept CD
Tuesday, June 22
1:30 p.m.–3:20 p.m.

documented in the literature. Recently, Ait-Sahalia and Jacod (2009) defined a general jump activity index to describe the degree of jump activities for semimartingales. An estimator of this index was also given and proven to be consistent and asymptotically normal. However, only the extremely large increments were used in their estimator so that the effective sample size is very small, resulting in large standard errors even for very large sample sizes. The purpose of this paper is to remedy this problem by proposing a new estimator of the jump activity index, which makes fuller use of increments. The newly proposed estimator shares similar asymptotic properties as that by Ait-Sahalia and Jacod (2009), but always has a smaller variance and MSE. Simulations justify the improvement.

Presenter: *Jones, Galin* **Affiliation:** *University of Minnesota*
Abstract Title: ***Variable at-a-time Markov Chain Monte Carlo***
Author(s): *Jones*
Abstract:

Session: M26
Cosmopolitan C
Monday, June 21
1:30 p.m.–3:20 p.m.

It is common practice in Markov chain Monte Carlo to update a high-dimensional chain one variable (or sub-block of variables) at a time, rather than conduct a single block update. When it is possible to draw from each full conditional distribution associated with the target this is just a Gibbs sampler. However, very often at least one of the Gibbs updates is replaced with a Metropolis-Hastings step yielding a Metropolis-Hastings-within-Gibbs strategy. While these strategies can make the implementation of MCMC easier, the theoretical convergence properties of the associated Markov chain have received limited attention. We present conditions under which the chain converges to its stationary distribution at a geometric rate. We pay particular attention to the Gibbs sampler and the case with state-independent component-wise proposals. We illustrate our results in two examples, a toy Bayesian inference problem and a practically relevant example involving maximum likelihood estimation for a generalized linear mixed model.

Presenter: *Kaizar, Eloise* **Affiliation:** *Ohio State University*
Abstract Title: ***Make No Mistake: Stepwise Multivariate Permutation Tests May Not Control Multiple Testing Error Rates***
Author(s): *Eloise Kaizar*, Jason Hsu, Yan Li*
Abstract:

Session: T17
Regency B
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Genomewide Association Studies (GWAS) often include multiple tests of independence between a large number of biomarkers and a single discrete phenotype. The large number of hypotheses to be tested coupled with current computational limitations necessitate the use of "shortcut" methods that make testing feasible. Step-down methods such as Holm's procedure are one common approach to shortcutting. Unfortunately, many of these methods do not incorporate correlation between test statistics, and thus tend to control multiple testing error rates at conservative levels. Multivariate permutation-based step-down tests have been proposed as one nonparametric solution to incorporate correlation into multiple testing procedures. Similar to the case of continuous response, multivariate permutation testing for discrete responses may not control FWER. Distinct from the continuous case, the discrete setting allows us to also consider the conditional FWER. We show that the conditional p-values reported under permutation testing can be surprisingly misleading, and as a result caution against the use of multivariate permutation tests to control the error rate in discrete data GWAS.

Presenter: *Kim, KyungMann* **Affiliation:** *Univerrsity of Wisconsin-Madison*
Abstract Title: ***A Bayesian Covariate-Adjusted Response-Adaptive Design with Biomarkers for Targeted Therapies in Cancer***
Author(s): *KyungMann Kim* and Jens Eickhoff*
Abstract:

Session: M21
Network

Pharmacogenomic biomarkers are a critical component of a targeted therapy as they can be used to identify patients who are more likely to benefit from treatment. New study designs may be helpful which can effectively evaluate both the prognosis based on

Monday, June 21
1:30 p.m.–3:20 p.m.

pharmacogenomic biomarkers and the response from the therapeutic intervention with targeted therapies. We propose a Bayesian response-adaptive design which utilizes individual pharmacogenomic profiles and patients' clinical outcomes as they became available during the course of the trial to assign most effective treatment to individual patients. A series of simulation studies were conducted to examine the operating characteristics of the proposed study design. The simulation studies show that the proposed design identifies patients who are more likely to benefit from a targeted therapy and that there are substantial savings in the sample size requirements.

Presenter: *Kim, Seongho* **Affiliation:** *University of Louisville*

Abstract Title: ***A Novel Bayesian Markov Chain Monte Carlo (MCMC) Method for Nonlinear Pharmacokinetics Models***

Author(s): *Seongho Kim* and Lang Li*

Abstract: The convergence rates of two different Bayesian MCMC schemes, single component MCMC (SCM) and group component MCMC (GCM), are compared given prior information. In this study, we verify that GCM is more efficient on identifiable model and SCM is more efficient on non-identifiable models. We then developed a novel MCMC scheme by taking the advantages of SCM and GCM. The simulation and experiment studies show that the proposed approach is more robust and efficient others.

Session: T23
Regency B
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Kong, Xinbing* **Affiliation:** *HKUST*

Abstract Title: ***Is there evidence for the high frequency data being purely discontinuous?***

Author(s): *Jing Bing-Yi, Kong* Xinbing, Liu Zhi*

Abstract: It is widely accepted that the asset price process consists of a jump component because of the heavy tails of the asset returns. Many literatures, theoretically or empirically, even suggested purely discontinuous process as the underlying model of asset prices. Is there any evidence from the high frequency financial data supporting the purely discontinuous models? or equivalently, does the continuous diffusion component vanish or not? In this paper, we formally propose a statistical test against the necessity of a continuous diffusion component. Under the null hypothesis that a continuous diffusion component does not vanish, our test statistics has a CLT when the discontinuous process has infinite variation, which facilitates control of type I error asymptotically. This is the advantage of our test over other tests developed recently. Simulation shows that the asymptotic size is close to the true value, and the test is powerful.

Session: W26
Studio One
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Laber, Eric* **Affiliation:** *University of Michigan*

Abstract Title: ***Statistical Inference In Dynamic Treatment Regimes via Q-Learning***

Author(s): *Eric B. Laber*, Min Qian, Susan A. Murphy*

Abstract: Clinical researchers wanting to make principled (evidence based) rules for tailoring treatment have run multi-stage randomized clinical trials in order to evaluate and compare different long-term treatment strategies. Q-learning and other extensions of regression to multi-stage data can be used effectively to construct decision rules that lead to a favorable clinical outcome. However, in order for these methods to be more widely adopted, one must be able to conduct statistical inference. In particular, one must be able to able to answer the following types of questions: Do two treatment strategies result in significantly different clinical outcomes? Which patient variables are relevant for tailoring treatment? The need to address these questions has led us to develop new statistical methodology that allows for the construction of valid confidence intervals for parameters in the regression models used in Q-learning. The focus of this talk will be on the use of this new methodology to construct meaningful and interpretable regression models via Q-learning and the application of confidence intervals to glean relevant scientific knowledge. We illustrate these ideas by means of a case study of the Study of the Adaptive Interventions for Children with ADHD

Session: T22
Concept AB
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Trial.

Presenter: *Lazar, Nicole* **Affiliation:** *University of Georgia*

Abstract Title: ***On the Use of Empirical Likelihood for the Analysis of Longitudinal Data***

Author(s): *Nicole Lazar*, Jien Chen*

Abstract: Longitudinal data are commonplace in medical, psychological, and sociological applications. As such, many methods, both parametric and non-parametric, have been developed for their analysis. In this talk, I discuss the use of the non-parametric method of empirical likelihood (EL) for analyzing longitudinal data. This topic has generated interest of late, with several authors exploring EL in the longitudinal data setting. I will focus on the use of EL specifically for selecting the covariance structure imposed on the model for longitudinal data.

Session: W25
Concept CD
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *LeBlanc, Michael* **Affiliation:** *Fred Hutchinson Cancer Research Center*

Abstract Title: ***Interactions in Association Models***

Author(s): *Michael LeBlanc*

Abstract: Genome-wide association studies, in which hundreds of thousands of single nucleotide polymorphisms (SNP) are measured large cohorts of subjects are now being analyzed. It is believed that interactions between SNPs or genes and interactions between genes and environmental factors could substantially contribute to the genetic risk of a disease. We describe strategies to identify drug/environment x gene or gene x gene interactions in high dimensional problems. We also consider methods that utilize potential gene × drug/environment or gene x gene or interactions in the data to increase the feasibility of detecting marginal associations. For instance, if not acknowledged in the analysis method, interactions can lead to reduced marginal effect size and hence reduce the power to detect true associations. (This is joint work with Charles Kooperberg)

Session: W16
Network
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Lee, J. Jack* **Affiliation:** *Univ. of Texas M. D. Anderson Cancer Ctr*

Abstract Title: ***Predictive Probability and Adaptive Randomization in Phase II Clinical Trials***

Author(s): *J. Jack Lee*, Guosheng Yin, Nan Chen*

Abstract: We propose a randomized phase II clinical trial design based on the Bayesian adaptive randomization scheme and predictive probability monitoring (BARPP). Adaptive randomization assigns more patients to a more efficacious treatment arm based on comparing the posterior probabilities of the efficacy between different arms. We continuously monitor the trial using the predictive probability. The trial is terminated early when it is shown that one treatment is overwhelmingly superior to the others or that all the treatments are equivalent. By coupling adaptive randomization and predictive probability approaches, the trial can assign more patients to a more efficacious treatment and allow for early stopping whenever sufficient information is obtained to conclude treatment superiority or equivalence. The design is efficient, flexible, and ethical. It controls both type I and type II errors and compares favorably with frequentist designs.

Session: W11
Regency A
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Lee, Yoonkyung* **Affiliation:** *Ohio State University*

Abstract Title: ***Does Modelling Lead to More Accurate Classification?***

Author(s): *Yoonkyung Lee* and Rui Wang*

Abstract: Classification is an important statistical problem with a wide range of applications. A variety of statistical tools have been developed for learning a classification rule from data. Understanding of their relative merits and comparisons help users to choose a proper method in practice. This talk focuses on comparison of model-based classification methods in statistics with algorithmic methods in machine learning in terms of the error rate. Extending Efron's comparison of logistic regression with the LDA under the normal setting, we contrast the support vector machine and boosting with the LDA and logistic regression and study their relative efficiencies based on the limiting behaviour of the classification boundary of each method. In addition to the theoretical study, we carry out numerical experiments for

Session: M15
Cosmopolitan D
Monday, June 21
10:20 a.m.–12:10 p.m.

more comprehensive comparison under different settings than the normal setting.

Presenter: *Levine, Michael* **Affiliation:** *Purdue University*

Abstract Title: ***MIXING DENSITY ESTIMATION USING THE NONPARAMETRIC PENALIZED LIKELIHOOD MAXIMIZATION***

Author(s): *Lei Liu, Michael Levine* and Yu Zhu*

Abstract: The problem of estimating continuous mixing density is of considerable practical importance. It is needed when estimating fluorescence life-times which is important in chemistry and biology. Continuous mixture models are also routinely used to model radioactive decay and many other practically important processes. It is well known that, when the true mixing distribution is continuous, its nonparametric maximum likelihood is degenerate. We propose to estimate the mixing density by maximizing a penalized likelihood and call the resulting estimate the nonparametric maximum penalized likelihood estimate (NPMPLE).Using theory and methods from the calculus of variations and differential equations, a new functional EM algorithm is derived for computing NPMPLE of the mixing density. In the algorithm, maximizers in M-steps are found by solving an ordinary differential equation with boundary conditions numerically. We verify that the proposed algorithm is a true EM algorithm by proving its ascent property. It is demonstrated by using simulation studies that the new algorithm outperforms other existing methods, such as the popular EMS algorithm.

Session: T38
Studio One
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Li, Yun* **Affiliation:** *University of North Carolina*

Abstract Title: ***Benefit from Sequencing Data in the Public Domain: In Silico CNP Genotyping from SNP Genotypes***

Author(s): *Yun Li, Robert E Handsaker, Steven A McCarroll, Gonçalo R Abecasis*

Abstract: Rapid advances in massively parallel sequencing technologies have made it possible to study the whole genome with digital resolution. We propose a method that integrates read depth, insert size, and LD among SNPs and CNPs to obtain highly accurate CNP genotypes. We applied our method to data from the 1000 Genomes Project with an average depth of ^5X. We first obtained preliminary SNP and CNP likelihoods and then refined genotype calls using LD among them. Comparing with the experimental counterparts, we show CNPs can be genotyped with a call rate of 99.30% and an error rate of 1.2% (at a LOD threshold of 1.3, corresponding to a 20X likelihood ratio in favor of the most likely CNP genotype). The refinement step improved upon a preliminary call rate of 98.2% and an error rate of 3.7%. Furthermore, the reconstructed haplotypes encompassing both SNPs and CNPs allow the prediction of CNP genotypes from SNP data alone in an independent sample. Our preliminary analysis estimated that ^77.7% (46.6%) of > 500,000 small indels can be predicted with r2 > 0.5 (0.8) based on SNP genotypes alone when using the above reconstructed haplotypes as external reference. We thus demonstrate that joint analysis of CNP and SNP data from sequencing data can generate more complete and accurate CNP genotypes and allow prediction of CNP genotypes from SNP genotypes. Our method is implemented in C/C++ and is at www.sph.umich.edu/csg/yli/mach/.

Session: T25
Network
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Li, Fang* **Affiliation:** *IUPUI*

Abstract Title: ***Comparing two nonparametric regression curves in the presence of long memory in covariates and errors***

Author(s): *Hira Koul, Fang Li **

Abstract: This paper discusses the problem of testing the equality of two nonparametric regression functions against two-sided alternatives in the presence of long memory in the common covariate and errors. The proposed test is based on a marked empirical process of the differences between the response variables. We discuss asymptotic null distribution of this process and consistency of the test for a class of general alternatives. We also conduct Monte

Session: M34
Concept CD
Monday, June 21
3:40 p.m.–5:30 p.m.

Carlo simulation to study the finite sample and power behavior of the test.

Presenter: *Li, Linyuan* **Affiliation:** *University of New Hampshire*

Abstract Title: ***Rate-Optimal Wavelet Estimation of Mean Regression with Long Memory Infinite Moving Average Errors***

Author(s): *Linyuan Li*

Abstract: We consider the wavelet-based estimators of mean regression function with long memory infinite moving average errors and investigate the rates of convergence of estimators based on thresholding of empirical wavelet coefficients. We show that these estimators achieve nearly optimal minimax convergence rates within a logarithmic term over a large class of function space. Therefore, in the presence of long memory moving average noise, wavelet estimators still achieve nearly optimal convergence rates. The theory is illustrated with some numerical examples.

Session: T24
Concept CD
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Li, Lang* **Affiliation:** *Indiana University School of Medicine*

Abstract Title: ***STAT1 regulates microRNA transcription in interferon γ – stimulated HeLa cells***

Author(s): *Guohua Wang, Yadong Wang, Mingxiang Teng, Denan Zhang, Yunlong Liu and Lang Li**

Abstract: Background: Constructing and modeling gene regulatory network is one of the central themes of systems biology. With the growing understanding on the mechanism of microRNA biogenesis and its biological function, establishing a microRNA-mediated gene regulatory network is not only desirable, but also achievable. Methodology: in this study, we propose a bioinformatics strategy to construct microRNA-mediated regulatory network using genome-wide binding patterns of transcription factor(s) and RNA polymerase II (Pol II), derived using chromatin immunoprecipitation following next generation sequencing (ChIP-seq) technology. Our strategy includes three key steps, identification of transcription start site and promoter regions of primary microRNA transcripts using RNA Pol II binding patterns, selecting cooperating transcription factors that collaboratively function with the transcription factors targeted by ChIP-seq assay, and construct network that contains regulatory cascades of both transcription factors and microRNAs. Principal Findings: Using CAMDA (Critical Assessment of Massive Data Analysis) 2009 data set that includes ChIP-seq data on RNA Pol II and STAT1 in HeLa S3 cells without and with interferon γ stimulation, we first identified promoter regions of 83 micorRNAs in HeLa cells. We then identified two potential STAT1 collaborating factors, AP-1 and C/EBP, and further established five feedback network elements that may regulate cellular response during interferon γ stimulation. Conclusions: This study offers a bioinformatics strategy to provide testable hypotheses on the mechanisms of microRNA-mediated transcriptional regulation, based upon genome-wide protein-DNA interaction data derived from ChIP-seq experiments.

Session: M35
Regency B
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Li, Mingyu* **Affiliation:** *Celgene Corp.*

Abstract Title: ***Comparisons of Procedures for Two-Stage Adaptive Designs in Clinical Trials***

Author(s): *Mingyu Li*, Hui Quan, Weichung Joe Shih, Kaihong Jiang and Peter S. Ouyang*

Abstract: There is an upward trend of applying adaptive designs in more and more clinical trials. The p-value combination procedure, weighted combination procedure, conditional error function method and likelihood ratio test are the four commonly used procedures in adaptive designs. They incorporate the features of potential early stopping, sample size re-estimation and overall Type I error rate control. In this research, we consider the generalize p-value combination procedure and the generalized circular conditional error function method and then compare the four procedures on the same ground of sample size, the significance and futility levels. Our computational results demonstrate that the four procedures have very similar power. However, the p-value combination and weighted combination procedures are more flexible and easy for use particularly compared to the likelihood ratio test procedure.

Session: M33
Network
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Li, Ni* **Affiliation:** *University of Missouri*

Abstract Title: ***Semiparametric Transformation Models for Panel Count Data with Dependent Observation Process***

Author(s): *Ni Li*, Liuquan Sun and Jianguo Sun*

Abstract: Panel count data usually occur in longitudinal follow-up studies that concern occurrence rates of certain recurrent events and in which study subjects can be observed only at discrete time points rather than continuously. In these situations, only the numbers of the events that occur between the observation times, not their occurrence times, are observed. Furthermore, the observation times or process may differ from subject to subject and more importantly, it may contain relevant information about the underlying recurrent event. This paper discusses regression analysis of such data and for the problem, a class of semiparametric transformation models is presented. The models are much more flexible than the existing ones and include many commonly used models as special cases. For estimation of regression parameters, some estimating equations are developed and the resulting estimators are shown to be consistent and asymptotically normal. An extensive simulation study was conducted and indicates that the proposed approach works well for practical situations. An illustrative example is provided.

Session: M23
Concept AB
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Li, Bing* **Affiliation:** *Pennsylvania State University*

Abstract Title: ***Groupwise Dimension Reduction***

Author(s): *Lexin Li, Bing Li*, and Lixing Zhu*

Abstract: In many regression applications, the predictors fall naturally into a number of groups or domains, and it is often desirable to establish a domain-specific relation between the predictors and the response. In this article, we consider dimension reduction that incorporates such domain knowledge. The proposed method is based on the derivative of the conditional mean, where the differential operator is constrained to the form of a direct sum. This formulation also accommodates the situations where dimension reduction is focused only on part of the predictors; as such it extends Partial Dimension Reduction to cases where the blocked predictors are continuous. Through simulation and real data analyses, we show that the proposed method achieves greater accuracy and interpretability than the dimension reduction methods that ignore group information. Furthermore, the new method does not require the stringent conditions on the predictor distribution that are required by existing methods.

Session: W28
Cosmopolitan D
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Li, Lingling* **Affiliation:** *Harvard Pilgrim Health Care Institute*

Abstract Title: ***Propensity-score based sensitivity analysis method for uncontrolled confounding***

Author(s): *Lingling Li*, Changyu Shen, Ann C. Wu, Xiaochun Li*

Abstract: In this talk, we introduce an alternative sensitivity analysis method for uncontrolled confounding in observational studies. The method was developed from a different perspective than the one discussed in the previous talk. This method uses a one-dimensional sensitivity function (SF) to quantify the residual confounding due to unmeasured confounders after adjusting for measured baseline characteristics. The SF only depends on a scalar random variable, the propensity score (PS), which is defined as the conditional probability of being treated given measured confounders. The strategy of using SF to remove hidden bias applies to multiple PS-based methods including inverse probability weighting and PS stratification. Having a one-dimensional SF facilitates the conduct of a comprehensive sensitivity analysis with varying assumptions. Moreover, it offers better chances for robust inference. A one-dimensional continuous function can be well approximated by low order polynomial terms. Therefore, even if the imposed SFs are practically certain to be incorrect, we can still hope to obtain robust inference by using polynomial SFs with varying orders and coefficients. We will illustrate the implementation

Session: T12
Concept CD
Tuesday, June 22
10:20 a.m.–12:10 p.m.

of this method with an asthma medication use example.

Presenter: *Li, Yi* **Affiliation:** *Harvard School of Public Health*

Abstract Title: ***Analysis of Survival Data With High Dimensional Predictors***

Author(s): *Yi Li, Lee Dicker and Dave Zhao*

Abstract: This talk introduces a new class of Dantzig variable selectors for linear regression models for right-censored outcomes. We first establish the finite sample error bound for the estimator when p>n and show the proposed selector is nearly optimal in the L2 sense and effectively reduces the dimension of predictors. To improve model selection performance, we further propose an adaptive Dantzig variable selector when p is less than n and discuss its large sample properties, namely, consistency in model selection and asymptotic normality of the estimator. The practical utility of the proposed adaptive Dantzig selectors is verified via extensive simulations. We apply the proposed methods to a myeloma clinical trial and identify important predictive genes for patients' survival.

Session: W17
Cosmopolitan D
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Li, Shaoyu* **Affiliation:** *Michigan State University*

Abstract Title: ***Indentifying novel pathway regulation in eQTL mapping***

Author(s): *Yuehua Cui and Shaoyu Li**

Abstract: The genetic bases of complex traits often involve multiple inherited genetic factors that function in a network basis. By changing the expression of functional genes related to a trait, gene regulations have been thought to be a major player in determining the trait variations. The combined analysis of genetic and gene expression, termed eQTL mapping, holds great promise in this regard. Known that genes function in a network basis, the detection of overall signal of the system could shed new light on the role of genetic regulation. We propose to identify novel regulators that mediate the expression changes by combining evidences to study gene regulations in an eQTL mapping framework. We hypothesize that gene expression changes are due to the regulation of a set of variants that belongs to a common system (e.g., network/pathway), and combine individual p-values in the system to form an overall signal while considering correlations between variants. Both simulation and real data analysis show the relative merits of the combined method. The proposed method provides an alternative strategy in addressing questions related to gene regulations from a systems biology perspective.

Session: T15
Regency A
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Li, Huilin* **Affiliation:** *NCI*

Abstract Title: ***Secondary Analysis of Case-control Data and Application in Genetic Association Studies***

Author(s): *Huilin Li* and Mitchell Gail*

Abstract: Case-control genome-wide association studies provide a vast amount of genetic information that may be used to investigate secondary phenotypes. We study the situation in which the secondary phenotype and genetic markers are dichotomous. We first prove that with disease rate is known, the inverse-probability-of-sampling-weighted (IPW) regression method is exactly the maximum likelihood estimation method using the full disease model. Those two methods are the most robust methods in term of guarding the possibility of interaction effect of genetic variants and secondary phenotype on the disease. When there is no interaction effect, the maximum likelihood estimation method with the no interaction assumption is the most efficient method. To strike a balance of the above methods, we proposed an adaptively weighted method that combines the IPW and MLE with reduced disease model. Our adaptively weighted method is always unbiased and has reduced mean square error for estimation with a pre-specified gene and increase the power to discover a new association in a genome-wide study when non-zero interaction is possible. Case-control with known population totals study design is also investigated in this paper.

Session: W16
Network
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Li, Lang* **Affiliation:** *Indiana University*
Abstract Title: ***A Personalized Drug Exposure Predictive Model Framework***
Author(s): *Lang Li*
Abstract: A personalized drug exposure model framework will be introduced. This personalized drug exposure model framework will utilize informatics approach such text mining; Bayes models and global optimization approaches; and experimental validations.

Session: T23
Regency B
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Li, Bo* **Affiliation:** *Purdue University*
Abstract Title: ***An approach to modeling asymmetric multivariate spatial covariance structure***
Author(s): *Bo Li* and Hao Zhang*
Abstract: We propose a framework to model the asymmetric multivariate covariance function that is often exhibited in real data. This general approach can endow any valid symmetric covariance function the ability of modeling asymmetric covariance structure and is very easy to implement. Our simulations and real data examples show that the asymmetric covariance function based on our approach can achieve remarkable improvement in prediction over the symmetric model.

Session: T36
Cosmopolitan C
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Liang, Faming* **Affiliation:** *Texas A&M University*
Abstract Title: ***A Monte Carlo Metropolis-Hastings Algorithm for Sampling from Distributions with Intractable Normalizing Constants***
Author(s): *Faming Liang*, Ick Hoon Jin*
Abstract: The problem of simulating from distributions with intractable normalizing constants has received much attention in recent literature. In this talk, we present a new algorithm, the so-called Monte Carlo Metropolis-Hastings (MCMH) algorithm, for tackling this problem. The MCMH algorithm is a Monte Carlo version of the Metropolis-Hastings algorithm. It replaces the unknown normalizing constant ratio by a Monte Carlo estimate in simulations, while still converges to the desired target distribution. The MCMH algorithm is illustrated with the spatial autologistic models and exponential random graph models. Unlike other auxiliary variable MCMC algorithms, such as the Mo{}ller and exchange algorithms, the MCMH algorithm avoids the requirement for perfect sampling, and thus can be applied to many statistical models for which perfect sampling is not available or very expensive. As discussed in the paper, the MCMH algorithm can also be applied to Bayesian inference for the random effect models and the missing data problems which involve simulations from a distribution with intractable integrals.

Session: M36
Cosmopolitan C
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Liang, Liming* **Affiliation:** *Harvard University*
Abstract Title: ***The power of imputation using variants identified by the 1000 Genomes Project***
Author(s): *Liming Liang*, Nilesh Morar, Anna L Dixon, G Mark Lathrop, Goncalo R Abecasis, Miriam F Moffatt, William O C Cookson*
Abstract: Genotype imputation has been shown to be a useful tool to improve power of genome-wide association studies. To date, most studies using genotype imputation have used one of the commercial genotyping arrays (from Illumina, Affymetrix, Perlegen Biosciences, among others) to genotype study samples and then used the HapMap samples as template reference panel. The 1000 Genomes Project aims at creating a most detailed map to date of human genetic variation. The research community is keen to know how this much denser category of SNPs identified from this project is going to help genome-wide association studies. Using global gene expression data from two independent datasets, which provides a broad spectrum of heritability and genetic architecture, we evaluate how imputation using the SNPs identified from the 1000 Genomes Project can improve power on genome-wide association

Session: T25
Network
Tuesday, June 22
1:30 p.m.–3:20 p.m.

studies. We show that with a similar sample size of the reference panel, the 1000 Genomes imputation gives ^5-8% more power on top of the HapMap imputation and ^12-17% increase of power compared to genotype platforms. We anticipate that the full scale of the 1000 Genomes Project which eventually sequence 1200 individuals will further increase the gain of power. At the same tine, results from this study provide a largely expanded eQTL database based on newly available samples, expression profiling techniques (Affymetrix U133Plus2 and Illumina Human-6) and statistical methods to measure gene expression and the densest to date SNP panels. This database will be available to the public through a convenient way that has helped researchers to study functions of many disease associated loci.

Presenter: *Lin, Xiwu* **Affiliation:** *GlaxoSmithKline*

Abstract Title: ***Truncated Robust Distance for Clinical Laboratory Safety Data Monitoring and Assessment***

Author(s): *Xiwu Lin*, Daniel Parks, Lei Zhu, Kwan Lee*

Abstract: Safety laboratory data are routinely collected in clinical studies for safety monitoring and assessment. We develop a truncated robust multivariate outlier detection method for identifying subjects with clinically relevant abnormal lab measurements. The proposed method can be applied to historical clinical data to establish a multivariate decision boundary for future clinical trial safety laboratory data monitoring and assessment. We perform numerical simulations to evaluate the performance of the proposed method in terms of detecting clinically relevant outliers. Simulations suggest that the proposed method performs as expected and has the ability to block the irrelevant outliers. Examples based on safety laboratory data from real clinical studies are shown to illustrate the use of the proposed method for identifying clinically relevant outliers.

Session: T32
Regency A
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Lin, Xiaodong* **Affiliation:** *Rutgers University*

Abstract Title: ***Penalized Maximum Likelihood Estimation for Stationary Time Series***

Author(s): *Yan Sun, Xiaodong Lin**

Abstract: The past decade has seen a rapid development of regularization techniques such as ridge regression, LASSO, SCAD, LARS and their extensions. However, these techniques have been developed mainly for circumstances where the observations are independent. In this talk, we will first describe extensions of the results of penalized methods for independent data to stationary multivariate time series. Under mild regularity conditions, our penalized estimators are sparse-consistent and possess well-known oracle properties. We demonstrate the utility of our results by developing a sparse version of the full factor GARCH model. Furthermore, we study the problem of regularization for AR(p) models with varying lags and demonstrate the connections between the regularized AR process with ARFIMA models. Finally, we show the applicability of our theory and methods via real and simulated data.

Session: W27
Cosmopolitan C
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Lin, Nan* **Affiliation:** *Washington University in St. Louis*

Abstract Title: ***Bayesian Regularization in Quantile Regression***

Author(s): *Qing Li, Ruibin Xi, Nan Lin**

Abstract: We study regularization in quantile regressions from a Bayesian perspective. By proposing a hierarchical model framework, we give a generic treatment to a set of regularization approaches, including lasso, group lasso and elastic net penalties. Gibbs samplers are derived for all cases. Both simulated and real data examples show that Bayesian regularized quantile regression methods often outperform quantile regression without regularization and their non-Bayesian counterparts with regularization.

Session: T28
Studio One
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Liu, Yufeng* **Affiliation:** *University of North Carolina*

Abstract Title: ***Multicategory Composite Least Squares Classifiers***

Author(s): *Yufeng Liu**

Abstract: Classification is a very useful statistical tool for information extraction. Although binary

Session: T18
Cosmopolitan D
Tuesday, June 22
10:20 a.m.–12:10 p.m.

classification problems are heavily studied, extensions to the multicategory case are much less so. In view of the increased complexity and volume of modern statistical problems, it is desirable to have multicategory classifiers that are able to handle problems with high dimensions and with a large number of classes. In this talk, I will present a new efficient multicategory composite least squares classifier (CLS classifier). Properties and performance of the proposed CLS classifier will be discussed. This talk is based on joint work with Seo Young Park.

Presenter: *Liu, Rong* **Affiliation:** *University of Toledo*

Abstract Title: ***SPLINE ESTIMATION OF A SEMIPARAMETRIC GARCH MODEL***

Author(s): *Rong Liu* Lijian Yang*

Abstract:

Session: T38
Studio One
Tuesday, June 22
3:40 p.m.–5:30 p.m.

The semiparametric GARCH model of Yang (2006) has combined the flexibility of a nonparametric link function with the dependence on infinitely many past observations of the classic GARCH model. We propose a cubic spline procedure to estimate the unknowns in the semiparametric GARCH model that is intuitively appealing due to its simplicity. The theoretical properties of the procedure is the same as the kernel procedure, while simulated and real data examples show that the numerical performance is either better than or comparable to the kernel method. The new method is computationally much more efficient than the kernel method and very useful for analyzing financial time series data.

Presenter: *Liu, Dungang* **Affiliation:** *Rutgers University*

Abstract Title: ***Combination of Confidence Distributions and an Efficient Approach for Meta-Analysis of Heterogeneous Studies***

Author(s): *Dungang Liu*, Regina Liu and Minge Xie*

Abstract:

Session: W13
Concept AB
Wednesday, June 23
8:00 a.m.–9:50 a.m.

We propose a new methodology for combining summary information from independent studies based on generalized linear models. This work is motivated by the need in meta-analysis to investigate the effect of covariates in clinical trials, especially when covariate designs are heterogeneous among the studies. To this end, we develop a combination approach for meta-analysis utilizing multiparameter normal confidence distributions (CDs). We show that the CD-based meta-analysis estimator asymptotically achieves the Cramer-Rao lower bound, and is as efficient as the Individual Patient Data (IPD) estimator derived from the complete individual data. But unlike the IPD estimator, the proposed approach only uses summary information and does not need any individual data. Additionally, it can adapt to heterogeneous studies, such as the case of missing design variables or non-estimable parameters, for which conventional meta-analysis is inadequate. We illustrate our approach with both simulation and real data analysis.

Presenter: *Liu, Desheng* **Affiliation:** *The Ohio State University*

Abstract Title: ***Reconstructing Land Cover Change Trajectories from Time Series Satellite Observations***

Author(s): *Desheng Liu*

Abstract:

Session: W15
Cosmopolitan C
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Land cover change trajectories over multiple time points are important indicators of complex human activities. Such change trajectories can be reconstructed from the classification of time series satellite observations. Current methods based on independent land cover classification often lead to a low accuracy in the resultant land cover change trajectories due to the propagation of classification errors at each time point. In this talk, we present a spatial-temporal approach to classifying time series satellite data jointly in an attempt to mitigate the impacts of independent classification errors on the accuracy of land cover change trajectories. Specifically, a spatial-temporal Markov Random Fields (MRF) model is developed to integrate spatial-temporal contextual information with spectral information for multi-temporal land cover classification. Promising results are obtained from the estimates of land cover change trajectories based on a time series of Landsat imagery in southeast Ohio. This work demonstrates the importance of spatial-temporal contextual information in

mapping land cover change trajectories.

Presenter: *Liu, Yushi* **Affiliation:** *the Ohio State University*

Abstract Title: ***the temporal profile analysis of gene expresison for female rainbow trout under normal and compressed cycle***

Author(s): *Yushi Liu*, Joe Verducci*

Abstract: New statistical procedures are introduced to investigate gene activity in support of the hypothalamus-pituitary-gonad-liver (HPGL) signaling network that provides the neuroendocrine regulation for reproduction in female, oviparous fishes. The methods include SCOOP (Shrunken Centroid Ordering by Orthogonal Projections) and a robust encoding of B-splines via Friedman's Generalized Elastic Net (GEN). Clustering of genes via GEN-transformed measurements produced much more stable results, and the mean time course pattern of each cluster provided biologists with a reliable summary from which to interpret systematic patterns. Ultimately, the genes selected by SCOOP and clustered though the GEN-transform strongly suggested supportive pathways involving immunology, muscle contraction, reproduction, protein transport, metabolism, and reduction/oxidation. Potential theoretical properties of SCOOP are still under vigorous investigation. Under certain multivariate normal models, SCOOP can be shown to delete the nuisance variables automatically. New optimization criterion has been developed for selecting sets of variables using maximum likelihood methods. Such criterion is based on minimizing certain dispersion functions and the variance of the leading principle component. We will extend this to multiple groups. Each of the group will possess a multivariate normal distribution. This can give insight into how to weigh between and within epoch matrices, therefore, leading the potential improvement of SCOOP.

Session: M18
Studio One
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Liu, Chuanhai* **Affiliation:** *Purdue University*

Abstract Title: ***Statistical Inference: Reconsideration for High-Dimensional Problems***

Author(s): *Chuanhai Liu*

Abstract: After the failure of the efforts in developing fiducial inference, it is perhaps now a most common view that combining Bayesian and frequentist methods is a promising strategy to solve very high dimensional problems. As more and more scientific investigations rely on statistical analysis, one may still be concerned with fundamental problems of statistical inference that have remained unsolved since Fisher (1930). Using the many-normal-means problem as an illustrative example, I present some baby steps we have taken to understand statistical inference in the presence of unknown quantities with no priors. I also discuss a new resolution to Stein's paradox and inferential issues in linear regression.

Session: T26
Cosmopolitan C
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Liu, Fei* **Affiliation:** *IBM*

Abstract Title: ***High-Dimensional Variable Selection in Meta Analysis for Censored Data***

Author(s): *Fei Liu, David Dunson , and Fei Zou*

Abstract: This article considers the problem of selecting predictors of time to an event from a high-dimensional set of candidate predictors using data from multiple studies. As an alternative to the current multi-stage testing approaches, we propose to model the study-to-study heterogeneity explicitly using a hierarchical model to borrow strength. Our method incorporates censored data through an accelerated failure time (AFT) model. Using a carefully-formulated prior specification, we develop a fast approach to predictor selection and shrinkage estimation for high-dimensional predictors. For model fitting, we develop a Monte Carlo Expectation Maximization (MC-EM) algorithm to accommodate censored data. The proposed approach, which is related to the relevance vector machine (RVM), relies on maximum a posteriori (MAP) estimation to rapidly obtain a sparse estimate. As for the typical RVM, there is an intrinsic thresholding property in which unimportant predictors tend to have their coefficients shrunk to zero. We compare our method with some commonly used

Session: T27
Cosmopolitan D
Tuesday, June 22
1:30 p.m.–3:20 p.m.

procedures through simulation studies. We also illustrate the method using the gene expression barcode data from three breast cancer studies. Key words: Accelerated failure time; EM algorithm; Lasso; MAP estimation; Meta analysis; Relevance vector machine; Shrinkage.

Presenter: *Liu, X. Shirley* **Affiliation:** *Dana-Farber Cancer Institute / Harvard*

Abstract Title: ***Discovering cis-regulatory regions from differential nucleosome occupancy patterns***

Author(s): *Clifford A. Meyer, Bo Jiang, Hansen H. He, Myles Brown, Jun S. Liu and X. Shirley Liu**

Abstract: We describe a method to detect in vivo cell type and condition specific transcription factor binding locations through an integrative analysis of differentiation- or stimulus-induced shifts in nucleosome occupancy and DNA sequence data. We use a Bayesian statistical framework to integrate ChIP-seq data on nucleosome occupancy shifts with DNA sequence and transcription factor binding motifs to estimate probabilities of transcription factor binding at sites flanked by two positioned nucleosomes. Using this method we screen a library of known transcription factor DNA binding motifs to infer transcription factor binding sites and the identity of transcription factors that are the key regulators of cell differentiation or a response to a stimulus. We demonstrate the effectiveness of this method through the determination of key transcription factors in the yeast heat shock response, the response of human CD4+ T-cells to infection and in the differentiation of human CD133+ to CD36+ cells. Our method provides an effective way of determining key transcription factors driving cell differentiation or stimulus response.

Session: M35
Regency B
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Lu, Yuefeng* **Affiliation:** *Eli Lilly and Co.*

Abstract Title: ***DCE-MRI Quantification and Modeling for Angiogenesis Tumor Animal Model Development***

Author(s): *Yuefeng Lu**

Abstract: Dynamic Contrast Enhanced MRI (DCE-MRI) imaging provides a non-invasive method to assess the tumor permeability and microvascular function in in-vivo animal models. Summarization, modeling and interpretation of DCE-MRI images need to be based on synthetic knowledge of MRI physics, physiological effect of the interaction between the dynamic contrast and the tissue, tumor biology, pharmacokinetics and statistical models. When design a DEC-MRI study, practitioners will need to decide what MRI parameters to measure and what statistics and models to use to have more reproducible results and most relevant biological interpretations. In this talk, I will review basic DCE-MRI knowledge and discuss design issues and modeling approaches for angiogenesis tumor animal model development.

Session: M25
Studio One
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Luo, Xiaolong* **Affiliation:** *Celgene Corporation*

Abstract Title: ***An Optimal Adaptive Design to Address Local Regulations in Global Clinical Trials***

Author(s): *Xiaolong Luo*, Weichung Joe Shih,S. Peter Ouyang,Bob DeLap*

Abstract: After Multi-Regional Clinical Trial (MRCTs) have demonstrated overall significant effects, evaluation for a region specific effect is often important. Recent guidance from regulatory authorities regarding evaluation for possible country specific effects has led to research on statistical designs that incorporate such evaluations in MRCTs. These statistical designs are intended to use MRCTs to address requirements for global registration of a medicinal product. Adding a regional requirement could change the probability for declaring positive effect for the region when there is indeed no treatment difference as well as when there is in fact a true difference within the region. In this paper, we first quantify those probability structures based on the guidance issued by the Ministry of Health, Labour and Welfare (MHLW) of Japan. An adaptive design is proposed to consider those probabilities and to optimize the efficiency for regional objectives. This two stage approach incorporates comprehensive global objectives into an integrated study design and may mitigate the need

Session: T11
Concept AB
Tuesday, June 22
10:20 a.m.–12:10 p.m.

for a separate local bridging study. A procedure is used to optimize region specific enrollment based on an objective function. The overall sample size requirement is assessed. We will use simulation analyses to illustrate the performance of the proposed study design.

Presenter: *Lv, Jinchi* **Affiliation:** *University of Southern California*

Abstract Title: ***Non-Concave Penalized Likelihood with NP-Dimensionality***

Author(s): *Jianqing Fan and Jinchi Lv**

Abstract: Penalized likelihood methods are fundamental to ultra-high dimensional variable selection. How high dimensionality such methods can handle remains largely unknown. In this paper, we show that in the context of generalized linear models, such methods possess model selection consistency with oracle properties even for dimensionality of Non-Polynomial (NP) order of sample size, for a class of penalized likelihood approaches using folded-concave penalty functions, which were introduced to ameliorate the bias problems of convex penalty functions. This fills a long-standing gap in the literature where the dimensionality is allowed to grow slowly with the sample size. Our results are also applicable to penalized likelihood with the L_1-penalty, which is a convex function at the boundary of the class of folded-concave penalty functions under consideration. The coordinate optimization is implemented for finding the solution paths, whose performance is evaluated by a few simulation examples and the real data analysis.

Session: M15
Cosmopolitan D
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Ma, Xiwen* **Affiliation:** *University of Wisconsin-Madison*

Abstract Title: ***Defining Responder in Alcohol Dependence Patients***

Author(s): *Xiwen Ma* ; Yun-Fei Chen; Haoda Fu; Conrad Wong*

Abstract: Alcohol Dependence is one of the leading causes of disability and mortality worldwide. Treatment for alcohol dependence varies for each individual and the investigation of the efficacy of pharmacotherapy is not always straightforward. While trial is conducted, the definition of a responder is not under consensus in the field. Current recommendation in allowing grace period to calculate the percentage of patients without heavy drinking for the rest of the trial does not take into account of baseline difference. We propose an innovative framework of responder analysis using the change from baseline relationship of the primary end point and one major health outcome measure with quantile regression from COMBINE data. We provide not only a meaningful definition to a responder, but also the quantile regression approach could be generalized for the statistical analysis in the field.

Session: M13
Network
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Ma, Shujie* **Affiliation:** *Michigan State University*

Abstract Title: ***A simultaneous confidence band for sparse longitudinal regression***

Author(s): *Shujie Ma*, Lijian Yang and Raymond J. Carroll*

Abstract: Functional data analysis has received considerable recent attention and a number of successful applications have been reported. In this paper, asymptotically simultaneous confidence bands are obtained for the mean function of the functional regression model, using piecewise constant spline estimation. Simulation experiments corroborate the asymptotic theory. The confidence band procedure is illustrated by analyzing the CD4 cell counts of HIV infected patients.

Session: T38
Studio One
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Macias, William* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Personlized Medicine and Tailored Therapeutics***

Author(s): *William L. Macias, MD, PhD*

Abstract: The use of tailored therapeutics, medications for which treatment decisions are based on the molecular profile of the patient, the disease, and/or the patient's response to treatment, should improve the precision of medical care. On average, only 50% of patients respond to prescribed medications meaning that the complement of patients is treated with an ineffective medication, sometimes in lieu of an effective medication. The beneficial impact of tailored therapeutics may be largest in the treatment of patients with severe illnesses or complex

Session: M32
Regency A
Monday, June 21
3:40 p.m.–5:30 p.m.

diseases where initial use of an ineffective medicine might be associated with mortality or major morbidity. The molecular marker used to identify appropriate patients will need to be very sensitive so as to not deny a potentially effective therapy to a marker negative patient. For therapeutics with potentially serious side effects or where other effective therapies are available, the marker may need to be both sensitive and specific. If a diagnostic test is necessary to measure the marker, the test itself will need to have both analytical and clinical validity. Substantial evidence will need to be obtained from adequate and well controlled clinical studies to allow inclusion of this information in product labeling. All of these requirements have significant implications as regards the clinical development of new chemical entities including the need to very early in development define the "tailoring hypothesis" so that prospective testing of the hypothesis can be conducted in pivotal trials.

Presenter: *Maiti, Taps* **Affiliation:** *Michigan State University*

Abstract Title: ***Small Area Estimation by Shrinking Means and Variances***

Author(s): *Taps Maiti*; Hao Ren*

Abstract: Shrinkage estimation plays a major role in small area estimation to overcome small sample issues. While the research is predominated by shrinking the area level means, or variances separately, there is hardly any research done involving simultaneous shrinkage of both means and variances. This talk will introduce a technique to address the above issue.

Session: T31
Concept AB
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Mallick, Bani* **Affiliation:** *Texas A&M*

Abstract Title: ***Bayesian Modeling of MPSS Data: Gene Expression Analysis of Bovine Salmonella Infection***

Author(s): *Soma S. Dhavala, Sujay Datta, Bani K. Mallick*, Raymond J. Carroll,Sangeeta Khare, Sara D. Lawhon and L. Garry Adams*

Abstract: Among the counting-based technologies available for gene expression profiling, massively parallel signature sequencing (MPSS) has some advantages over competitors such as serial analysis of gene expression (SAGE) or direct sequencing of cDNA and is ideal for building complex relational databases for gene expression. The goal of our present study was a comparison between the {it in vivo} global gene expression profiles of %sout{bovine ligated ileal loops} {tissues} infected with different strains of {it Salmonella} obtained using the MPSS technology. In this article, we develop an exact ANOVA type model for this count data using a zero inflated Poisson (ZIP) distribution, different from the existing methods that assume continuous densities. We adopt two Bayesian hierarchical models---one parametric and the other semiparametric with a Dirichlet process prior that has the ability to ``borrow strength'' across related signatures, where signature is a specific arrangement of the nucleotides, usually 16-21 base-pairs long. Modeling each signature-count by a ZIP, we assume a normal density for the log-transformed mean parameter of the Poisson part. The mean of this normal density is assumed to have a linear model structure with parameters capturing the signature effect and the treatment effect. In the parametric model these parameters are given the usual conjugate prior distributions, whereas in the semiparametric case, the `treatment effect' parameter is given a Dirichlet process prior with a normal baseline distribution, resulting in automatic clustering of the signatures. The deviance information criterion (DIC) is used for model choice and inference on differential expression is based on the posteriors of the `treatment effect' parameters. To this end, symmetrized Kullback-Leibler (KL) divergences with bootstrapped cut-off values are used, as well as the Kruskall-Wallis test for the equality of medians. Among the genes associated with the differentially expressed signatures identified by our semiparametric model, there are several important Gene Ontology categories that are consistent with the existing biological knowledge about the host response to Salmonella infection. We conclude with a summary

Session: T37
Cosmopolitan D
Tuesday, June 22
3:40 p.m.–5:30 p.m.

of the biological significance of our discoveries.

Presenter: *Mallinckrodt, Craig* **Affiliation:** *Eli Lilly*

Abstract Title: ***Observational and Causal Components of Placebo Response***

Author(s): *Craig Mallinckrodt*

Abstract:

Signal detection is the term often used to describe the ability to differentiate between an effective drug and placebo; that is, to find a treatment effect when one exists. In many disease states increasing placebo response has been implicated as the primary cause of decreasing signal detection. The observational and causal components of placebo response can be used to better understand and control placebo response. Examples from the development of antidepressants and antipsychotics are used to illustrate how design features can be manipulated to reduce placebo response along with other means of improving signal detection.

Session: W23
Regency B
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Mandrekar, Sumithra* **Affiliation:** *Mayo Clinic*

Abstract Title: ***Clinical Trial Designs for Predictive Biomarker Validation***

Author(s): *Sumithra J. Mandrekar*, Daniel J. Sargent*

Abstract:

Biomarkers can guide patient-specific treatment selection by providing an integrated approach to prediction using the genetic makeup of the tumor and the genotype of the patient. Designs for predictive marker validation are broadly classified as retrospective (i.e., using data from previously well-conducted randomized controlled trials (RCT)) versus prospective (enrichment, all-comers or unselected, hybrid, or adaptive analysis). Well designed retrospective analysis can bring forward effective treatments to marker defined subgroup of patients in a timely manner (e.g. K-RAS and colorectal cancer). Prospective enrichment designs are appropriate when compelling preliminary evidence suggests that not all patients will benefit from the study treatment, however this may sometimes leave questions unanswered (e.g. Trastuzumab and breast cancer). An unselected design is optimal where preliminary evidence regarding treatment benefit and assay reproducibility is uncertain (e.g. EGFR and lung cancer). Hybrid designs are appropriate when preliminary evidence demonstrate the efficacy of certain treatments for a marker defined subgroup, making it unethical to randomize patients with that marker status to other treatments (e.g. multigene assay and breast cancer). Adaptive analysis designs allow for pre-specified marker defined subgroup analyses. The implementation of these design strategies will lead to a more rapid clinical validation of biomarker guided therapy.

Session: M21
Network
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Marino, Miguel* **Affiliation:** *Harvard Univ. Dept. of Biostatistics*

Abstract Title: ***Statistical Inference in Factor Analysis for High-Dimensional, Low-Sample Size Data***

Author(s): *Miguel Marino and Yi Li*

Abstract:

Cancer researchers are keen on tracking trends in cancer mortality rates and studying the cross relationship of these trends not only for scientific reasons of understanding the cancers as a complex dynamical system, but also for practical reasons such as prevention, planning and resource allocation. Factor analysis which studies such cross-correlation matrices is an effective means of data reduction, whose inference typically requires the number of random variables, p, to be relatively small and fixed, and the sample size, n, to be approaching infinity. However, contemporary surveillance techniques have yielded large matrices in both dimensions, limiting the usage of existing factor analysis techniques due to the poor estimate of the covariance/correlation matrix. We develop methods, in the framework of random matrix theory, to study the cross-correlation of cancer mortality annual rate changes in the setting where $p > n$. We propose methodology to test complete independence across cancer sites. We develop an approach based on group sequential theory to determine the number of significant factors in a factor model. Sparse principal components analysis is studied on the principal components deemed to be significantly different than random matrix theory

Session: W21
Regenecy A
Wednesday, June 23
10:10 a.m.–12:00 p.m.

prediction to aid in the interpretation of the underlying factors. Methods are implemented on SEER cancer mortality rates from 1969 through 2005.

Presenter: *Martin, Ryan* **Affiliation:** *IUPUI*

Abstract Title: ***On likelihood and probabilistic inference without priors***

Author(s): *Jing-Shiang Hwang, Chuanhai Liu, and Ryan Martin**

Abstract: A common theme in contemporary research on foundations of statistical inference is some sort of synergy between Bayesian and frequentist ideas. Recent progress has been made along these lines based on so-called inferential models (IMs), which can achieve frequency-calibrated probabilistic inference without priors. I will review this IM framework, and present some recent developments based on a re-interpretation of the likelihood as a predictive probability measure on an auxiliary space. An interesting by-product of this re-interpretation is a new form of marginal likelihood inference, where integration over the parameter space is replaced by a union over subsets of the auxiliary space. Simple normal examples will illustrate the new approach, with particular attention paid to the famous Behrens-Fisher problem, which has been one of the focal points of controversy between the Neyman-Pearson and Fisherian approaches to statistical inference.

Session: T26
Cosmopolitan C
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Mehta, Cyrus* **Affiliation:** *Cytel Inc*

Abstract Title: ***Adaptive Sample Size Re-estimation in Randomized Clinical Trials***

Author(s): *Cyrus Mehta*

Abstract: This talk will discuss adaptive sample size re-estimation in major Phase III trials based on unblinded interim estimates of the primary effect size. This controversial topic generates two areas of concern: 1) the need for a robust statistical methodology for sample size re-estimation and its consequences for making inferences from the final trial data, and 2) the practical organization of such an adaptive approach paying due regard to the confidentiality of interim data and the need to preserve the integrity of the trial's conduct throughout. We will discuss these issues, and illustrate their importance through examples of actual clinical trials.

Session: T21
Regency A
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Michailidis, George* **Affiliation:** *The University of Michigan*

Abstract Title: ***Adaptive sampling procedures for estimating function thresholds***

Author(s): *George Michailidis*

Abstract: In this talk, we discuss multi (two) - stage sampling procedures for estimating a threshold for a regression function, such as a point where the function crosses some critical value. It will be shown how the proposed procedures, which involve sampling a pre-fixed budget (number) of points (covariate-response pairs) at two stages -the first, a learning stage using an agnostic sampling design, and the second a "zoom-in" stage with sampling in the vicinity of the initial estimate obtained from the learning stage - lead to accelerated convergence rates over one-stage procedures, in certain cases even allowing the parametric square-root-n rate to be achieved or exceeded for a nonparametric problem. The proposed procedure is illustrated on synthetic and real data.

Session: M15
Cosmopolitan D
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Millen, Brian* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Statistical Considerations for Trials Incorporating Tailored Subgroups***

Author(s): *Brian A. Millen, Ph.D.**

Abstract: With emphasis on tailored therapies, clinical trial sponsors are increasingly interested in proving a treatment's efficacy within a pre-specified subpopulation while evaluating its effects in the broader patient population. In such trials, the sponsor is often interested in obtaining an indication for a broad population with additional information in product labeling reflecting the increased effect within the subpopulation. Several statistical issues associated with such trials must be considered. These include issues of design (e.g., enrichment strategies) and analysis (e.g., multiple testing methods and methods for assessing treatment

Session: M21
Network
Monday, June 21
1:30 p.m.–3:20 p.m.

---

and interaction effects). In this talk, we explore these issues and various approaches to addressing the concerns. A specific clinical trial example is used throughout the discussion.

Presenter: *Mitra, Ritendranath* **Affiliation:** *UT MD Anderson Cancer Centre*

Abstract Title: ***Bayesian graphical model for histone modifications***

Author(s): *Riten Mitra* Yuan Ji* Peter Muller*

Abstract: Understanding histone modifications is a challenging and extremely important question in computational biology. These modifications play a role in altering chromatin structure and thus influence various chromatin-dependent processes including replication, DNA-repair, and transcription. There is a growing need to deduce a 'histone code' that unravels the interrelationship between these modifications. Earlier attempts to infer this complex dependence structure had relied mostly on Bayesian networks. One fundamental problem to this approach is that simple Bayesian networks allow only directed acyclic graphs. The biological reality is much more complex , and as in a typical pathway, the space of potential structures contain cycles, loops, directed and undirected edges. The Bayesian graphical model introduced by us takes these into account and models a more generalized form of histone dependence , while balancing the computational cost. This is achieved through the method of reciprocal graphs and imposing a prior on the most likely pathways. We employ Besag's formulation of a Markov random field to impose a very general dependence structure. However it is not feasible to compute certain normalization constants . We get around this problem by implementing a novel augmented MCMC scheme with a suitable proposal.

Session: M18
Studio One
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Mueller, Peter* **Affiliation:** *M.D. Anderson Cancer Center*

Abstract Title: ***Modeling Dependent Gene Expression***

Author(s): *Peter Mueller*, Donatello Telesca, Giovanni Parmigiani, Ralph Freedman*

Abstract: We propose a Bayesian modeling approach for inference about dependence of high throughput gene expression. Our goals are to use prior knowledge about pathways to anchor inference about dependence among genes; to account for this dependence while making inferences about differences in gene's behavior across phenotypes; and to explore differences in the dependence itself across phenotypes. Useful features of the proposed approach are a model-based parsimonious representation of expression as an ordinal outcome, a novel and flexible representation of prior information on the nature of dependencies, and the use of a coherent probability model over both the structure and strength of the dependencies of interest. We evaluate our approach through simulations and in the analysis of data on expression of genes in the Complement and Coagulation Cascade pathway in ovarian cancer.

Session: T37
Cosmopolitan D
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Müller, Hans-Georg* **Affiliation:** *UC Davis*

Abstract Title: ***Beyond Functional Linear Regression***

Author(s): *Hans-Georg Müller**

Abstract: Functional regression has emerged as a useful approach for the analysis of complex data with functional or longitudinal predictors and scalar or functional responses. A major emphasis has been the functional linear regression model, which allows to implement dimension reduction in a simple and straightforward way but may be too restrictive. We will discuss flexible extensions of this model. These extensions include global models, such as functional polynomial and functional additive models; and local models, where the focus is on the dependency of a Gaussian process or its derivatives at a given time on the value of a predictor process at the same or a different time. The methods will be illustrated with densely as well as sparsely sampled functional data. This talk is based on joint work with Wenjing Yang and Fang Yao.

Session: T14
Cosmopolitan C
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Naranjo, Joshua* **Affiliation:** *Western Michigan University*
Abstract Title: ***A simple diagnostic for proportional hazards and the logrank test***
Author(s): *Joshua Naranjo*, Ruvie Lou Martinez*
Abstract: The logrank test is commonly used to compare two survival curves. Ideally, the logrank test is used after checking for validity of the proportional hazards assumption. When PH assumption is violated, the generalized Wilcoxon test is a frequent alternative. We show that proportional hazards implies relatively late maximum separation between the two curves (within the range of comparison). We investigate a statistic that quantifies lateness of separation. The statistic may be used as a pretest for using either the logrank or Wilcoxon.

Session: W18
Studio One
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Nettleton, Dan* **Affiliation:** *Iowa State University*
Abstract Title: ***Testing for Heterosis in Gene Expression***
Author(s): *Dan Nettleton**
Abstract: Heterosis, also known as hybrid vigor, occurs when the mean trait value of offspring is more extreme than that of either parent. For maize, this phenomenon was first documented in the late 1800s and is the basis of the seed corn industry today. In an effort to understand the molecular genetic mechanisms responsible for heterosis, researchers have begun to measure the expression levels of thousands of genes in parental maize lines and their offspring. We will discuss statistical tests that can be used to identify genes that exhibit heterosis in their expression. The testing problem is nonstandard because the null hypothesis of no heterosis is a union of two essentially disjoint closed convex cones that is neither a cone nor convex. The roles of likelihood ratio, intersection union, and union intersection tests will be presented. The challenge of simultaneously testing thousands of nonstandard null hypotheses will be discussed.

Session: T17
Regency B
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Nguyen, Thuan* **Affiliation:** *Oregon Health & Science University*
Abstract Title: ***Fence Methods for gene mapping under the Backcross Experiments***
Author(s): *Thuan Nguyen*, Jie Peng, and Jiming Jiang*
Abstract: Gene mapping techniques aim to identify genes. Furthermore, gene mapping elucidates how these genes function interactively with each other and with the environment, in the process of contributing to phenotypic variations. Gene mapping is a long and complicated process, and as an important first step, the goal is to identify the genomic regions that harbor trait susceptible genes. For many simple traits, strong correlation between the phenotype and marker phynotype have been established. However, for complex traits, the contribution of any particular gene is usually quite small. This results in difficulties in mapping such genes due to weak correlation(s) between their genotypes and the phenotype under the study. Several statistical methods have arisen to take advantage of the availability of numerous informative markers, in order to map genes more successfully in complex traits. Model selection techniques play an important role simultaneously mapping trait susceptible genes simultaneously. The fence method (Jiang et al. 2008) was motivated to deal with a number of limitations of the traditional information criteria based model selection approaches, such as BIC and AIC. In this work, the fence method is applied to perform gene mapping under the context of backcross experiments in model organisms. Its performance is evaluated by simulation studies and it is also compared with the BIC_delta method proposed by Broman and Speed (2002). KEYWORDS: Quantitative Trait Loci, Backcross, Restricted fence method.

Session: M18
Studio One
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Nguyen, Thuan* **Affiliation:** *Oregon Health & Science University*
Abstract Title: ***Fence Method for Nonparametric Small Area Estimation***
Author(s): *Jiming Jiang, Thuan Nguyen*, and Sunil J. Rao*
Abstract: We consider the problem of selecting nonparametric models for small area estimation, which recently have received much attention. We develop a procedure based on the idea of fence

Session: T31
Concept AB
Tuesday, June 22
3:40 p.m.–5:30 p.m.

method (Jiang et al. 2008) for selecting the mean function for the small areas from a class of approximating splines. Simulation results show impressive performance of the new procedure even when the number of small areas is fairly small. The method is applied to a hospital graft failure dataset for selecting a nonparametric Fay-Herriot type model.

Presenter: *Norton, Jonathan* **Affiliation:** *FDA/CDER*

Abstract Title: ***A Longitudinal Model for Medical Benefit-Risk Analysis, with Case Study***

Author(s): *Jonathan D Norton**

Abstract:

Session: M38
Studio One
Monday, June 21
3:40 p.m.–5:30 p.m.

The regulatory decision for a medical product should be based on a comparative assessment of the benefits and risks from its use. There is little consensus, however, about how this assessment should be conducted. I present a generalization of a model introduced by Chuang-Stein et al. in 1991. This model assumes that a study subject's benefit-risk profile is one of five discrete clinical states, one being premature withdrawal from the study. The paradigm of treating withdrawal as an outcome in itself is particularly appropriate for drugs that provide only symptomatic relief, because a subject's perception of utility is particularly germane to the benefit-risk assessment. My revision of the Chuang-Stein model allows the state to change during the course of the trial. In principle, this offers a more complete assessment of how the effects of a treatment may change over time, as well as showing any temporal trend that may not be treatment-related (i.e., it also appears in the placebo arm). A statistical graphic that shows each subject's benefit-risk profile over time is also introduced. An approved drug for chronic pain is presented as a worked example. This class of drugs provides a good illustration of a benefit-risk tradeoff because they often produce gastrointestinal side effects along with the desired analgesic effect. Since the treatment is for a chronic condition, one would ideally hope for long-lasting benefits and transient side effects, if any.

Presenter: *Obenchain, Bob* **Affiliation:** *Risk Benefit Statistics LLC*

Abstract Title: ***Making Fair and Efficient Comparisons***

Author(s): *Bob Obenchain**

Abstract:

Session: M28
Cosmopolitan D
Monday, June 21
1:30 p.m.–3:20 p.m.

Federal funding for Comparative Effectiveness Research is now reality. To be genuinely useful, this new research needs improved planning and analysis strategies. Frankly, much US health care outcomes research or pharmacoepidemiology studies of the last 50 years lack credibility. Many relatively sensational publications claimed findings that could not be replicated in follow-up studies. For there to be any genuine hope of improved credibility, major paradigm shifts are mandatory. Researchers need to abandon reliance on global, parametric models that make strong and unrealistic assumptions. Specifically, it's not appropriate to arbitrarily restrict attention to models with only main effects for treatment. Such analyzes (and their ubiquitous p-values) are much too easy to misuse or abuse. At the least, researchers should forsake least squares means and view, instead, a distribution of least squares counter-factual difference estimates. After all, treatment interaction effects form the very basis for patient differential response. Treatment interactions are thus mandatory components of initiatives to inform evidence based medicine and targeted therapeutics. Unfortunately, introducing interaction terms into parametric models one degree-of-freedom at a time can cause blatant model selection bias. Thus we focus here on methods that form patient subgroups to make direct, local comparisons. The corresponding models are simple, non-parametric, nested ANOVAs. These approaches reveal a full distribution of observed treatment effects that can be validated by comparing it with a random permutation distribution and then interpreted much like a Bayesian posterior. Finally, we will also discuss strategies to address the problem of unmeasured confounders.

Presenter: *Offen, Walter* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Flexible Dose Study Designs: Tailoring the Dose to the Patient***

Author(s): *Offen *, Walter W*

Abstract: There is much activity in tailoring the right drug to an individual patient. However, tailoring treatments to optimize outcomes for an individual is a function of not only the right drug, but also the right dose of that right drug. Traditionally the evaluation of dose-effect relationships is conducted by utilizing one or more multiple-arm fixed dose designs. However, such designs only allow inference to average dose effects, which is a sub-optimal inference because typically a single dose is not optimal for all patients with a specific disease. For many diseases and drugs, patients require different doses to achieve their own personalized optimal benefit-risk balance. There are a number of advantages to conducting flexible dose designs which allow such optimization within a patient. They include generally greater drug-placebo differences and fewer dropouts, both of which lead to greater power and/or smaller sample size requirements. Primary statistical comparisons for such designs are simply to compare the patients randomized to study drug with those randomized to placebo. We acknowledge there is valuable information that can be obtained from fixed dose designs. Consequently we are not recommending that all Phase 2 and/or Phase 3 trials be flexible dose designs. Rather, we recommend that regulatory agencies encourage greater utilization of flexible dose designs by sponsors, when appropriate, than the current paradigm.

Session: M21
Network
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Ogawa, Sachio* **Affiliation:** *Eli Lilly Japan K.K.*

Abstract Title: ***Points to Consider when Determining the Sample Size of a Given Region in an Oncology Multi-Regional Clinical Trial***

Author(s): *Sachio Ogawa*, Risa Sekiguchi*

Abstract: Multi-Regional Clinical Trials (MRCT) have increased for simultaneous new drug development in many therapeutic areas. A key issue in MRCTs is how to obtain consistent results among regions. In order to achieve this goal, two types of consistency criteria to assess consistency of efficacy in the region of interest and two types of efficacy criteria to assess efficacy in the region of interest have been proposed for use in MRCTs. Previously, we evaluated the relationship between Japanese subgroup sample size and the probability that the Japan results in an oncology MRCT would satisfy the above efficacy criteria. In that evaluation, a time-to-event variable such as survival time was used as the primary endpoint. In this first approach we wanted to obtain a basic understanding of the methodology and simplified our assumptions in that the control and test treatment hazards were the same for all regions. However, when designing an actual oncology clinical trial, more realistic conditions need to be considered. Specifically, hazards for the control treatment will differ among regions and patients in a given region will enter the trial later than those of other regions. In this presentation, we show the results of a new simulation which uses a more realistic setting and evaluates the impact of these conditions.

Session: T11
Concept AB
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *O'Malley, A James* **Affiliation:** *Harvard Medical School*

Abstract Title: ***Network Meta Analysis Design Adapted to a Historically Controlled Clinical Trial***

Author(s): *A James O'Malley**

Abstract: I describe a novel design for a single armed trial of a new drug coated coronary-artery stent. In order to make comparisons against coronary artery bypass graft (CABG) surgery, two historical trials involving a third treatment (bare metal stenting) are used. Because one of the historical trials is a small study comparing the drug coated stent to a bare metal stent and the other is a large randomized trial comparing the same bare metal stent to CABG, the pairs of treatments form a connected network between the trials. I will discuss the virtues of the network meta-analysis design when individual level data are available and will present the results from the actual analysis of the trial. In particular, I will demonstrate that a different

Session: W27
Cosmopolitan C
Wednesday, June 23
10:10 a.m.–12:00 p.m.

conclusion is obtained than for the standard historical-control analysis.

Presenter: *O'Malley, A James* **Affiliation:** *Harvard Medical School*

Abstract Title: ***Novel Bayesian multivariate hierarchical models of random coefficients of regression models for survey data***

Author(s): *A James O'Malley*; Alan M Zaslavsky*

Abstract: Surveys of healthcare experiences, an important tool for assessment and improvement of healthcare quality, may be adjusted by linear models for differences in characteristics among patients of various healthcare units. Regression coefficients vary across units, inducing a multilevel model structure for individual outcomes. Two models are considered for the level-2 correlation structure of the random coefficients: completely unstructured and separable. The latter assumes that the covariance structure of regression coefficients of six outcome variables on four casemix components can be expressed as the Kronecker product of matrices for the associations across outcome measures and across predictors, potentially making the results more interpretable (31 distinct parameters instead of 300). We also extend this model with a third level to assess the stability of the coefficients across regions or across time. Again, the most general model allows unstructured correlation matrices at levels 2 and 3 while more restrictive and interpretable results are obtained by imposing separable structures at one or both levels or by assuming equal correlation matrices at both levels. We present Bayesian methods for fitting these models, selecting a parsimonious model, comparing models and interpreting results

Session: T16
Studio One
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Park, Jincheol* **Affiliation:** *Texas A&M University*

Abstract Title: ***Bayesian Analysis of Geostatistical Models with an Auxiliary Lattice***

Author(s): *Jincheol Park* and Faming Liang*

Abstract: The Gaussian geostatistical model has been widely used for modeling spatial data. However, this model suffers from a server difficulty in computation: It requires to invert a large covariance matrix. This is not feasible when the number of observations is large. In this talk, we will propose an auxiliary lattice-based approach for tackling the matrix obstacle. In our approach, we introduce an auxiliary lattice to the data and define a Gaussian Markov random field on the auxiliary lattice. By making use of some analytical results on Gaussian Markov random fields, our approach completely avoids the matrix inversion, and thus can be applied to very large datasets.

Session: M36
Cosmopolitan C
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Pena, Edsel A* **Affiliation:** *University of South Carolina*

Abstract Title: ***Classes of Multiple Decision Functions Strongly Controlling FWER and FDR***

Author(s): *Edsel A. Pena*, Joshua Habiger, Wensong Wu*

Abstract: In this talk I will discuss two general classes of multiple decision functions, where each member of the first class strongly controls the family wise error rate (FWER), while each member of the second class strongly controls the false discovery rate (FDR). These classes offer the possibility that an optimal multiple decision function with respect to a pre-specified criterion, such as missed discovery rate (MDR), could be found within these classes. Such multiple decision functions can be utilized in multiple testing, specifically, but not limited to, the analysis of "large M, small n" microarray data sets. The utility of these classes will be demonstrated by obtaining "power-enhanced" multiple decision functions.

Session: T17
Regency B
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Pennell, Michael* **Affiliation:** *The Ohio State University*

Abstract Title: ***Cutoff Based Designs for Community Intervention Studies***

Author(s): *Michael Pennell*, Erinn Hade, David Murray, Dale Rhoda*

Abstract: Public health interventions are often designed to target communities defined either geographically (e.g., cities, counties) or socially (e.g., schools or workplaces). The group randomized trial (GRT) is regarded as the gold standard for evaluating these interventions. However, community leaders may object to randomization as groups may be denied a

Session: M38
Studio One

Monday, June 21
3:40 p.m.–5:30 p.m.

potentially beneficial intervention. Under a regression discontinuity design (RDD), individuals may be assigned to treatment based on the levels of a pretest measure, thereby allowing those most in need of the treatment to receive it. In this talk, I will consider analysis, power, and sample size issues in applying the RDD and related cutoff designs to community-based intervention studies. I will examine the power of these designs as a function of intraclass correlation, number of groups, and number of members per group and compare results to the traditional GRT.

Presenter: *Qiao, Xingye* **Affiliation:** *UNC Chapel Hill*

Abstract Title: ***Pairwise variable selection for classification***

Author(s): *Xingye Qiao*, Yufeng Liu and J. S. Marron*

Abstract:

Session: W21
Regenecy A
Wednesday, June 23
10:10 a.m.–12:00 p.m.

While traditional marginal variable selection methods have the merits of convenient implementation and good interpretability, they do not take the joint effects among variables into account. In some situations, variables which have strong joint effects can be passed over by marginal methods because of their small marginal effects. In the context of binary classification in supervised learning, we develop a novel method of pairwise variable selection, based on a within-class permutation test to evaluate the statistical significance of joint effects. Moreover, we introduce a new notion of variable selection quality, bivariate False Discovery Rate (biFDR), and provide an estimation procedure for biFDR. Simulated examples and real data applications are analyzed to demonstrate the usefulness of the proposed approach. This is a joint work with Yufeng Liu and J. S. Marron

---

Presenter: *Randolph, Tim* **Affiliation:** *Fred Hutchinson Cancer Research Center*

Abstract Title: ***Structured penalties for functional linear models---partially empirical eigenvectors for regression***

Author(s): *Tim Randolph*, Jaroslaw Harezlak, Ziding Feng*

Abstract:

Session: T34
Concept CD
Tuesday, June 22
3:40 p.m.–5:30 p.m.

A challenge with functional data is incorporating the spatial structure of the functions into the analysis. Common approaches to this ill-posed problem reduce the dimension by regressing on principal components or by projection onto the span of a prescribed basis. We present an approach to functional linear models which recognizes that the spatial structure obtained by the joint spectral properties of the predictors and a linear penalty operator provides a rigorous way to incorporate structure directly into the estimation process. In this sense, the components in the regression are `partially empirical'. The framework for this is provided by the generalized singular value decomposition which clarifies the penalized estimation process: it informs the choice of penalty by making explicit the joint influence of the penalty and predictors, including the bias, variance and performance of the estimated coefficient function.

Presenter: *Reiss, Philip* **Affiliation:** *New York University & Nathan Kline Inst.*

Abstract Title: ***Extensions of function-on-scalar linear regression***

Author(s): *Philip T. Reiss* and Lei Huang*

Abstract:

Session: T34
Concept CD
Tuesday, June 22
3:40 p.m.–5:30 p.m.

A popular way to regress functional responses on scalar predictors is the basis function/roughness penalty estimator introduced by Ramsay and Silverman. We show how a modified derivation of the estimator leads to multiple smoothing parameter selection that is computationally feasible. Our approach also motivates two extensions of the functional linear model. The first yields fitted values that have smooth dependence on both the function argument and the scalar predictor(s), which, as we demonstrate, can have major advantages over linearity with respect to the predictors. The second uses an elastic net-type penalty to obtain sparse estimates of the coefficient functions.

Presenter: *Reiter, Jerome* **Affiliation:** *Duke University*
Abstract Title: ***Adapting Multiple Imputation to Protect Confidential Data***
Author(s): *Jerome Reiter*
Abstract: When sharing data with the public, statistical agencies can replace sensitive or identifying values with multiple imputations drawn from models estimated with the original data. This approach, often called synthetic data, generally requires methods of inference that differ from the combining rules developed by Rubin (1987) for missing data. I present a recent adaptation of multiple imputation useful for creating public use data from censuses or large administrative databases. Along the way, I explain why new combining rules are needed for this and other adaptations of multiple imputation.

Session: M17
Regency B
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Robins, James* **Affiliation:** *Harvard School of Public Health*
Abstract Title: ***Estimation of Optimal treatment Strategies from Longitudinal Data***
Author(s): *James M. Robins*, Andrea Rotnitzky*
Abstract: We review recent developments in the estimation of an optimal treatment strategy or regime from longitudinal data collected in an observational study. We also propose novel methods for using the data obtained from an observational database in one health-care system to determine the optimal treatment regime for biologically similar subjects in a second health-care system when, for cultural, logistical, or financial reasons, the two health-care systems differ (and will continue to differ) in the frequency of, and reasons for, both laboratory tests and physician visits. Finally, we propose a novel method for estimating the optimal timing of expensive and/or painful diagnostic or prognostic tests. Diagnostic or prognostic tests are only useful in so far as they help a physician to determine the optimal dosing strategy, by providing information on both the current health state and the prognosis of a patient because, in contrast to drug therapies, these tests have no direct causal effect on disease progression. Our new method explicitly incorporates this no direct effect restriction.

Session: T12
Concept CD
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Rocha, Guilherme* **Affiliation:** *Indiana University*
Abstract Title: ***Model selection consistency for l1-penalized M-estimators***
Author(s): *Guilherme *Rocha, Xing Wang and Bin Yu*
Abstract: Since its early use in least squares regression problems, the l1-penalization framework for variable selection has been employed in conjunction with a wide range of loss functions encompassing regression, classification and survival analysis. While a well developed theory exists for the l1-penalized least squares estimates, few results concern the model selection behavior of l1-penalized estimates for general loss functions. We derive explicit necessary and sufficient generalized irrepresentability (GI) conditions for l1-penalized parametric M-estimates to consistently select the components of a model as well as their sign. In general, the GI conditions depend on the Hessian of the risk function at the true value of the unknown parameter. Under Gaussian predictors, we obtain a set of conditions allowing the GI conditions to be re-expressed solely in terms of the second moment of the predictors. Applications to classification using l1-penalized SVMs and logistic regression are used as examples.

Session: T28
Studio One
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Rosales, Karen* **Affiliation:** *Western Michigan University*
Abstract Title: ***Confidence Intervals for Difference of Means of Zero-Clustered Data***
Author(s): *Karen Rosales*, Joshua Naranjo*
Abstract: In many applications, the data has a cluster of observations at zero, but is otherwise continuous and non-negative. This presents difficulties for both classical and nonparametric methods because of the large number of ties at zero, and the skewness of the nonzero observations. We discuss proposals for estimating the difference in means, and investigate their properties through a simulation study.

Session: T33
Network
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Rothman, Adam* **Affiliation:** *University of Michigan*
Abstract Title: ***Sparse multivariate regression with covariance estimation***
Author(s): *Adam J. Rothman*, Elizaveta Levina, Ji Zhu*
Abstract: We propose a procedure for constructing a sparse estimator of a multivariate regression coefficient matrix that accounts for correlation of the response variables. This method, which we call multivariate regression with covariance estimation (MRCE), involves penalized likelihood with simultaneous estimation of the regression coefficients and the covariance structure. An efficient optimization algorithm and a fast approximation are developed for computing MRCE. Using simulation studies, we show that the proposed method outperforms relevant competitors when the responses are highly correlated. We also apply the new method to a finance example on predicting asset returns.

Session: T27
Cosmopolitan D
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Rubin, Donald* **Affiliation:** *Harvard University*
Abstract Title: ***FOR OBJECTIVE CAUSAL INFERENCE, DESIGN TRUMPS ANALYSIS***
Author(s): *Donald B. Rubin*
Abstract: For obtaining causal inferences that are objective, and therefore have the best chance of revealing scientific truths, carefully designed and executed randomized experiments are generally considered to be the gold standard. Observational studies, in contrast, are generally fraught with problems that compromise any claim for objectivity of the resulting causal inferences. The thesis here is that observational studies have to be carefully designed to approximate randomized experiments, in particular, without examining any final outcome data. Often a candidate data set will have to be rejected as inadequate because of lack of data on key covariates, or because of lack of overlap in the distributions of key covariates between treatment and control groups, often revealed by careful propensity score analyses. Sometimes the template for the approximating randomized experiment will have to be altered, and the use of principal stratification can be helpful in doing this. These issues are discussed and illustrated using the framework of potential outcomes to define causal effects, which greatly clarifies critical issues.

Session: M28
Cosmopolitan D
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Sanil, Ashish* **Affiliation:** *Berry Consultants*
Abstract Title: ***Aspects of Development of Statistical Software in the Design and Execution of Adaptive Bayesian Clin***
Author(s): *Ashish Sanil*
Abstract: The Bayesian approach is particularly well-suited for use in adaptive clinical trials since it provides a mathematically rigorous and principled methodology for incorporating information from accruing data to update uncertainty about model parameters in order to make optimal decisions. However, the design and analysis of all but the simplest of trials in the Bayesian setting demand a significant computational effort that usually require development of customized software. This talk will cover the design and implementation of such software constructed primarily for evaluating the operating characteristics of the designs via trial simulation. Key software components as well as particular implementation issues will be discussed.

Session: W11
Regency A
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Schaubel, Douglas* **Affiliation:** *University of Michigan*
Abstract Title: ***Estimating the Average Effect on the Survival Function of a Time-Dependent Treatment***
Author(s): *D.E. Schaubel* and J.D. Kalbfleisch*
Abstract: In several areas of medicine, the treatment of interest is time-dependent. For example, patients with end-stage renal disease typically receive dialysis for a period of time before undergoing kidney transplantation. When the time to failure (e.g., death) is potentially censored, analyses of such data has traditionally involved Cox regression using time-dependent treatment indicators, with non-proportionality addressed through a time-dependent treatment effect. However, investigators usually prefer treatments to be contrasted in terms

Session: M14
Cosmopolitan C
Monday, June 21
10:20 a.m.–12:10 p.m.

of survival (not hazard) functions; particularly in cases where the treatment effect on the hazard function is not constant over time. We develop semiparametric methods to estimate the effect of a time-dependent treatment on survival and restricted mean residual lifetime. Asymptotic properties of the proposed estimators are derived, with finite-sample characteristics evaluated through simulation. The proposed methods are applied to end-stage renal disease data from a national organ transplant registry.

Presenter: *Sethuraman, Venkat* **Affiliation:** *Novartis Oncology*

Abstract Title: ***Statistical Consideration in Testing for Assay Sensitivity in a 'thorough' QT Study***

Author(s): *Venkat Sethuraman*, Shuang Wu and Jixian Wang*

Abstract: The current guidelines, ICH E14, for the evaluation of non-antiarrhythmic compounds recommends a "thorough" QT study (TQT) to be conducted during clinical development. One of the treatment arms in a TQT study is a positive control, e.g., Moxifloxacin. The current hypothesis for assay sensitivity involves testing at pre-selected time points, where electrocardiograms (ECG) are obtained, if the mean baseline adjusted differences between positive control and placebo are greater than 5 ms. Since multiple time points are tested, the overall type I error rate should be adjusted using an appropriate multiple comparison procedure. We consider some recently proposed tests (Zhang, 2008) and explore the use of some standard global test procedures such as O'Brien's ordinary least squares (OLS) and Läuter's standardized sum (SS) tests for testing sensitivity of positive control. The power of several test procedures are evaluated using simulation.

Session: T13
Network
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Shao, Qin* **Affiliation:** *The University of Toledo*

Abstract Title: ***Seasonality Analysis of Time Series in Partial Linear Models***

Author(s): *Qin Shao*

Abstract: Seasonality analysis is one of the classic topics in time series. The talk will focus on the techniques for seasonality analysis when the trend function is unspecified and the application to the monthly global land-ocean temperature anomaly indexes.

Session: W26
Studio One
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Shao, Jun* **Affiliation:** *University of Wisconsin-Madison*

Abstract Title: ***A multiple testing procedure for combination drugs with two study endpoints***

Author(s): *Jun Shao*, Alan Chiang, Sheng Zhang, Jiwei Zhao*

Abstract: We consider a combination drug containing two agents in a clinical trial with some combinations of dose levels. The hypothesis of interest is whether a combination drug with low dose levels is better than two single-agent drugs with the best dose levels. In addition, two endpoints are considered, such as the low-density lipoprotein (bad cholesterol) and the high-density lipoprotein (good cholesterol). We propose a multiple testing procedure that consists of the MIN test to obtain p-values from different dose combinations and a bootstrap step-down method, an improved version of Holm's step-down testing procedure. The proposed test controls the family-wise error rate and is less conservative and hence more powerful than Holm's method.

Session: M13
Network
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Shen, Lei* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Modern Statistical Methods for Subgroup Identification in Clinical Trials***

Author(s): *Lei Shen*

Abstract: Traditional subgroup analyses in clinical trials based primarily on interaction testing suffer from a number of statistical and strategic issues, such as multiplicity, inadequate power and inability to study higher order interactions. To identify meaningful subgroups in clinical trials as a way of developing tailored therapies, it is therefore critical that more advanced methodologies be utilized for this effort. I will present some approaches to identify and

Session: M22
Regency A
Monday, June 21

1:30 p.m.–3:20 p.m.

confirm subgroups that have been applied to clinical trials. Various aspects of type I error control will be discussed, with both proposed methods and open problems.

Presenter: *Shen, Gang* **Affiliation:** *North Dakota State University*

Abstract Title: ***Developing A New BIC for Detecting Change-points***

Author(s): *Shen, Gang & Ghosh, Jayanta*

Abstract:

Session: W26
Studio One
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Usual derivation of BIC for the marginal likelihood of a model or hypothesis via Laplace approximation does not hold for a change-point which is a discrete parameter. We provide an analogue lBIC, which is a lower bound to the marginal likelihood of a model with change points and has an approximation error up to Op(1) like standard Schwartz BIC. Several applications are provided covering simulated r.v.'s and real financial figures on short-term interest rate.

Presenter: *Shen, Changyu* **Affiliation:** *Indiana University*

Abstract Title: ***Sensitivity analysis for causal inference using inverse probability weighting***

Author(s): *Changyu Shen*, Xiaochun Li, Lingling Li, Martin Were*

Abstract:

Session: T12
Concept CD
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Evaluation of impact of potential uncontrolled confounding is an important step for causal inference based on observational studies. In this talk, we introduce a sensitivity analysis approach for causal inference that is based on inverse probability weighting. Specifically, inferential bias due to uncontrolled confounding is driven by two parameters that govern the magnitude of the variation of the multiplicative errors of the propensity score and their correlations with the potential outcomes. Plausible values of the two governing parameters allow one to examine potential bias. We will discuss the general methodology that allows both parametric and non-parametric analysis. We will also introduce a specific parametric model that offers a mechanistic view on how the uncontrolled confounding may bias the inference and thereby provides some guidance on how to postulate non-estimable parameters. Our method can be readily applied to both binary and continuous outcomes and depends on the covariates only through the propensity score that can be estimated by any parametric or non-parametric method. We illustrate our method with one real data set.

Presenter: *Shen, Jeremy* **Affiliation:** *Stanford University*

Abstract Title: ***Copy Number Profiling Using Next-Generation DNA Sequencing with Change-Point Methods***

Author(s): *Jeremy J. Shen*, Nancy R. Zhang*

Abstract:

Session: T35
Regency B
Tuesday, June 22
3:40 p.m.–5:30 p.m.

As DNA sequencing capacity continue to grow at an exciting pace, there is increasing interest to use sequencing to study genome structural aberrations including copy number variation (CNV). We will present a change-point approach to the detection of CNV using the read depth information from Next-Generation Sequencing of matched case-control DNA samples. We model the positions of mapped reads as non-homogeneous Poisson processes and formulate the problem as the calling of change points in the rate difference between the case and control processes. Instead of binning the range of read positions, we propose the use of sliding window of varying sizes and use Circular Binary Segmentation (CBS) to estimate change points. We derived score statistic and a binomial exact statistic for the scan. As in all change-point problems, one seeks not only to estimate the boundary of change points but also the number of change points. We devised a modified BIC (mBIC) for change points on Poisson Processes to select the appropriate change-point model. We will also discuss the computational aspect of our algorithm, and demonstrate its performance on simulation and existing cancer genome sequencing data.

Presenter: *Skrivanek, Zachary* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***A Seamless 2/3 Design Incorporating a Clinical Utility Index***

Author(s): *Zachary Skrivanek*, PhD; Mary Jane Geiger, MD, PhD; Brenda Gaydos, PhD*

Abstract: Diabetes is a disease with well understood and validated biomarkers that have been used in

Session: W22
Concept AB
Wednesday, June 23
10:10 a.m.–12:00 p.m.

clinical trials for decades to assess the safety and efficacy of diabetes therapies. Consequently, adaptive designs are well suited for learning about the dose response of a diabetes drug and providing confirmatory evidence for the safety and efficacy of the optimal dose(s). We will discuss the design of a novel adaptive, inferentially seamless phase 2/3 study for an experimental drug to treat diabetes. The design employs a Bayesian analytical approach to allocate patients to a set of doses of the experimental drug and to determine if there are 1-2 doses that could be continued to be studied to confirm safety and efficacy of those doses. That is, this single study comprises both dose decision and dose confirmation, traditionally determined by separate, fixed-design phase 2 and 3 studies. The preference for a dose is determined by a Clinical Utility Index, which balances the select efficacy and safety measures. The algorithm is completely prespecified, and the operating characteristics were assessed via simulation. This design was developed through iterative simulations that involved the same scientists who would have input as to what doses would be selected for traditional, separate confirmatory trials. It involved much more a priori planning than would be required for a typical fixed design. We will discuss the differences between this approach of designing a study and the traditional fixed-design approach. We will also discuss the mathematical form of the Clinical Utility Index and compare it to alternative derivations.

Presenter: *Small, Dylan* **Affiliation:** *University of Pennsylvania*

Abstract Title: ***Causal Inference for Continuous Time Longitudinal Data When Covariates Are Observed Only at Discrete Times***

Author(s): *Mingyuan Zhang, Marshall Joffe, Dylan Small**

Abstract: Most of the work on g-estimation for causal inference in longitudinal data assumes a discrete time underlying data generating process. However, in some studies, it is more reasonable to assume that the data are generated from a continuous time process, but the covariates are only observable at discrete times. For this setting, we study the assumptions needed for discrete time g-estimation to provide consistent estimates and present a new method that provides consistent estimates under weaker assumptions than usual discrete time g-estimation. We use our new method to study the effect of diarrhea on children's height, using a data set collected following a massive flood in Bangladesh.

Session: M31
Concept AB
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Song, Yang* **Affiliation:** *Merck Research Labs*

Abstract Title: ***Optimal Cost-effective Designs for Phase II Proof of Concept Trials***

Author(s): *Cong Chen, Yang Song**

Abstract: Unlike a confirmatory registration trial, a Phase II Proof of Concept trial is exploratory in nature and sponsors of such trials have the liberty to choose the type I error rate and the power. The decision is largely driven by the perceived probability of having a truly active treatment per patient exposure (a surrogate measure to development cost), which is naturally captured in an efficiency score that measures cost-effectiveness of the trial. Optimization of the score function leads to type I error rate and power (and therefore sample size) for the trial that is most cost-effective. This in turn leads to cost-effective Go-No Go criteria for development decisions. The idea is applied to derive optimal trial-level, program-level and franchise-level design strategies. The idea is also generalized to evaluate cost effectiveness of different trial designs in situations where the treatment being developed is potentially more efficacious in a subset of patient population defined by biomarkers.

Session: W22
Concept AB
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Stamey, James* **Affiliation:** *Baylor University*

Abstract Title: ***Bayesian approaches to handle pharmacoepidemiological data with an unmeasured confounder and response misclassification***

Author(s): *James Stamey*, John Seaman, Karen Price, Doug Faries*

Abstract: Epidemiological data is well known to suffer from biased due to both measurement error and variables that are unmeasured. We consider the case of Poisson response models where the

Session: T32

Regency A
Tuesday, June 22
3:40 p.m.–5:30 p.m.

observed count is subject to underreporting and a binary confounder is unmeasured. We propose a Bayesian approach using validation data to correct for both sources of error. Strengths and limitations of the method are discussed.

Presenter: *Stein, Michael* **Affiliation:** *University of Chicago*

Abstract Title: ***On a class of space-time intrinsic random functions***

Author(s): *Michael Stein**

Abstract:

Session: T36
Cosmopolitan C
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Power law generalized covariance functions provide a simple model for describing the local behavior of an isotropic random field. This work seeks to extend this class of covariance functions to spatial-temporal processes for which the degree of smoothness in space and in time may differ while maintaining other desirable properties for the covariance functions including computational tractability.

Presenter: *Stevens, John* **Affiliation:** *Utah State University*

Abstract Title: ***Assessing and Accounting for Dependence in Gene Expression Summaries***

Author(s): *John R. Stevens*, Gabriel Nicholas*

Abstract:

Session: W16
Network
Wednesday, June 23
8:00 a.m.–9:50 a.m.

In studies using oligonucleotide arrays, statistical methods to test for differential expression traditionally assume that each gene's expression summaries are independent across arrays. When certain preprocessing methods are used to obtain those estimates, this assumption is not necessarily true. We introduce a distance measure to assess the dependence of summaries across arrays for each gene, for any preprocessing method. We discuss the relative performance of several common preprocessing methods with respect to this measure, and show how accounting for these dependencies in a test for differential expression affects power.

Presenter: *Stroud, Jonathan* **Affiliation:** *George Washington University*

Abstract Title: ***Parameter Estimation for Large Gridded Spatial Datasets***

Author(s): *Jonathan Stroud*

Abstract:

Session: W15
Cosmopolitan C
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Spatial lattice data arise in many applications, including environmental science and medical imaging. The datasets in these fields are often large, making likelihood-based parameter estimation quite difficult. This paper proposes a new approach to maximum likelihood and Bayesian parameter estimation for Gaussian processes observed on a large lattice. Our approach uses data augmentation and circulant embedding of the covariance matrix, and provides exact inference for the parameters and the interpolations. We propose an MCMC approach for Bayesian inference, and a stochastic EM algorithm for likelihood inference. Using simulated data and a real application to satellite images, we show that the method outperforms standard methods including approximate likelihood and the Whittle approximation.

Presenter: *Su, Haiyan* **Affiliation:** *Montclair State University*

Abstract Title: ***Semiparametric hybrid empirical likelihood inference for two sample comparison with censored data***

Author(s): *Haiyan Su*, Mai Zhou, Hua Liang*

Abstract:

Session: W18
Studio One
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Two-sample comparison problems are often encountered in practical projects and have widely been studied in literature. Owing to practical demands, the research for this topic under special settings such as a semiparametric framework have also attracted great attentions. Zhou and Liang (2005) proposed an empirical likelihood-based semi-parametric inference for the comparison of treatment effects in a two-sample problem with censored data. However, their approach is actually a pseudo-empirical likelihood and the method may not be fully efficient. In this study, we develop a new empirical likelihood-based inference under more general framework by using the hazard formulation of censored data for two sample semi-parametric hybrid models. We demonstrate that our empirical likelihood

statistic converges to a standard chi-squared distribution under the null hypothesis. We further illustrate the use of the proposed test by testing the ROC curve with censored data, among others. Numerical performance of the proposed method is also examined.

Presenter: *Sun, Wenguang* **Affiliation:** *North Carolina State University*

Abstract Title: ***Optimal Screening for Sparse Signals***

Author(s): *T. Tony Cai, Wenguang Sun**

Abstract: In large scale statistical inference problems,it is common that signals are sparse and it is desirable to significantly reduce the original large data set to a much smaller subset for further study. In this talk, we consider two related data screening problems: One is to find the smallest subset such that it contains all signals with high probability and another is to find the largest subset so that it virtually contains only signals (i.e. the proportion of the nulls in the subset is negligible.) These screening problems are closely connected to but distinct from the more conventional detection or multiple testing problems. We shall discuss precise conditions under which these goals are achievable and construct screening procedures that have near optimality properties.

Session: M27
Regency B
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Sun, Wei* **Affiliation:** *University of North Carolina*

Abstract Title: ***Statistical methods for RNA-seq data analysis***

Author(s): *Wei Sun**

Abstract: RNA-seq data is not only more accurate and more sensitive to measure gene expression, but also provides important allele-specific information, which can be used to identify cis-eQTL (gene expression quantitative trait loci) and imprinting. We first discuss statistical methods for analyzing RNA-seq data from F1 cross of inbred mice, which is a relatively simple situation since two parental haplotypes are known. Next we study possible strategies for analyzing RNA-seq data from human population.

Session: T15
Regency A
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Sun, Yanqing* **Affiliation:** *The University of North Carolina at Char*

Abstract Title: ***Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood***

Author(s): **Yanqing Sun, Rajeshwari Sundaram and Yichuan Zhao*

Abstract: The Cox model with time-dependent coefficients has been studied by a number of authors recently. In this paper, we develop empirical likelihood (EL) pointwise confidence regions for the time-dependent regression coefficients via local partial likelihood smoothing. The EL simultaneous confidence bands for a linear combination of the coefficients are also derived based on the strong approximation methods. The empirical likelihood ratio is formulated through the local partial log-likelihood for the regression coefficient functions. Our numerical studies indicate that the EL pointwise/simultaneous confidence regions/bands have satisfactory finite sample performances. Compared with the confidence regions derived directly based on the asymptotic normal distribution of the local constant estimator, the EL confidence regions are overall tighter and can better capture the curvature of the underlying regression coefficient functions. Two data sets, the gastric cancer data and the Mayo Clinic primary biliary cirrhosis data, are analyzed using the proposed method.

Session: W25
Concept CD
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Sun, Shuxia* **Affiliation:** *Wright State University*

Abstract Title: ***Asymptotic Properties for Lp-Norms of Error Density Estimators in Nonlinear Autoregressive Time Series Models***

Author(s): *Shuxia Sun*, Fuxia Cheng*

Abstract: Horvath and Zitikis (2003) showed that, in the first-order autoregressive models, the asymptotic behavior of the Lp-distance of the kernel density estimator of residuals and the density function itself is the same as in the i.i.d. case. In this talk, we present the asymptotic property of the Lp-norm of the distance between the error density function and its kernel estimator in nonlinear autoregressive time series models. This is joint work with Fuxia

Session: T24
Concept CD
Tuesday, June 22

1:30 p.m.–3:20 p.m. Cheng.

Presenter: *Takeuchi, Masahiro* **Affiliation:** *Kitasato University*

Abstract Title: ***Simultaneous Global Clinical Trials-Statistical Practice on the Evaluation of Drug Profile in Asian Studies***

Author(s): *Masahiro Takeuchi**

Abstract: Global clinical trials with a common protocol, based on the assumption of no racial differences among regions, have become popular for new drug applications. We will investigate statistical strategies and study designs adjusting for intrinsic/extrinsic factors in global clinical trials.

Session: T11
Concept AB
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Tamura, Roy* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Placebo Response and the Sequential Parallel Design - Statistical Considerations***

Author(s): *Roy N. Tamura*

Abstract: A difficult problem in many psychiatric diseases is the highly variable placebo response rate. High placebo response contributes to a high Type II error rate for known active compounds. The sequential parallel design has been proposed as an alternative to the traditional parallel design clinical trial. Usage of the sequential parallel design in psychiatric clinical trials is growing. In this talk, I will review the design and the underlying assumptions in the design, discuss the efficiency of the design, and lay out some remaining statistical issues which are still unresolved.

Session: W23
Regency B
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Tarpey, Thaddeus* **Affiliation:** *Wright State University*

Abstract Title: ***Optimal Partitioning for Linear Mixed Effects Models: Applications to Identifying Placebo Responders***

Author(s): *Thaddeus Tarpey*, Eva Petkova, Yimeng Lu, and Usha Govindarajulu*

Abstract: A long--standing problem in clinical research is distinguishing drug treated subjects that respond due to specific effects of the drug from those that respond to non-specific (or placebo) effects of the treatment. Linear mixed effect models are commonly used to model longitudinal clinical trial data. A solution to this problem is presented using an optimal partitioning methodology for linear mixed effects models. The approach is compared and contrasted with a growth mixture model approach. The methodology is applied to a two-phase depression clinical trial where subjects in a first phase were treated openly for 12 weeks with fluoxetine followed by a double blind discontinuation phase where responders to treatment in the first phase were randomized to either stay on fluoxetine or switched to a placebo.

Session: W23
Regency B
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Taylor, Jeremy* **Affiliation:** *University of Michigan*

Abstract Title: ***Using joint longitudinal and survival models for individual prediction.***

Author(s): *Jeremy M G Taylor*

Abstract: In this talk I will illustrate the use of joint longitudinal and survival models for individual prediction of future longitudinal and event time data. The model is motivated by a prostate cancer application for patients previously treated with radiation therapy, PSA is the longitudinal variable and the event of interest is recurrence of the cancer. Features of the model include random effects, non-linear profiles for the longitudinal variable, proportional hazards model with values and derivatives of PSA as time dependent variables, dependent censoring. Estimation is via Markov chain Monte Carlo methods. I will discuss how to predict the residual time distribution for a new individual with some longitudinal data, and validation of the model using training and testing datasets.

Session: M12
Concept CD
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Taylor, Jeremy* **Affiliation:** *University of Michigan*
Abstract Title: ***Finding and validating subgroups in clinical trials***
Author(s): *Jeremy M G Taylor, Jared C Foster, Stephen J Ruberg*
Abstract: We consider the problem of subgroups of patients who may have an enhanced treatment effect in a randomized clinical trial, and it is desirable that the subgroup be defined by a limited number of covariates. The development of a standard, pre-determined strategy may help to avoid the well-known dangers of subset analysis. We present two methods developed to find subgroups of enhanced treatment effect. The first method involves the use of logistic regression and forward selection, with the largest possible model being that with all main effects, one and two-way interaction terms of the covariates and the treatment group indicator. The second method, referred to as "Virtual Twins", involves predicting response probabilities for treatment and control "twins" for each subject. The difference in these probabilities is then used as the outcome in a tree, which can potentially include any set of the covariates. Simulation studies are presented for situations in which there are and are not true subgroups of enhanced treatment effect, and the methods are compared using a variety of metrics, including area under the curve, sensitivity, specificity, positive and negative predicted values, and a cross-validation-based estimate of treatment effect.

Session: M22
Regency A
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Tchetgen Tchetgen, Eric* **Affiliation:** *Harvard University*
Abstract Title: ***Double-robust adjustment for confounding in cohort and case-control studies***
Author(s): *Eric J Tchetgen Tchetgen* ; Andrea Rotnitzky*
Abstract: Modern epidemiologic studies often aim to evaluate the causal e¤ect of a point exposure on the risk of a disease from cohort or case-control observational data. Because confounding bias is of serious concern in such non-experimental studies, investigators routinely adjust for a large number of potential confounders in a logistic regression analysis of the e¤ect of exposure on disease outcome. Unfortunately, when confounders are not correctly modeled, standard logistic regression is likely biased in its estimate of the effect of exposure, potentially leading to erroneous conclusions. We partially resolve this serious limitation of standard logistic regression analysis with a new approach which carefully combines standard logistic regression with a logistic regression analysis in which exposure is the dependent variable and the outcome and confounders are the independent variables. As a result, we obtain a correct estimate of the exposure-outcome odds ratio, if either the standard logistic regression of the outcome given exposure and confounding factors is correct, or the regression model of exposure given the outcome and confounding factors is correct but not necessarily both. Thus, the resulting estimator is so- called double-robust, and is also often most e¢ cient among all double-robust estimators. The approach is general in that it applies to both cohort and case-control studies whether the design of the study is matched or unmatched on a subset of covariates.

Session: T12
Concept CD
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Thall, Peter* **Affiliation:** *M.D. Anderson Cancer Center*
Abstract Title: ***A Bayesian Geometric Phase II-III Select-and-Test Design Based On Treatment Failure Time and Toxicity***
Author(s): *Peter F. Thall*, Leiko Wooten, Xuemei Wang, Hoang Nguyen, Johannes Wolff*
Abstract: The problem of comparing several experimental treatments to a standard treatment is common in medical research. To avoid selection bias, various randomized two-stage and multi-stage phase II/III designs have been proposed that select one or more promising experimental treatments and compare them to the standard while controlling overall type I and type II error rates. These designs accommodate univariate outcomes and assume that patients are homogeneous. We address the more complex problem where the goal is to increase average treatment failure time and control toxicity while accounting for patient heterogeneity. We present a hybrid two-stage phase II/III design based on two-dimensional

Session: M11
Concept AB
Monday, June 21
10:20 a.m.–12:10 p.m.

treatment effect parameters obtained from Bayesian regression models for failure time and the probability of toxicity that account for prognostic covariates. Decisions at each stage of the trial are based on the posterior odds of alternative and null two-dimensional covariate-adjusted treatment parameter sets. The alternative parameter set is constructed from elicited target parameter pairs. Design parameters are chosen to minimize expected sample size subject to overall size and generalized power constraints. The method is illustrated by a trial of several chemotherapy combinations for treating pediatric brain tumors.

Presenter: *Tong, Liping* **Affiliation:** *Loyola University Chicago*

Abstract Title: ***A Joint Statistical Model of Social Behavior and Social Network***

Author(s): *Liping Tong, David Shoham, Jie Yang, Richard S Cooper*

Abstract: Individual social behavior, such as smoking, drinking, choice of meals, can be strongly influenced by the behavior of their friends. Meantime, the choice of friends can also be influenced by the status of their social behavior. The two effects, selection (effect of social behavior on the formation of a network) and influence (effect of social network on the status of social behavior), have been identified in many studies using various models. However, selection and influence never take effect independently. To study the interdependence of social behavior and social network, Snidjers et al. has developed a family of co-evolution models, which is the only statistical modeling method currently available to deal with longitudinal social network and behavior data. Unfortunately, the theoretical properties of their model are not clear yet. Besides, it is hard to say whether this stochastic process has a stationary distribution or not. In this paper, we first show how to modify Snidjers' model to guarantee a converging property. Then we propose a joint statistical model of social network and behavior variables, from which we are able to construct a stochastic process that converges to a desired distribution.

Session: W26
Studio One
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Tsai, Kao-Tai* **Affiliation:** *Frontier Informatics Service*

Abstract Title: ***The Ultimate Professional Statisticians: Thinkers, Practitioners, and Leaders***

Author(s): *Kao-Tai Tsai*

Abstract: In this fast growth of advanced information and global competition era, the statistics profession is at a critically important point in its history for strategic decision-making. In the decades up to the 1970s, the statistics profession primarily focused on the development of the statistical theory to solve fundamental problems encountered in R&D, manufacturing and other optimization processes. However, in order to meet current societal needs, the statistics profession needs to adjust practices and perform more like engineering with critical statistical thinking and focus on how to best utilize known scientific and statistical principles for the greater benefit of organizations and the general public. In addition, the evidence-based practice of statistics to integrate the best research evidence with statistical expertise and value to our clients should also be at the frontier of consideration. In this presentation, we discuss how well the statistics profession does in the industrial environment, how the statistics profession has been perceived by other professions, the challenges that face the statistics profession, and how to better steer the statistics profession to fit current and future needs of our clients in various organizations. We emphasize the critical reality that, in order to have a prosperous and respectable profession, statisticians need to modify the traditional model of thinking and practice, expand our circle of influence through learning and interaction with other disciplines, be more proactive in outreach to society and provide effective leadership.

Session: M23
Concept AB
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Tseng, George* **Affiliation:** *Dept of Biostatistics, U of Pittsburgh*

Abstract Title: ***Genomic meta-analysis for dimension reduction and gene clustering***

Author(s): *Dongwan Kang and George C Tseng**

Abstract: As increasing number of gene expression profiles are accumulated in the public domain, statistical meta-analysis for combining multiple genomic studies have drawn increasing

Session: T35
Regency B
Tuesday, June 22
3:40 p.m.–5:30 p.m.

attention. Currently, major attention in the literature has been focused on biomarker detection. In this talk, we will present some on-going work of genomic meta-analysis for dimension reduction and gene clustering. Applications to yeast cell cycle data, yeast environmental perturbation data and cancer data sets will be demonstrated.

Presenter: *Tsong, Yi* **Affiliation:** *FDA*

Abstract Title: ***Advanced and Adaptive Designs of thorough QT studies***

Author(s): *Tsong, Yi*

Abstract:

Session: M23
Concept AB
Monday, June 21
1:30 p.m.–3:20 p.m.

The objective of a thorough QT clinical trial is to demonstrate that a test treatment will not induce prolongation of QT interval in healthy subjects. This is carried out by showing the maximum mean effect on QT interval adjusted by the time-matched placebo effect is less than 10 ms. On the other hand, in order to minimize the chance of false negative claim when the selected study subjects do not respond to treatment with known QT prolongation effect, the trial is designed with an active control treatment. The negative conclusion of QT prolongation induced by the test treatment is validated if it is demonstrated that the maximum QT prolongation effect induced by the positive control treatment after adjusted for the placebo effect at the matched time point is greater than 5 ms. Sample size requirement for the two tests may be different. In order to improve the efficiency of trial and to reduce the number of subjects exposed to the positive control, alternative designs were proposed. Such designs include hybrid design which combines parallel-arm and crossover and trial with unequal number of subjects exposed to positive control and test treatments. Furthermore, since non-inferiority and validation tests are to be carried out with equal type I error rate of 0.5%, they may be carried out hierarchically in an adaptive fashion. In this presentation, we will discuss the advantages and limitations of various advanced and adaptive designs.

Presenter: *Tu, Wanzhu* **Affiliation:** *Indiana University School of Medicine*

Abstract Title: ***Adolescent Blood Pressure Development: A Functional Data Analysis Perspective***

Author(s): *Wanzhu Tu, Jaroslaw Harezlak*

Abstract:

Session: T34
Concept CD
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Essential hypertension is a disease of unknown etiology. What most agree on, however, is that retention of excess sodium underlies the sustained increase in blood pressure (BP). Sodium retention can be estimated from the extent to which there is suppression of the renin-angiotensin aldosterone system (RAAS). If the BP-RAAS link is confirmed, measurements of the RAAS can be used as markers for increased risk of hypertension. In this research, we propose a joint semiparametric modeling structure for the assessment of the effects of aldosterone and renin on longitudinally measured systolic and diastolic blood pressure in a group of children. The modeling structure offers the flexibility to accommodate repeatedly measured bivariate outcome data, as well as the joint effect of two independent variables. Scientifically, by focusing on the renin-aldosterone axis, we attempt to directly quantify the effect of sodium retention on blood pressure by the kidney through distal nephron. The proposed model has the potential to be further extended for other applications in studies of biological systems.

Presenter: *Wahed, Abdus* **Affiliation:** *University of Pittsburgh*

Abstract Title: ***Statistical inference for treatment strategies from two-stage randomization designs when second randomization is delayed***

Author(s): *Abdus S. Wahed*, Jesse Y. Hsu*

Abstract:

Session: T22
Concept AB
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Two-stage randomization designs are common in treatment of complex diseases such as cancer and depression. In such designs patients are randomized to an induction treatment in the first stage and then based on their response (to the induction treatment) and consent (to maintenance treatment) status, are randomized to a maintenance treatment in the second stage. In some settings, this second randomization may be delayed due to unavoidable reasons such as patient scheduling and resolution of adverse events resulting from induction

treatment. In many cases, the ultimate goal is to compare the effects of combination of induction and maintenance treatments based on overall survival. Statistical methods proposed in the literature handle non-responders and non-consenters alike. However, patients who do not respond to the initial treatment may be prognostically different from those who respond but do not consent to further treatment. In this study, we estimate the casual effect of treatment strategies by introducing consistent and efficient estimators of survival distributions, where non-responders and non-consenters are treated separately. In addition we account for the delay in second randomization in our methods. Large-sample properties of the proposed estimators are derived and compared in a simulation study. The estimators are then applied to a leukemia clinical trial data set.

Presenter: *Wang, Lifeng* **Affiliation:** *Michigan State University*

Abstract Title: ***Boosting for nonparametric high-dimensional models***

Author(s): *Qi Yan, Lifeng Wang**

Abstract: In regression analysis, variables can often be combined into groups based on prior knowledge. Such a group structure of the predictor variables can be effectively utilized in regression analysis in order to improve identification of relevant groups of variables and to improve the prediction performance. In this paper, we propose a boosting method to perform nonparametric regression and feature selection for high-dimensional group additive models. We investigate the learning theory for the proposed boosting algorithm, and illustrate its finite sample performance via both simulated and real data.

Session: M15
Cosmopolitan D
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Wang, Sue-Jane* **Affiliation:** *FDA*

Abstract Title: ***Regulatory Perspectives on Genomic Biomarker classifier in Therapeutic Development***

Author(s): *Sue-Jane Wang*

Abstract: The predictive utility of a biomarker classifier has tremendous clinical appeal. It is envisioned that there will be a growing number of examples in which use of a companion diagnostic will need to be considered and may become an integral part for practicing medicines. The credible mechanism to test the clinical utility of a genomic classifier is crucial for choices of trial designs in late phase registration trials. In this talk, we present some current FDA thinking on biomarker classifier as a predictive biomarker in pharmacogenomics clinical trials. The study design issues and analyses approaches on the use of genomic biomarker classifier throughout a drug development program will also be discussed including use of tree-based approaches.

Session: M22
Regency A
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Wang, Sijian* **Affiliation:** *University of Wisconsin, Madison*

Abstract Title: ***Regularized REML for Estimation and Selection of Fixed and Random Effects in Linear Mixed-Effects Models***

Author(s): *Sijian Wang*, Peter XK Song, Ji Zhu*

Abstract: The linear mixed effects model (LMM) is widely used in the analysis of clustered or longitudinal data. In the practice of LMM, inference on the structure of random effects component is of great importance not only to yield proper interpretation of subject-specific effects but also to draw valid statistical conclusions. This task of inference becomes significantly challenging when a large number of fixed effects and random effects are involved in the analysis. The difficulty of variable selection arises from the need of simultaneously regularizing both mean model and covariance structures, with possible parameter constraints between the two. In this paper, we propose a novel method of regularized restricted maximum likelihood to select fixed and random effects simultaneously in the LMM. The Cholesky decomposition is invoked to ensure the positive-definiteness of the selected covariance matrix of random effects, and selected random effects are invariant with respect to the ordering of predictors appearing in the model. We develop a new algorithm that solves the related optimization problem effectively, in which the

Session: T28
Studio One
Tuesday, June 22
1:30 p.m.–3:20 p.m.

computational load turns out to be comparable with that of the Newton-Raphson algorithm for MLE or REML in the LMM. We also investigate large sample properties for the proposed estimation, including the oracle property. Both simulation studies and data analysis are included for illustration.

Presenter: *Wang, Lu* **Affiliation:** *University of Michigan*

Abstract Title: ***Semiparametric Regression with Missing Outcomes Using Weighted Kernel-Profile Estimating Equations***

Author(s): *Lu Wang*, Xihong Lin, and Andrea Rotnitzky*

Abstract: We consider semiparametric generalized partial linear regression models when the outcome is missing at random (MAR), some covariate effects are modeled parametrically and one covariate is modeled nonparametrically. We propose a class of augmented inverse probability weighted (AIPW) kernel-profile estimating equations, where the nonparametric component is estimated as solution to AIPW kernel estimating equations and the parametric regression coefficients are estimated as solutions to AIPW profile estimating equations. The AIPW kernel estimating equations require input estimates of the missingness probabilities and the conditional mean of the outcome given covariates and auxiliaries under working parametric models. We show that the AIPW estimators of the nonparametric and the parametric component are double-robust, i.e. they are consistent provided one of the working models is correct, but not necessarily both. In addition, the AIPW estimator of the parametric component is locally semiparametric efficient in the following sense. It is consistent and asymptotically normal under the semiparametric model defined by the semiparametric generalized partial linear model on the full-data, and the restrictions on the missingness mechanism that the data are MAR and follow the assumed parametric model. In addition, if the model for the outcome mean is correctly specified, its asymptotic variance achieves the semiparametric variance bound. We conduct simulations to evaluate the finite sample performance of the AIPW estimators, and apply the proposed methods to data analyzed to investigate risk factors of myocardial ischemia.

Session: M24
Concept CD
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Wang, Jane-Ling* **Affiliation:** *Univ. of California, Davis*

Abstract Title: ***Modeling Left-truncated and Right Censored Survival data with longitudinal covariates***

Author(s): *Yuru Su and Jane-Ling Wang **

Abstract: In this talk, we explore the modeling of survival data in the presence of longitudinal covariates. In particular, we consider survival data that are subject to both left truncation and right censoring. It is well known that traditional approaches, such as the partial likelihood approach for the Cox proportional hazards model encounter difficulties when longitudinal covariates are involved in the modeling of the survival data. We propose a joint likelihood approach to overcome these difficulties and establish asymptotic theory. The new approach will also be illustrated numerically and with a data.

Session: M12
Concept CD
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Wang, Jiyong* **Affiliation:** *BioPIER*

Abstract Title: ***BioPIER Clinical Workbench***

Author(s): *Simon Gao, Ji-yong Wang*

Abstract: BioPIER Clinical Workbench is a clinical data repository system, built with powerful real-time clinical data reporting and analysis. It is best used as a centralized in-house data review and validation platform for outsourced/EDC studies:

- Team-based clinical data review
- Clinical data validation
- Clinical safety and statistical analysis
- Integrated patient profiling
- Integrated study document viewing
- Clinical statistical reporting (CSR) automation and standardization

Session: M38
Studio One
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Wang, Jiantian* **Affiliation:** *Kean University*
Abstract Title: ***A proof for the underestimation of the Greenwood's type of estimation***
Author(s): *Jiantian Wang*
Abstract: We will investigate some small sample properties of the Kaplan-Meier estimator(KM) and the Nelson-Aalen estimator(NA) under the Koziol-Green Model. Explicit and elegant bias expressions of these estimators will be derived. By revealing the asymptotic structures of the variances of the KM and the NA estimators, we will prove the underestimation of the Greenwood-type variance estimators of KM and the NA.

Session: W18
Studio One
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Wang, Duolao* **Affiliation:** *London School of Hygiene & Tropical Medi*
Abstract Title: ***Comparison of Six Commonly Used QT Correction Models and Their Parameter Estimation Methods***
Author(s): *Duolao Wang*,Jorg Taubel, Radivoj Arezina*
Abstract: QT interval has become a widely used surrogate endpoint to assess drug safety in drug development. QT interval is often corrected by heart rate to QTc (corrected QT interval) so that QTc interval becomes independent of the heart rate. Although parametric QT correction models exist, there is a lack of assessment of sensitivity of model parameter estimates and performance of QTc intervals to parameter estimation methods. This study compares six commonly used QT correction models (linear, hyperbolic, parabolic, logarithmic, shifted logarithmic, and exponential models) and three available parameter estimation methods (golden section, least square regression model, and mixed model) using five indices for QTc evaluation based on real and simulated ECG datasets. The results show that the golden section approach always finds the correction factor making QTc interval independent of heart rate for all 6 formulae. However, the correction formulae derived from least square regression model and mixed model often fail to make QTc interval independent of heart rate. Parabolic correction formula with its correction parameter being estimated using the golden section approach is recommended to correct QT for heart rate in clinical studies.

Session: T13
Network
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Wang, Jing* **Affiliation:** *University of Illinois at Chicago*
Abstract Title: ***Modeling time series via spline confidence bands and block bootstrap***
Author(s): *University of illinois at Chicago*
Abstract: An efficient block-bootstrap method is applied to obtain simultaneous spline confidence band for the trend function of time series with alpha-mixing errors. Simulation study confirms that the bands have satisfactory coverage of the true trend function. The method is applied to a real data.

Session: M34
Concept CD
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Wang, Lily* **Affiliation:** *University of Georgia*
Abstract Title: ***Estimation and variable selection for generalized additive partial linear models***
Author(s): *Li Wang*, Xiang Liu, Hua Liang, Raymond J. Carroll*
Abstract: A class of generalized additive partial linear models is investigated. We propose the use of polynomial spline smoothing for estimation of nonparametric functions, and derive quasi-likelihood based estimators for the linear parameters. We establish asymptotic normality for the estimators of the parametric components. The procedure avoids solving big system of equations as in kernel-based procedures and thus results in gains in computational simplicity. We further develop a class of variable selection procedures for the linear parameters by employing a nonconcave penalized likelihood, which is shown to have an oracle property. Monte Carlo simulations and an analysis of a dataset from Pima Indian diabetes study are presented for illustration.

Session: M37
Cosmopolitan D
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Wang, Haonan* **Affiliation:** *Colorado State University*

Abstract Title: ***Dynamical Multiple-Input, Single-Output Model of Neural Spike***

Author(s): *Catherine Y. Tu, F. Jay Breidt, Dong Song, Theodore W. Berger and Haonan Wang**

Abstract: In this talk, we consider the problem of modeling neural signal transformation. A dynamic Multiple-Input, Single-Output model of neural information communication is proposed. Each input neuron and the output neuron have a functional relationship which is approximated by polynomial splines. A penalized likelihood approach is implemented for simultaneous parameter estimation and variable selection. The notion of sparsity in parameter estimation has been generalized to function estimation. Two different types of functional sparsity are of particular interest: global sparsity and local sparsity. Computation of the penalized approach is rather challenging. The one-step estimator based on the group bridge approach for maximizing the penalized likelihood is proposed. The performance of the proposed method is assessed using Monte Carlo simulation studies.

Session: M37
Cosmopolitan D
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Wathen, J. Kyle* **Affiliation:** *M. D. Anderson Cancer Center*

Abstract Title: ***Personalized Medicine - Using Biomarker Signatures to Predict Response to New Therapies***

Author(s): *J. Kyle Wathen*, Donald Berry, Neby Bekele, Nola Hylton, Anna Barker, Laura Esserman*

Abstract: The ISPY2 process is a new approach to conducting clinical research that utilizes a patient's biomarkers measurements to predict which treatment is most likely to provide benefit. Patients will be adaptively randomized and the treatment assignment probabilities will be altered to favor the treatment that, on average, appears superior for a given patient's biomarker characteristics. In contrast to the traditional phase II clinical trial, which has a fixed number of treatments, the ISPY2 process will allow new agents to enter the trial as they become available and will "graduate" treatments based on the likelihood of future success in a subset of the patient population. A simulation study is presented and examples given to demonstrate the adaptive nature of the design.

Session: M32
Regency A
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Wehrenberg, Scott* **Affiliation:** *Boston Scientific Corporation*

Abstract Title: ***Case Study of a Bayesian Clinical Study with a Hierarchical Prior in a Medical Device Trial***

Author(s): *Scott Wehrenberg*

Abstract: In the medical device industry it is common to run a clinical trial to compare a new device to an old approved device. For this purpose, a Bayesian hierarchical model is a useful method for incorporating historical data on the old device, potentially increasing the precision of parameter estimates. In this case study, I present the application of such a model to a randomized, single blind, non-inferiority trial in a regulatory setting. Sample size and operating characteristics were determined through simulations, using specialized Fortran software developed by Professor Ming-Hui Chen. Revisions to the statistical methods were made based on discussions with the FDA. A more conservative "conditional borrowing" strategy was employed, restricting the scenarios under which historical data could be borrowed. The final method resulted in a reduced sample size of 460 patients compared to the frequentist alternative. After unblinding, the criterion for borrowing historical data was not met. Analyses were carried out using both Fortran and WinBugs programming in order to confirm the results. The posterior probability of non-inferiority was 0.9996, and therefore non-inferiority of the new device to the old device was concluded.

Session: W11
Regency A
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Wei, Ying* **Affiliation:** *Columbia University*

Abstract Title: ***Analysis on Censored Quantile Residual Life Model via Spline Smoothing***

Author(s): *Yanyuan Ma and Ying Wei*

Abstract: We propose a general class of quantile residual life models, where a speci¯c quantile of the residual life time, conditional on an individual has survived up to time t, is a function of

Session: M31
Concept AB
Monday, June 21
3:40 p.m.–5:30 p.m.

certain covariates with their coe±cients varying over time. Those varying coe±cients are assumed to be smooth unspeci¯ed functions of t. We propose to estimate those coe±cient functions using spline approximation. Incorporating the spline representation directly into a set of unbiased estimating equations, we obtain a one-step estimation procedure of the coe±cient functions. We show that the estimation procedure leads to a uniformly consistent estimator. To obtain further computational simpli¯cation, we incorporate a modi¯cation and propose a two-step estimation approach, where we estimate the coe±cients on a series of time points ¯rst, and follow by spline smoothing. We compare the two methods in terms of their asymptotic e±ciency and computational complexity. We further developed inference tools to test the signi¯cance of the covariate e®ect on residual life. The ¯nite sample performance on the estimation and testing procedures are further illustrated through numerical experiments. We also applied the methods to a real data from a neurological study.

Presenter: *Wellner, Jon* **Affiliation:** *University of Washington, Statistics*
Abstract Title: ***Estimation for two - phase designs: semiparametric models and Z-theorems***
Author(s): *Jon A. Wellner*
Abstract: I will discuss problems concerning estimation in two-phase designs with covariate information missing by design. The study of several new estimators will depend on a new Z-theorem with estimated nuisance parameters and related results for empirical processes indexed by functions depending on estimated parameters. We will also make connections with the theory of generalized empirical likelihood estimators along the lines of Newey and Smith (2004).

Session: M14
Cosmopolitan C
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Wu, Jeremy* **Affiliation:** *US Census*
Abstract Title: ***Longitudinal Data Systems: New Frontier in Statistics***
Author(s): *Jeremy Wu*
Abstract: Census results appear in Chinese history as far back as the Western Han Dynasty more than two thousand years ago. Random sampling and mathematical statistics have been a central part of statistical study for the past century. The U.S. Census Bureau started the Longitudinal Employer-Household Dynamics (LEHD) program to build a longitudinal data system on jobs in 1999. LEHD is not a census, a random survey, or a set of administrative records. Instead, it is an integration of all these data sources, creating innovative data and products about the U.S. jobs market that have never been available to the public before. There are now intensive efforts to start similar longitudinal data systems on education. This talk focuses on the challenges of this new frontier in statistics.

Session: M17
Regency B
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Wu, Lang* **Affiliation:** *University of British Columbia*
Abstract Title: ***Joint modelling longitudinal and survival data with missing values and measurement errors***
Author(s): *Lang Wu*, Wei Liu, Joan Hu*
Abstract: Joint modelling longitudinal data and survival data has received great attention in recent years. We consider a nonlinear mixed effects model for the longitudinal data and a frailty model for the survival data, with shared random effects. We discuss some computational issues, including the EM algorithm and Laplacian approximations. We also consider commonly used two-step methods. Some examples are present for illustration.

Session: M12
Concept CD
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Wu, Changbao* **Affiliation:** *University of Waterloo*
Abstract Title: ***Pseudo Empirical Likelihood Inference for Multiple Frame Surveys***
Author(s): *Changbao Wu* and J.N.K. Rao*
Abstract: This paper presents a pseudo empirical likelihood approach to inference for multiple frame surveys. We establish a unified framework for point and interval estimation of finite population parameters, and show that inferences on the parameters of interest making

Session: W25

Concept CD
Wednesday, June 23
10:10 a.m.–12:00 p.m.

effective use of different types of auxiliary population information can be conveniently carried out through the constrained maximization of the pseudo empirical likelihood function. Confidence intervals are constructed using either the asymptotic chisquare distribution of an adjusted pseudo empirical likelihood ratio statistic or a bootstrap calibration method. Simulation results based on Statistics Canada's Family Expenditure Survey data showed that the proposed methods perform well in finite samples for both point and interval estimation. In particular, a multiplicity-based pseudo empirical likelihood method is proposed. It is easy to use this method for multiple frame surveys with more than two frames and it does not require complete frame membership information. The proposed pseudo empirical likelihood ratio confidence intervals have a clear advantage over the conventional normal approximation-based intervals in estimating population proportions of rare items, a scenario that often motivates the use of multiple frame surveys. All related computational problems can be handled using existing algorithms for pseudo empirical likelihood methods with single frame surveys.

Presenter: *Wu, Yujun* **Affiliation:** *Sanofi-Aventis*

Abstract Title: ***Interim Treatment Selection with a Flexible Selection Margin in Clinical Trials***

Author(s): *Yujun Wu* and Peng-liang Zhao*

Abstract:

Session: M33
Network
Monday, June 21
3:40 p.m.–5:30 p.m.

When several treatment arms are administered along with a control arm in a trial, dropping the non-promising treatments at an early stage helps to save the resources and expedite the trial.Drop-the-losers design (Sampson and Sill, 2005) and Winner design (Shun, Lan, and Soo, 2008) are such two-stage adaptive designs that select the treatment with the numerically highest mean response at the interim stage. We modified this interim treatment selection rule by introducing a flexible selection margin to judge the acceptable treatment difference. Other treatments could be selected at the interim stage along with the empirically best one if their treatment difference does not exceed this margin. Hypothesis testing procedure is developed to assess the selected treatment(s) by taking into account the interim selection process. The modified selection rule makes the design more flexible and practical.

Presenter: *Wu, Yichao* **Affiliation:** *NCSU*

Abstract Title: ***An ordinary differential equation based solution path algorithm***

Author(s): *Yichao Wu*

Abstract:

Session: T18
Cosmopolitan D
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Efron, Hastie, Johnstone and Tibshirani (2004) proposed Least Angle Regression (LAR), a solution path algorithm for the least squares regression. They pointed out that a slight modification of the LAR gives the LASSO (Tibshirani, 1996) solution path. However it is largely unknown how to extend this solution path algorithm to models beyond the least squares regression. In this work, we propose an extension of the LAR for generalized linear models and the quasi-likelihood model by showing that the corresponding solution path is piecewise given by solutions of ordinary differential equation systems. Our contribution is twofold. First, we provide a theoretical understanding on how the corresponding solution path propagates. Second, we propose an ordinary differential equation based algorithm to obtain the whole solution path.

Presenter: *Wu, Rongling* **Affiliation:** *Pennsylvania State University*

Abstract Title: ***A statistical model for mapping morphological shapes***

Author(s): *Guifang Fu, Kiranmoy Das, Jiahan Li, Runze Li, and Rongling Wu**

Abstract:

Session: W24
Network
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Living things come in all shapes and sizes, from bacteria, plants, and animals to humans. Knowledge about the genetic mechanisms for biological shape has been far-reaching implications for a range spectrum of scientific disciplines including anthropology, agronomy, forestry, ecology, developmental biology, evolution and biomedicine. We derived a statistical model for mapping specific genes or quantitative trait loci (QTLs) that control morphological shape. The model is founded on recent developments of DNA-based

molecular markers for various biological systems, shape correspondence analysis, and advanced statistical algorithms derived to analyze and interpret observational data of high complexity. The EM algorithm was implemented to estimate the biological shape of different genotypes at a shape QTL. We formulate a series of quantitative testable hypotheses about the interplay between genetic actions and the origin, properties, and functions of morphological shape. Computer simulation was used to investigate the statistical property of the model. By identifying specific QTLs for morphological shape, the model developed will help to ask, disseminate and address many major integrative biological and genetic questions and challenges in the genetic control of biological shape and function.

Presenter: *Xie, Jun* **Affiliation:** *Purdue University*

Abstract Title: ***Large-Scale Multinomial Inference and Its Applications in Genome-Wide Association Studies***

Author(s): *Chuanhai Liu and Jun Xie**

Abstract: Despite progresses in statistical analyses of SNP data, many statistical challenges in these data sets are unsolved. SNPs are single base differences in DNA sequence among individuals. The data type is categorical, with three possible genotypes in a single SNP. When we consider a block of 10 SNPs, there are about 60,000 categories, much larger than sample size of typical studies. In addition, multiple testing for simultaneous studies of hundreds and thousands of SNPs is another difficult issue. I will present some preliminary analysis for large scale SNP data and introduce a new concept of hypothesis testing motivated by Dempster-Shafer theory for inference. We introduce auxiliary data generating models and infer unknown multinomial cell probabilities by reasoning through unobserved realizations of auxiliary random variables. We apply the method to identify SNPs that are potentialy associated with a given disease.

Session: T26
Cosmopolitan C
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Xie, Dawei* **Affiliation:** *University of Pennsylvania*

Abstract Title: ***Small area estimation by combining two surveys with special consideration of cell only households***

Author(s): *Dawei Xie*, Van Parsons, Nat Schenker, Trivellore Raghunathan, Michell Town, Benmei Liu, Eric Feuer*

Abstract: Cancer surveillance research requires estimates of the prevalence of cancer risk factors and screening for small areas such as states and counties. To obtain such estimates, Raghunathan et al (2007) has developed an approach to combine two national surveys, i.e., the BRFSS, a large telephone survey, and the NHIS, a face-to face interview survey with smaller sample size. In this approach, the population in each small area was divided into two strata, households with landline telephones and households without. Estimates from the years of 1997-2000 were obtained for the two sub-populations from a Bayesian hierarchical model. The potential noncoverage and noresponse biases in the BRFSS were took into account. When applying this approach to newer data in the year of 2004-2006, we had to consider another stratum of the population which is consisted of the rapid-growing cell only households. We extended the above approach to obtain estimates for three sub-populations for each small area and weighted them by the proportions in each sub-population. A difficulty in this extension is that there are no reliable estimates for these proportions, i.e., the proportion of cell only households, landline households and no landline or cell phone households for the small areas. Therefore we had to estimate them first. Estimate from several approaches were compared. The variances of these estimates were incorporated into the Markov Chain Monte Carlo method which was used to obtain the final estimates of prevalence of cancer risk factors and screening for small areas.

Session: T16
Studio One
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Xu, Ronghui* **Affiliation:** *University of California, San Diego*
Abstract Title: ***Mixed-effects model selection using conditional AIC***
Author(s): *Donohue, Xu*, Vaida, Haut*
Abstract: We consider model selection for longitudinal data, when the focus is on the prediction of the random effects. The conditional Akaike information (cAI) can be used for this purpose, defined as the expected predictive log-likelihood of the observed data given the random effects. Under normal linear mixed models, the cAI can be directly estimated (even in finite samples) by a conditional AIC (cAIC). Outside of this setting we show that it is helpful to view the estimation of the fixed and the random effects as via an estimating function, obtained through a penalized likelihood. An asymptotic expansion of the log penalized likelihood leads to an approximately unbiased estimate of the cAI, i.e. the cAIC. Simulation studies will be presented and the method will be illustrated on a longitudinal skin cancer data set. Time permitting we will also discuss the handling of nuisance parameters, and a link through the generalized linear mixed model to the proportional hazards mixed model.

Session: M31
Concept AB
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Xu, Jane* **Affiliation:** *Sepracor*
Abstract Title: ***Hommel-based gatekeeping multiple comparison procedure: an example***
Author(s): *Jane Xu*; Thomas Brechenmacher*
Abstract: In recent years increasingly more and more clinical trials are conducted where the objectives are hierarchically ordered. In such cases it is important that appropriate multiple comparison adjustment methods are applied to ensure the family-wise error rate (FWER) is controlled in the strong sense. In this talk, a motivating example will be given to illustrate a Hommel-based gatekeeping multiple comparison procedure that controls the FWER. This method is derived from the work done by Dmitrienko, and Tamhane, et. al. [2010]. Details of the procedure will be described and how it is applied to the motivating clinical trial will be illustrated. The results from this procedure will be compared with results using the Bonferroni-based tree gate-keeping procedure described in Dmitrienko, et. al. [2007].

Session: W12
Regency B
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Xue, Lan* **Affiliation:** *Oregon State University*
Abstract Title: ***Consistent Model Selection for Marginal Generalized Additive Model for Correlated Data***
Author(s): *Lan Xue*, Annie Qu, Jianhui Zhou*
Abstract: We consider the generalized additive model when responses from the same cluster are correlated. Incorporating correlation in the estimation of nonparametric components for the generalized additive model is important since it improves estimation efficiency and increases statistical power for model selection. In our setting, there is no specified likelihood function for the generalized additive model since the outcomes could be non-normal and discrete, which makes estimation and model selection very challenging problems. We propose consistent estimation and model selection which incorporate the correlation structure. We establish an asymptotic property with L_{2}-norm consistency for the nonparametric components, which achieves the optimal rate of convergence. In addition, the proposed model selection strategy is able to select the correct generalized additive model consistently. That is, with probability approaching to 1, the estimators for the zero function components converge to 0 almost surely. We will illustrate our method using numerical studies with both continuous and binary responses, and a real data application of binary periodontal data.

Session: M37
Cosmopolitan D
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Xue, Hongqi* **Affiliation:** *University of Rochester*
Abstract Title: ***Sieve Estimation in Nonlinear Ordinary Differential Equation Models***
Author(s): *Hongqi Xue*, Hongyu Miao and Hulin Wu*
Abstract: We consider estimation of constant and time-varying coefficients in nonlinear ordinary differential equation (ODE) models where analytic closed-form solutions are not available. The numerical solution-based nonlinear least squares (NLS) estimator is proposed. A numerical algorithm such as the Runge-Kutta algorithm is used to approximate the ODE

Session: W13
Concept AB

Wednesday, June 23
8:00 a.m.–9:50 a.m.

solution. The asymptotic properties are established for the proposed estimators with consideration of both numerical error and measurement error. The B-spline approach is used to approximate the time-varying coefficients and the corresponding asymptotic theories in this case are investigated under the framework of the sieve approach. Our results show that if the maximum step size of the numerical algorithm goes to zero faster than a special rate, then the numerical error is negligible compared to the measurement error. This provides a theoretical guidance in selection of the step size for numerical evaluations of ODEs. Moreover, we have shown that the numerical solution-based NLS estimator and the sieve NLS estimator are strongly consistent. The sieve estimator of constant parameters is asymptotically normal with the same asymptotic co-variance as that of the case where the true ODE solution is exactly known, while the estimator of the time-varying parameter has the optimal convergence rate under some regularity conditions. We illustrate our approach with both simulation studies and clinical data on HIV viral dynamics.

Presenter: *Yang, Yuhong* **Affiliation:** *University of Minnesota*

Abstract Title: ***Parametricness, Model Identifiability and Adaptation***

Author(s): *Wei Liu and Yuhong Yang* *

Abstract:

Parametric and nonparametric models are convenient mathematical tools to describe characteristics of data with different degrees of simplification. When a model is to be selected from a number of candidates, not surprisingly, differences occur when the data generating process is assumed to be parametric or nonparametric. In this talk, in a regression context, we will consider the question if and how we can distinguish between parametric and nonparametric situations and discuss feasibility of adaptive estimation to handle both parametric and nonparametric scenarios optimally.

Session: M37
Cosmopolitan D
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Yang, Jie* **Affiliation:** *University of Illinois at Chicago*

Abstract Title: ***Optimal Design for Two-Level Factorial Experiments with Binary Response***

Author(s): *Abhyuday Mandal, Jie Yang*, and Dibyen Majumdar*

Abstract:

We consider the problem of obtaining locally D-optimal designs for factorial experiments with qualitative factors at two levels each with binary response. For the 2^2 factorial experiment with main effects model we obtain optimal designs analytically in special cases and demonstrate how to obtain a solution in the general case using cylindrical algebraic decomposition. The optimal designs are shown to be robust to the choice of the assumed values of the prior and when there is no basis to make an informed choice of the assumed values we recommend the use of the uniform design, i.e., the design that assigns equal number of observations to each of the four points. For the general 2^k case we show that the uniform design has a maximin property.

Session: W14
Concept CD
Wednesday, June 23
8:00 a.m.–9:50 a.m.

Presenter: *Yi, Grace* **Affiliation:** *University of Waterloo*

Abstract Title: ***A Pairwise Likelihood Method for Clustered Binary Data under Generalized Partially Linear Single-Index Models***

Author(s): *Wenqing He, Grace Yi*

Abstract:

Clustered data arise commonly in practice and it is often of interest to estimate the mean parameters as well as the association parameters. However,much research has been directed to inference about the mean parameters with the association parameters treated as nuisance. There is little work concerning both the marginal and association structures, especially in the semiparametric framework. In this talk, I will discuss a semiparametric method based on the pairwise likelihood formulation that handles inference on mean and association parameters simultaneously. Generalized partially linear single-index models are employed to allow for the flexibility of modeling the marginal mean structure.

Session: M31
Concept AB
Monday, June 21
3:40 p.m.–5:30 p.m.

Presenter: *Yin, Xiangrong* **Affiliation:** *University of Georgia*
Abstract Title: ***Sufficient dimension reduction based on an ensemble of minimum average variance estimators***
Author(s): *Xiangrong Yin *; Bing Li*
Abstract: A class of sufficient dimension reduction method family via forward approach is developed. This general class includes several well-known methods such as minimum average variance estimate (MAVE; Xia, et al. 2002), density MAVE (dMAVE; Xia, 2007) and sliced regression (SR; Wang and Xia, 2008) as special cases. In particular, among the family we suggest a new method using characteristic function. Comparing with inverse dimension reduction methods the new method doesn't require linearity conditions, while inherits the advantage of SR with robustness against extreme values as well as exhaustiveness of the recovery of the central subspace. Contrast with SR where slicing is used, the new method repeatedly uses the data, and thus it is much more robust in the tuning parameter selection. In addition, case of multivariate responses is naturally incorporated in the new method. Simulation study shows that our method is the best among methods that recover the central subspace.

Session: T28
Studio One
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Yu, Menggang* **Affiliation:** *Indiana University*
Abstract Title: ***Predicting Drug-Drug Interaction***
Author(s): *Menggang Yu*
Abstract: A meta-analysis from multiple data resources is critical to obtaining a comprehensive understanding of a drug's pharmacokinetics (PK). Often time and subject specific drug concentration data from clinical PK studies on inhibitor/inducer or substrate's PK are not directly available, though there always exist rich published PK information in the literature. In this talk, an innovative Bayesian meta-analytic model framework is presented to analyze multiple published sample mean data sets for predicting drug-drug interaction (DDI). An MCMC PK parameter estimation procedure is developed, and DDI prediction is conducted based on the PK models of two drugs and posterior distributions of the PK parameters. The performance of Bayesian meta-analysis in DDI research is demonstrated through a ketoconazole-midazolam example.

Session: T23
Regency B
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Yu, Lili* **Affiliation:** *Georgia Southern University*
Abstract Title: ***Extended quasi-likelihood in the frame of AFT model***
Author(s): *Lili Yu**
Abstract: Yu et al has proposed a simple and efficient analysis method in the generalized linear model for right censored data by using Quasi-likelihood. However, they assumed known mean and variance relationship. In this paper, we relax this assumption by estimating the unknown variance function parametrically using fractional polynomials. Simulation shows that the new method is doing well. The new method is illustrated by two real datasets.

Session: M23
Concept AB
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Yu, Kai* **Affiliation:** *National Cancer Institute*
Abstract Title: ***Efficient P-value Evaluation for Resampling-based Tests***
Author(s): *Kai Yu*, Faming Liang*
Abstract: The resampling-based test, which is based on permutation or bootstrap procedures, has been widely used for statistical hypothesis testing when the asymptotic distribution of the test statistic is unavailable or unreliable. It requires repeated calculations of the test statistic on a large number of simulated datasets for its significance level assessment, and thus it is very computationally intensive. In this paper, we propose an efficient p-value evaluation procedure that utilizes the recently developed stochastic approximation Markov chain Monte Carlo algorithm. The new procedure can be used easily for estimating the p-value for any resampling-based test. We demonstrate through numeric examples that the proposed

Session: M36
Cosmopolitan C
Monday, June 21
3:40 p.m.–5:30 p.m.

procedure can achieve 100 to 1000 times as high an efficiency as the standard resampling-based procedure when evaluating a test statistic with a small p-value (e.g., less than). With its computational burden reduced by this proposed procedure, the versatile resampling-based test would become computationally feasible for a much wider range of applications.

Presenter: *Yuan, Ming* **Affiliation:** *Georgia Tech*

Abstract Title: ***Sparse Regularization for High Dimensional Additive Models***

Author(s): *Ming Yuan*

Abstract: Motivated by practical demands, sparse regularization for high dimensional additive models has attracted considerable attention in recent years. Although a number of approaches have been introduced, it still remains unclear whether or not these methods can take full advantage of sparsity often associated with this type of problem, and more fundamentally to what extent sparsity can determine the difficulty of high dimensional additive model estimation. To gain insights on these questions, we study the behavior of the l1 type of regularization under a sequence model. Our results suggest remarkable similarities and differences between linear regression and additive models in high dimensional settings. In particular, our analysis indicates that, unlike in linear regression, l1 regularization does not yield optimal estimation for additive models of high dimensionality. This surprising observation prompts us to introduce a new regularization technique that can be shown to be optimal in the minimax sense.

Session: T27
Cosmopolitan D
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Yuan, Ying* **Affiliation:** *Univ. of Texas MD Anderson Cancer Center*

Abstract Title: ***Continual Reassessment Method for Late-Onset Toxicities Using Data Augmentation***

Author(s): *Ying Yuan and Guosheng Yin*

Abstract: The continual reassessment method (CRM) is a commonly used dose-finding design for phase I clinical trials. In practice, it requires that the toxicity outcome be observed shortly after the initiation of the treatment, which, however, may not be the case for late-onset toxicities. To accommodate the setting in which toxicities may occur long after the treatment, we propose the late-onset toxicity CRM (LOT-CRM). We naturally treat the unobserved toxicities as missing data, and show that such missing data are nonignorable in the sense that the missingness depends on the unobserved outcomes. We address the missing toxicity data using the data augmentation approach and thus we do not need to interrupt patient accrual to wait for the full observation of the outcomes of patients already in the trial. We evaluate the operating characteristics of the LOT-CRMs through extensive simulation studies. We show that the proposed designs satisfactorily resolve the issues related to late-onset toxicities.

Session: M11
Concept AB
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Zeng, Peng* **Affiliation:** *Auburn University*

Abstract Title: ***SIM-Lasso for estimation and variable selection in single-index models***

Author(s): *Peng Zeng *, Tianhong He, Yu Zhu*

Abstract: The single index model is a natural extension of the linear regression model for applications where linearity between the response variable and the predictor variables may not hold. In this talk, we propose a penalized local linear smoothing method called sim-lasso for estimation and variable selection in the single index model. The sim-lasso method incorporates an L1 penalty of the derivative of the link function into the loss function of local linear smoothing and can be considered as an extension of the usual lasso to the single index model. We develop several algorithms to calculate the sim-lasso estimates and solution paths. The properties of the solution paths are investigated. In simulation study and a real data application, sim-lasso demonstrates excellent performance.

Session: W28
Cosmopolitan D
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Zeng, Donglin* **Affiliation:** *University of North Carolina*

Abstract Title: ***Semiparametric Inference for Partly Single-Index Hazards Model***

Author(s): *Kai Ding, Donglin Zeng*, Michael Kosorok*

Abstract: We propose a partly single-index hazards model for modelling censored data. The covaraites

Session: M24
Concept CD
Monday, June 21
1:30 p.m.–3:20 p.m.

of interest enter the model parametrically while confounding covariates are modelled via a single index. We show that the usual local likelihood method gives bias estimation. We further propose a corrected approach which yields consistent and efficient estimators. The proposed method is demonstrated by simulation studies and real data application.

Presenter: *Zhang, Lingsong* **Affiliation:** *Purdue University*
Abstract Title: ***Statistical Methods for Copy Number Alternation***
Author(s): *Lingsong Zhang*
Abstract: We will compare several methods for detection of copy number alternation. A new method based on smoothing techniques will be proposed. Connections and differences between the new method and existing methods will be explored.

Session: T35
Regency B
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Zhang, Hao* **Affiliation:** *North Carolina State University*
Abstract Title: ***FIRST: Combining forward iterative selection and shrinkage in high dimensional linear models***
Author(s): *Hao Helen Zhang**
Abstract: We propose a new class of variable selection techniques for linear regression in high dimensional models based on a forward selection version of shrinkage methods, to be called the forward iterative regression and shrinkage technique (FIRST). The FIRST naturally combines traditional search-based methods with modern penalization approaches in one unified framework and inherits advantages from both. One main attractive property of the FIRST is its computational efficiency, which employs the explicit form of the solution in the univariate case and updates the regression parameters iteratively until convergence. By carefully considering the relationship between estimators at successive stages, we develop fast algorithms to compute the estimators. The performance of FIRST is illustrated for large-scale high dimensional sparse models.

Session: M27
Regency B
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Zhang, Dabao* **Affiliation:** *Purdue University*
Abstract Title: ***Penalized orthogonal-components regression for large p small n data***
Author(s): *Dabao Zhang*, Yanzhu Lin, Min Zhang*
Abstract: Here we propose a penalized orthogonal-components regression (POCRE) for large p small n data. Orthogonal components are sequentially constructed to maximize, upon standardization, their correlation to the response residuals. A new penalization framework, implemented via empirical Bayes thresholding, is presented to effectively identify sparse predictors of each component. POCRE is computationally efficient owing to its sequential construction of leading sparse principal components. In addition, such construction offers other properties such as grouping highly correlated predictors and allowing for collinear or nearly collinear predictors. With multivariate responses, POCRE can construct common components and thus build up latent-variable models for large p small n data.

Session: T27
Cosmopolitan D
Tuesday, June 22
1:30 p.m.–3:20 p.m.

Presenter: *Zhang, Guangyu* **Affiliation:** *University of Maryland, College Park*
Abstract Title: ***Indentifying implausible gestational ages in preterm babies with Bayesian mixture models***
Author(s): *Guangyu Zhang*, Nathaniel Schenker, Jennifer D. Parker, Dan Liao*
Abstract: Body weight and gestational ages are two important variables in obstetrics research and clinical practice. Weight of a newborn can be measured accurately; however, gestational ages of US birth data are calculated based on a mother's recall of the last menstrual period, which has been shown to introduce random or systematic errors and consequentially lower the data quality. In order to mitigate some of those errors, we use Bayesian mixture models to identify implausible gestational ages of preterm babies. First, we develop and compare

Session: T16
Studio One
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Bayesian versions of two mixture models proposed by Tentoni et al (2004) and Platt et al (2001). Then, we extend these methods by using reversible Jump MCMC. We conduct simulation studies and apply our method to the 2002 US birth data. Our research results provide useful statistical tools to correct the misspecification problem and improve the data quality.

Presenter: *Zhang, Min* **Affiliation:** *Purdue University*

Abstract Title: ***An Efficient Approach for Genome-Wide SNP Selection***

Author(s): *Min Zhang*, Vitara Pungpapong, William Muir, Dabao Zhang*

Abstract: With the recent advances in high-throughput genotyping technology, a large number of single nucleotide polymorphisms are genotyped in many subjects. Therefore, genome-wide selection is gaining increasing popularity in both plant and animal breeding. A series of statistical methods have been proposed to estimate the genome-wide breeding values on the basis of the marker genotype information, however, these methods are either computation intensive or with low prediction accuracy. We proposed a penalized orthogonal components regression approach that significantly reduces the computing time while attains similar or even better prediction accuracy. The method is evaluated by comparing to existing approaches via simulation studies.

Session: M18
Studio One
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Zhang, Xiaoxi* **Affiliation:** *Pfizer Inc.*

Abstract Title: ***Longitudinal Image Analysis of Tumor/Healthy Brain Change in Contrast Uptake Induced by Radiation***

Author(s): *Xiaoxi Zhang*, Timothy D Johnson, Roderick J A Little, Yue Cao*

Abstract: This work is motivated by a quantitative Magnetic Resonance Imaging study of the differential tumor/healthy tissue change in contrast uptake induced by radiation. The goal is to determine the time in which there is maximal contrast uptake (a surrogate for permeability) in the tumor relative to healthy tissue. A notable feature of the data is its spatial heterogeneity. Zhang, Johnson, Little, and Cao (2008 and 2010) discuss two parallel approaches to "denoise" a single image of change in contrast uptake from baseline to one follow-up visit of interest. In this work we extend the image model to explore the longitudinal profile of the tumor/healthy tissue contrast uptake in multiple images over time. We fit a two-stage model. First, we propose a longitudinal image model for each subject. This model simultaneously accounts for the spatial and temporal correlation and denoises the observed images by borrowing strength both across neighboring pixels and over time. We propose to use the Mann-Whitney U statistic to summarize the tumor contrast uptake relative to healthy tissue. In the second stage, we fit a population model to the U statistic and estimate when it achieves its maximum. Our initial findings suggest that the maximal contrast uptake of the tumor core relative to healthy tissue peaks around three weeks after initiation of radiotherapy, though this warrants further investigation.

Session: M25
Studio One
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Zhang, Jianchun* **Affiliation:** *Purdue University*

Abstract Title: ***A Weak Belief Approach to One Sample Inference***

Author(s): *Jianchun *Zhang Chuanhai *Liu*

Abstract: The lack of several desirable properties has hindered Fisher's fiducial argument and its successor, Dempster-Shafer (DS) inference from entering the mainstream statistical community. As a variation of DS inference, a new statistical inferential approach, called weak belief (WB) approach, has been proposed recently by the authors. The WB approach is aiming to provide direct probabilistic inference with desired frequency properties. An essential task of WB inference is to propose efficient predictive random set (PRS) in light of the targeted inferential problems. In this talk, we propose a new efficient PRS for one sample inferential problem. Our proposed PRS is intuitively easier to interpret and computationally simpler compared with the existing ones. The resulting PRS is applied to the well-known one

Session: T26
Cosmopolitan C
Tuesday, June 22
1:30 p.m.–3:20 p.m.

sample goodness-of-fit test problem. Simulation study shows that our method is more powerful compared with the classical tests such as Kolmogorov-Smirnov test, Anderson-Darling test and Cramer-von Mises test and as good as those proposed recently based on likelihood ratio by Zhang (2002).

Presenter: *Zhang, Heping* **Affiliation:** *Yale University*

Abstract Title: ***Decision Trees for Identifying Predictors of Treatment Effectiveness in Clinical Trials***

Author(s): *Heping Zhang*

Abstract: Double-blind, randomized clinical trials are the preferred approach to demonstrate the effectiveness of one treatment against another. The comparison is, however, made on the average group effects. While patients and clinicians have always struggled to understand why patients respond differently to the same treatment, and while much hope has been held out for the nascent field of predictive biomarkers (e.g., genetic markers), there is still much utility in exploring whether it is possible to identify personal characteristics that distinguish treatment effectiveness. To this end, we developed a node-splitting rule to build decision tree models that reflected within-node and within-treatment responses. We illustrated our method with data from clinical trials.

Session: M22
Regency A
Monday, June 21
1:30 p.m.–3:20 p.m.

Presenter: *Zhao, Hongyu* **Affiliation:** *Yale University*

Abstract Title: ***Risk modeling in personalized medicine***

Author(s): *Jia Kang, Judy Cho, Hongyu Zhao (*)*

Abstract: Recent genome wide association studies have identified many genetic variants affecting complex human diseases. It is of great interest to build disease risk prediction models based on these data. In this presentation, we will present the statistical challenges in using genome wide association data for risk predictions, and discuss different methods through both simulation studies and applications to real-world data.

Session: W24
Network
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Zhao, Hongyu* **Affiliation:** *Yale University*

Abstract Title: ***Bayesian Analysis of Proteomics Data with Nonrandom Missingness***

Author(s): *Ruiyan Luo, Christopher Colangelo, William Sessa, Hongyu Zhao*

Abstract: There are multiple proteomics platforms to identify proteins with different expression levels with iTRAQ (isobaric Tags for Relative and Absolute Quantitation) as one commonly used technique that allows simultaneous quantitation of proteins in multiple samples. In this talk, we discuss a Bayesian hierarchical model-based method to infer the relative protein expression levels and hence to identify differentially expressed proteins from iTRAQ data. Our model assumes that the measured peptide intensities are affected by both protein expression levels and peptide specific effects. The values of these two effects across experiments are modeled as random effects. The nonrandom missingness of peptide data is modeled with a logistic regression which relates the missingness probability for a peptide with the expression level of the protein that produces this peptide.We propose a Markov chain Monte Carlo method for the inference of model parameters, including the relative expression levels across samples. Our simulation results suggest that the estimates of relative protein expression levels based on the MCMC samples have smaller bias than those estimated from ANOVA models or fold changes. We apply our method to an iTRAQ dataset studying the roles of Caveolae for postnatal cardiovascular function.

Session: M16
Regency A
Monday, June 21
10:20 a.m.–12:10 p.m.

Presenter: *Zhao, Sihai* **Affiliation:** *Harvard Univ. Dept. of Biostatistics*

Abstract Title: ***Principled sure independence screening for Cox models with ultra-high-dimensional covariates***

Author(s): *Sihai Dave Zhao*, Yi Li*

Abstract: It is rather challenging for current variable selectors to handle situations where the number of covariates under consideration is ultra-high. Consider a motivating study of clinical trials of

Session: W21
Regenecy A
Wednesday, June 23
10:10 a.m.–12:00 p.m.

bortezomib for the treatment of multiple myeloma, where overall survival and expression levels of 44760 probesets, encompassing more than 22000 genes, were measured for each of 188 patients with the goal of identifying genes that predict survival after treatment. This dataset defies analysis even with regularized regression. Some remedies have been proposed for the linear model and for generalized linear models, but there are few solutions in the survival setting and, to our knowledge, no theoretical support. Furthermore, existing strategies often involve tuning parameters that are difficult to interpret. In this paper we propose and theoretically justify a principled method for reducing dimensionality in the analysis of censored data by selecting only the important covariates. We describe a new, necessary condition on the dependence between the covariates and the censoring distribution. Our procedure involves a tuning parameter that has a simple interpretation as the desired false positive rate. We present simulation results and apply the proposed procedure to analyze the aforementioned myeloma study.

Presenter: *Zheng, Hanzhe* **Affiliation:** *Merck.& Co. Inc*

Abstract Title: ***Introduction of a New Generalized Similarity Index And Its Distribution***

Author(s): *Hanzhe Zheng*

Abstract:

Session: T33
Network
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Although many similarity indices have been proposed, they are designed for the measure of the resemblance between two samples. We introduce a new generalized similarity index that can describe the degree of resemblance among samples when the number of samples is more than 2 and it will reduce to the usual similarity index when the number of samples becomes 2. Also based on the random allocation theories and large sample theories, we investigate the distribution of this new index in the large sample situation.

Presenter: *Zheng, Tian* **Affiliation:** *Columbia University*

Abstract Title: ***Studying Co-regulation and Inter-regulation of Genes via eQTL Mapping***

Author(s): *Tian Zheng**

Abstract:

Session: T15
Regency A
Tuesday, June 22
10:20 a.m.–12:10 p.m.

eQTL mapping is to find loci on human genome that have demonstrated linkage to or association with the expression of a gene in microarray hybridization experiments. Such identified loci may contain important information on the regulatory factors of the given gene under study. In this talk, I will discuss co-regulation and inter-regulation patterns identified via similar strategies.

Presenter: *Zhou, Qing* **Affiliation:** *UCLA Department of Statistics*

Abstract Title: ***Decomposition of Multimodal Distributions via Multi-Domain Sampling***

Author(s): *Qing Zhou **

Abstract:

Session: M36
Cosmopolitan C
Monday, June 21
3:40 p.m.–5:30 p.m.

When a posterior distribution has multiple modes, the posterior mean does not offer an informative summary of the distribution. Motivated by this problem, we propose to decompose the sample space of a multimodal distribution into domains of attraction of local modes. Domain-based representations are defined to summarize the probability masses of and conditional expectations on domains of attraction, which are much more informative than the mean and other overall expectations. A computational method, the multi-domain sampler, is developed to sample from and construct domain-based representations for a multimodal distribution simultaneously. To achieve efficient sampling from the domains of major modes, the multi-domain sampler utilizes the iterative weighting scheme in the Wang-Landau algorithm, incorporates gradient descent optimization, and employs adaptive global moves constructed from online estimation of domain-based representations. Applications to examples in Euclidean spaces and structural inference of Bayesian networks demonstrate the effectiveness of the multi-domain sampler in statistical estimation and construction of domain-based representations.

Presenter: *Zhou, Harrison* **Affiliation:** *Yale University*
Abstract Title: ***FUNCTIONAL REGRESSION FOR GENERAL EXPONENTIAL FAMILIES***
Author(s): *Wei Dou, David Polard, Harrison Zhou**
Abstract: Abstract: We derive a minimax lower bound for rates of convergence for an infinite-dimensional parameter in an exponential family model. An estimator that achieves the optimal rate is constructed by maximum likelihood on finite-dimensional approximations with parameter dimension that grows with sample size.

Session: T14
Cosmopolitan C
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Zhou, Lan* **Affiliation:** *Texas A&M University*
Abstract Title: ***Reduced rank models for spatially correlated functional data***
Author(s): *Lan Zhou*
Abstract: We present a new method to analyze data from an experiment using rodent models to investigate the role of p27, an important cell cycle mediator, in early colon carcinogenesis. The responses modeled here are essentially functions nested within a two-stage hierarchy. Standard functional data analysis literature focuses on a single stage of hierarchy and conditionally independent functions with near white noise. However, in our experiment, there is substantial biological motivation for the existence of spatial correlation among the functions, which arises from the locations of biological structures called colonic crypts: this possible functional correlation is a phenomenon we term crypt signaling. Thus, as a point of general methodology, we require an analysis that allows for functions to be correlated at the deepest level of the hierarchy. We developed a reduced rank functional mixed effects model and uses splines to model functions. Our methodology uses two sets of functional principal components for dimension reduction to effectively overcomes the difficulty in modeling the covariance kernel of a random function and the difficulty in modeling the correlation between functions. Analysis of this data set gives new insights into the structure of p27 expression in early colon carcinogenesis and suggests the existence of significant crypt signaling. This is based on the joint work with Jianhua Huang, Josue Martinez, Arnab Maity, Veerabhadran Baladandayuthapani and Raymond Carroll.

Session: T14
Cosmopolitan C
Tuesday, June 22
10:20 a.m.–12:10 p.m.

Presenter: *Zhu, Jun* **Affiliation:** *Colorado State University*
Abstract Title: ***On Variable Selection in Spatial Linear Regression***
Author(s): *Tingjin Chu, Haonan Wang, and Jun Zhu**
Abstract: Penalized methods are considered here for simultaneous variable selection and parameter estimation in spatial linear regression. I discuss asymptotic properties of parameter estimation and computationally feasible algorithms including a resampling method. Finite-sample properties of the proposed methods are investigated by simulation study. The methods are further illustrated by an ecological dataset.

Session: T36
Cosmopolitan C
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Zhu, Zhengyuan* **Affiliation:** *Iowa State University*
Abstract Title: ***Statistical inference under spatial preferential sampling***
Author(s): *Zhengyuan Zhu*
Abstract: Classical geostatistical methods implicitly assumes that the process of selecting the spatial sampling locations is independent of the spatial process observed at those locations. However, this assumption may not hold in applications such as mineral exploration and pollution monitoring, where the sampling rates depend on the spatial process being observed. Conventional inference methods can lead to deceiving results under such sampling designs which are referred to as preferential sampling. We propose a likelihood based approach which model the sampling process and the spatial process jointly. Efficient algorithms are developed to fit the model.

Session: T36
Cosmopolitan C
Tuesday, June 22
3:40 p.m.–5:30 p.m.

Presenter: *Zhu, Yanni* **Affiliation:** *Eli Lilly and Company*

Abstract Title: ***Tailoring Therapies for Complex Diseases based on Multiple Genotypic Markers***

Author(s): *Yanni Zhu *, Lei Shen*

Abstract: The focus of tailored therapy is to identify the right patients who will benefit from a treatment. The concept of tailoring is changing the traditional "one size fits all" approach in drug development; and the need to tailor treatments to the patient is increasingly being recognized. The central problem in developing tailored therapies is how to characterize and demonstrate differential efficacy and safety responses to drugs by patients. A common approach is to identify patient sub-populations defined by genetic markers, particularly single-nucleotide polymorphisms (SNP). Historically, much of pharmacogenomic analyses have been restricted to single marker exploration. For complex diseases such as diabetes and cardiovascular diseases, tailored therapies critically depend on the identification and integration of a number of genetic markers, each of which may only have a small effect. However, individual clinical studies often do not have sufficient sample size to enable sufficient power to identify these small effects, and it becomes even more difficult to effectively construct a multi-marker classifier to define the best patient subpopulation. In this talk I will present challenges in real world applications and discuss a number of approaches to identify a multi-marker signature with clinically meaningful combined effect, which may take the form of gene-gene interaction, haplotype, and genetic network. A number of open problems can be fertile ground for academia-industry collaborations.

Session: W24
Network
Wednesday, June 23
10:10 a.m.–12:00 p.m.

Presenter: *Zou, Hui* **Affiliation:** *University of Minnesota*

Abstract Title: ***A unified algorithm for computing ultra high dimensional penalized models***

Author(s): *Yi Yang, Hui Zou**

Abstract: With the advent of modern technology high dimensional data have frequently appeared in diverse fields such as genomics, drug discovery, signal processing, finance, and so on. Penalized methods have become increasingly popular for variable selection with high dimensions. In this talk we present a new simple yet powerful algorithm that computes the solution paths of a wide class of penalized models of the form "Loss+Penalty". The algorithm covers least squares loss, logistic regression loss, partial likelihood, Huber loss and other losses. The penalty functions can be either L1 or concave.

Session: M27
Regency B
Monday, June 21
1:30 p.m.–3:20 p.m.

PARTICIPANTS INDEX

Name	Affiliation	E-mail	Sessions
Abowd, M. John	Cornell University	john.abowd@cornell.edu	M17
Alosh, Mohamed	Food and Drug Administration	mohamed.alosh@fda.hhs.gov	W12
An, Lingling	University of Arizona	anling@email.arizona.edu	M35
Austin, Peter	Institute for Clinical Evaluative Sciences	peter.austin@ices.on.ca	Short course #3,M28
Banerjee, Moulinath	University of Michigan, Ann Arbor	moulib@umich.edu	M24
Banks, David	Duke University	banks@stat.duke.edu	Short course #2
Blocker, W. Alexander	Harvard Statistics Department	ablocker@fas.harvard.edu	M17
Bock, E. Mary	Purdue University Dept. of Statistics	mbock@purdue.edu	
Buckman, ShaAvhree	FDA	ShaAvhree.Buckman@fda.hhs.gov	K01
Bunea, Florentina	Florida State University	flori@stat.fsu.edu	M24
Campbell, Gregory	U.S. Food and Drug Administration	greg.campbell@fda.hhs.gov	K02
Carroll, J. Raymond	Texas A&M University	joyce@stat.tamu.edu	
Carvalho, E. Luis	Boston University	lecarval@math.bu.edu	T37
Case, G. Michael	Eli Lilly and Company	mgcase@lilly.com	
Chao, C. Edward	Data Numerica Institute	ez.ap@163.com	T33
Charnigo, Richard	University of Kentucky	jst227@email.uky.edu	
Chen, Adam	Georgiasouthern University	cc01450@georgiasouthern.edu	
Chen, Din	Georgia Southern University	dchen@georgiasouthern.edu	
Chen, Jiahua	University of British Columbia	jhchen@stat.ubc.ca	W25
Chen, Jian	Eli Lilly and Company	chenjianx1@lilly.com	
Chen, Ling	Food and Drug Administration	ling.chen@fda.hhs.gov	M18,M38
Chen, Ming-Hui	University of Connecticut	mhchen@stat.uconn.edu	T37,W11
Chen, Wei	Eli Lilly and Company	chen_wei_j@lilly.com	
Chen, Wei	Wayne State University	chenw@karmanos.org	M32
Chen, Xiang	Yale University	xiang.chen@yale.edu	M16
Chen, Yeh-Fong	FDA	yehfong.chen@fda.hhs.gov	W23
Chen, Yuguo	University of Illinois	yuguo@illinois.edu	M26
Chen, Yun-Fei	Eli Lilly and Company	yfchen@lilly.com	M13
Cheng, Fuxia	Illinois State University	fcheng@ilstu.edu	M34,T24, T38
Cheng, Guang	Dept of Statistics, Purdue University	chengg@stat.purdue.edu	M14,M24, W28
Chiang, Y. Alan	Eli Lilly and Company	chiangay@lilly.com	
Cho, Meehyung	Sanofi-aventis	meehyung.cho@sanofi-aventis.com	M33
Choi, Suktae	Eli Lilly and Company	choisu@lilly.com	
Chuang-Stein, Christy	Pfizer	christy.j.chuang-stein@pfizer.com	K01,L9
Chun, Hyonho	yale	hyonho.chun@yale.edu	T15
Clark, W. Scott	Eli Lilly and Company	wsclark@lilly.com	M13

Name	Affiliation	E-mail	Sessions
Connor, Jason	Berry Consultants	jason@berryconsultant.com	T21
Cook, R. Dennis	School of Statistics, Univ. of Minnesota	dennis@stat.umn.edu	W28
Costigan, M. Timothy	Eli Lilly and Company	tcostigan@lilly.com	
Crainiceanu, M. Ciprian	Johns Hopkins University Biostatistics	ccrainic@jhsph.edu	T14
Craiu, V. Radu	University of Toronto	craiu@utstat.toronto.edu	M26
Crowe, Brenda	Eli Lilly and Company	bjcrowe@lilly.com	L3
Cui, Yuehua	Michigan State University	cuiy@msu.edu	
Cui, Zhanglin	Eli Lilly and Company	Cui_Zhanglin@Lilly.com	
Dean, Charmaine	Simon Fraser University	dean@sfu.ca	M12
Deng, Michelle	Eli Lilly and Company	denghuam@lilly.com	
Denne, Jonathan	Eli Lilly and Company	jon_denne@lilly.com	
Dhavala, S. Soma	TAMU	soma@stat.tamu.edu	
Di Lucca, Maria Anna	UT MD Anderson Cancer Center	pcunning@mdanderson.org	
Dmitrienko, Alex	Eli Lilly and Company	alexei@lilly.com	W12
Donohue, C. Michael	University of California, San Diego	mdonohue@ucsd.edu	W17
Doss, Hani	University of Florida	doss@stat.ufl.edu	M26
Du, Juan	Dept. of Stat. Kansas State University	dujuan@ksu.edu	W15
Dutta, Ritabrata	Purdue University	rdutta@purdue.edu	
Duvvuru, Suman	Eli Lilly and Company	duvvurusu@lilly.com	
Fan, Richard (Yiying)	Case Western Reserve University	richard_fan90@yahoo.com	
Fan, Yingying	University of Southern California	fanyingy@marshall.usc.edu	M27,T27
Fan, Youran	Purdue University	yfan@purdue.edu	
Faries, Doug	Eli Lilly and Company	d.faries@lilly.com	M28
Feng, Yang	UIUC	yfeng4@illinois.edu	W14
Ferguson, Kristie	BSC	kwon@bsci.com	
Fiecas, A. Mark Joseph	Brown University	mfiecas@stat.brown.edu	M25
Fu, Haoda	Eli Lilly and Company	fu_haoda@lilly.com	
Fu, Wenjiang	Michigan State University	fuw@msu.edu	W16
Gao, Lixin	BioPIER	simon.gao@biopier.com	
Gaydos, Brenda	Eli Lilly and Company	blg@lilly.com	T21
Goldberg, Yair	UNC-CH	yair.goldy@gmail.com	T22
Gould, Allan Lawrence	Merck & CO., Inc.	larry_gould@merck.com	Sh. course #4
Grubbs, L. David	Chapman & Hall/CRC	david.grubbs@taylorandfrancis.com	
Guo, Wenge	New Jersey Institute of Technology	wenge.guo@njit.edu	T17
HA, IL DO	Daegu Haany University	idha@dhu.ac.kr	W17
Han, Baoguang	Eli Lilly and Company	hanba@lilly.com	
Han, Xiang	Purdue University	xiang.xhan@gmail.com	
Haraldsdottir, H. Ragnheidur	Columbia University	ragnheidur@stat.columbia.edu	M25
Harezlak, Jaroslaw	Indiana University School of	harezlak@iupui.edu	T34

Name	Affiliation	E-mail	Sessions
	Medicine		
Harvill, L. Jane	Baylor University	jane_harvill@baylor.edu	M34
He, Xian	Purdue University	hexian0119@gmail.com	
He, Xu	Univeristy of Wisconsin-Madison	hexu@stat.wisc.edu	M13
He, Xuming	University of Illinois at U-C	x-he@uiuc.edu	
He, Yulei	Harvard Medical School	he@hcp.med.harvard.edu	T16
He, Yunxiao	Yale University	yunxiao.he@yale.edu	T33,W27
Higgs, Rick	Eli Lilly and Company	higgs@lilly.com	
Ho, Shuyen	GlaxoSmithKline	syh71094@gsk.com	
Holland, Chris	MacroGenics	hollandc@macrogenics.com	W12
Hong, Quan	Eli Lilly and Company	hong_quan@lilly.com	W22
Hong, Sam	MedImmune, Inc	hongs@medimmune.com	M18,M38,T33,W13,W14,W18,W26
Hsu, Chyi-Hung	Novartis Pharmaceuticals	chyihung.hsu@novartis.com	W22
Hsu, Jason	The Ohio State University	Hsu.1@osu.edu	Short course #5
Hu, Ming	University of Michigan	hming@umich.edu	T25
Huang, Chiang Ching	Northwestern University	huangcc@northwestern.edu	T25
Huang, Lei	NYU Child Study Center	huangracer@gmail.com	
Huang, Spencer	Northwestern University	huangcc@northwestern.edu	T25
Hung, H.M. James	FDA	hsienming.hung@fda.hha.gov	
James, Gareth	University of Southern California	gareth@usc.edu	T34
Jeng, X Jessie	University of Pennsylvania	xjeng@upenn.edu	T35
Jeong, Jong-Hyeon	University of Pittsburgh	jeong@nsabp.pitt.edu	
Ji, Yuan	UT M. D. Anderson Cancer Center	yuanji@mdanderson.org	M11,T37
Jia, Haomiao	Columbia University	hj2198@columbia.edu	T31
Jiang, Hongmei	Northwestern University	hongmei@northwestern.edu	M35
Jiang, Jiming	University of California, Davis	jiang@wald.ucdavis.edu	T31
Jiang, Yuan	Yale University	yuan.jiang@yale.edu	T18
Jing, Bing-Yi	Hong Kong Univ. of Science & Technology	majing@ust.hk	T24
Joshi, Adarsh	Eli Lilly and Company	joshi_adarsh@lilly.com	
Kaizar, Eloise	Ohio State University	ekaizar@stat.osu.edu	T17, W14
Kats, Irina	BioInforMedica, Inc.	ikats@comcast.net	
Kim, KyungMann	University of Wisconsin-Madison	kmkim@biostat.wisc.edu	M21
Kim, Yongman	US FDA	yongman.kim@fda.hhs.gov	
Kulkarni, M. Pandurang	Eli Lilly and Company	kulkarni@lilly.com	L8
Kundu, G. Madan	IUPUI	mgkundu@iupui.edu	
Lam, Peter	Boston Scientific	lamp@bsci.com	W11
Lan, G. Kuang-Kuo	Johnson & Johnson	glan@its.jnj.com	L8
Landry, John	Eli Lilly Canada Inc.	landry_john@lilly.com	
Lazar, Nicole	University of Georgia	nlazar@stat.uga.edu	W25
LeBlanc, Michael	Fred Hutchinson Cancer Research Center	mleblanc@fhcrc.org	W16

Name	Affiliation	E-mail	Sessions
Lee, J. Jack	M. D. Anderson Cancer Center	jjlee@mdanderson.org	M32,W11
Lee, Yoonkyung	Ohio State University	yklee@stat.osu.edu	M15
Levine, Michael	Purdue University	mlevins@purdue.edu	T38
Li, Bo	Purdue University	boli@purdue.edu	T36,W15
Li, Fang	IUPUI	fli@math.iupui.edu	M34
Li, Grace	Eli Lilly and Company	li_Ying_Grace@lilly.com	Keynote Banquet
Li, Huilin	NCI	lih5@mail.nih.gov	W16
Li, Jie	Eli Lilly and Company	lijiex1@lilly.com	
Li, Lang	Indiana University School of Medicine	lali@iupui.edu	M35,T23
Li, Lingling	Harvard Pilgrim Health Care Institute	lingling_li@post.harvard.edu	T12
Li, Linyuan	University of New Hampshire	linyuan@math.unh.edu	T24
LI, MINGYU	Celgene	mli@celgene.com	M33
LI, NI	University of Missouri	lini95@gmail.com	
Li, Shaoyu	MSU	lishaoyu@stt.msu.edu	T15
Li, Xiaochun	Indiana University School of Medicine	xiaochun@iupui.edu	T12,W16
Li, Yehua	University of Georgia	yehuali@uga.edu	T14
Li, Yi	Harvard School of Public Health	yili@jimmy.harvard.edu	W17
Li, Yingfu	UHCL	Li@uhcl.edu	
Li, Yue	Eli Lilly and Company	li_yue@lilly.com	
Li, Yun	University of North Carolina	yunli@bios.unc.edu	T25
Liang, Faming	Texas A&M University	fliang@stat.tamu.edu	M36,T26
Liang, Hong	Marshfield Clinic	liang.hong@mcrf.mfldclin.edu	
Liang, Liming	Harvard University	lliang@hsph.harvard.edu	T25
Lin, Nan	Washington University in St. Louis	nlin@math.wustl.edu	T28
Lin, Wen	Allergan, Inc.	lin_wen@allergan.com	
Lin, Xiaodong	Rutgers University	lin@business.rutgers.edu	W27
Lin, Xihong	Harvard University	xlin@hsph.harvard.edu	M16,K02
Lin, Xiwu	GlaxoSmithKline	xiwu.2.lin@gsk.com	T32
Lin, Yanzhu	Purdue University	LIN43@PURDUE.EDU	
Lipkovich, A. Ilya	Eli Lilly and Company	lipkovichia@lilly.com	M22
Liu, Desheng	The Ohio State University	liu.738@osu.edu	W15
Liu, Dungang	Rutgers University	dungang@stat.rutgers.edu	W13
Liu, Hai	Indiana University School of Medicine	liuhai@iupui.edu	
Liu, Jingyi	Eli Lilly and Company	liu_jingyi@lilly.com	T31
Liu, Mengling	New York University	mengling.liu@nyu.edu	
Liu, Minzhi	MacroStat	minzhi.liu@macrostat.com	
Liu, Rong	Eli Lilly and Company	liurongrl@lilly.com	
Liu, Wenlei	Eli Lilly and Company	liu_wenlei@lilly.com	
Liu, X. Shirley	Dana-Farber Cancer Institute / Harvard	xsliu@jimmy.harvard.edu	M35
Liu, Yufeng	University of North Carolina	yfliu@email.unc.edu	M15,T18
Liu, Yushi	the Ohio State Unversity	liuyushi@stat.osu.edu	M18
Liu-Seifert, Hong	Eli Lilly and Company	Liu-Seifert_hong@lilly.com	M21
Lu, Huan	Eli Lilly and Company	lu_huan@lilly.com	

Name	Affiliation	E-mail	Sessions
Lu, Yuefeng	Eli Lilly and Company	lu_yuefeng@lilly.com	M22,M25
Lunagomez, Simon	M. D. Anderson Cancer Center	SLunagomez@mdanderson.org	
Luo, Junxiang	Eli Lilly and Company	luoju@lilly.com	
Luo, Xiaolong	Celgene Corporation	xluo@celgene.com	T11
Lv, Jinchi	University of Southern California	jinchilv@marshall.usc.edu	M15
Ma, Shujie	Michigan State University	mashujie@stt.msu.edu	T38
Ma, Xiwen	University of Wisconsin-Madison	xiwenma@stat.wisc.edu	M13
Maiti, Taps	Michigan State University	maiti@msu.edu	T31
Mallinckrodt, H. Craig	Eli Lilly and Company	cmallinc@lilly.com	W23
Mandrekar, J. Sumithra	Mayo Clinic	mandrekar.sumithra@mayo.edu	M21
Marino, Miguel	Harvard Univ. Dept. of Biostatistics	mmarino@hsph.harvard.edu	W21
Martin, Ryan	IUPUI	rgmartin@math.iupui.edu	T26
Mehta, Cyrus	Cytel Inc	mehta@cytel.com	L1,T21
MENG, XIAOLI	Dept of Statistics, Harvard University	meng@stat.harvard.edu	M17
Michailidis, George	The University of Michigan	gmichail@umich.edu	M15
Mitra, Ritendranath	UT M. D. Anderson Cancer Center	RMitra1@mdanderson.org	
Mou, Jiani	UC Davis	jianimou@gmail.com	
Mueller, Peter	UT MD Anderson Cancer Center	pcunning@mdanderson.org	
Müller, Hans-Georg	UC Davis	mueller95616@netscape.net	
Naranjo, D. Joshua	Western Michigan University	naranjo@wmich.edu	
Natanegara, Fanni	Eli Lilly and Company	natanegara_fanni@lilly.com	
Nettleton, Dan	Iowa State University	dnett@iastate.edu	T17
Nguyen, Thuan	Oregon Health & Science University	nguythua@ohsu.edu	M18,T31
Norton, D. Jonathan	US Food and Drug Admin	Jonathan.Norton@fda.hhs.gov	M38
Obenchain, L. Robert	Risk Benefit Statistics LLC	wizbob@att.net	M28
Offen, W. Walt	Eli Lilly and Company	wwo@lilly.com	M21,K02
Ogawa, Sachio	Eli Lilly Japan K.K.	ogawa_sachio@lilly.com	T11
O'Malley, A James	Harvard Medical School	omalley@hcp.med.harvard.edu	W27
Park, Jincheol	Texas A&M University	jpark@stat.tamu.edu	M36
Pena, A. Edsel	University of South Carolina	pena@stat.sc.edu	T17
Pennell, Michael	The Ohio State Univeristy	mpennell@cph.osu.edu	M38
Petullo, M. David	FDA	david.petullo@fda.hhs.gov	
Pohl, M. Gerhardt	Eli Lilly and Company	pohl_gerhardt@lilly.com	
Price, L. Karen	Eli Lilly and Company	price_k@lilly.com	T32
QI, QI	University of Missouri	qiqi789@gmail.com	
Qian, Hui-Rong	Eli Lilly and Company	qianhu@lilly.com	
Qiao, Meihua	Eli Lilly and Company	qiao_meihua@lilly.com	
Qiao, Xingye	UNC Chapel Hill	qiao@unc.edu	W21
Qu, Yongming	Eli Lilly and Company	quyo@lilly.com	
Randolph, Tim	Fred Hutchinson Cancer Research Center	trandolp@fhcrc.org	T34

Name	Affiliation	E-mail	Sessions
Rashad, Mushfiqur	FDA	mushfiqur.rashid@fda.hhs.gov	
Rayamajhi, Jyoti	Eli Lilly and Company	rayamajhi_jyoti@lilly.com	
Reiss, Philip	New York University & Nathan Kline Inst.	phil.reiss@nyumc.org	T34
Reiter, Jerome	Duke University	jerry@stat.duke.edu	M17
Robins, James	Harvard School of Public Health	robins@hsph.harvard.edu	T12,T22
Rosales, V. Karen Grace	Western Michigan University	k4villar@wmich.edu	
Rotelli, Matt	Eli Lilly and Company	rotel@lilly.com	
Rothman, J. Adam	University of Michigan	ajrothma@umich.edu	T27
Ruberg, J. Stephen	Eli Lilly and Company	sruberg@lilly.com	K01,M22
Rubin, B. Donald	Harvard University	rubin@stat.harvard.edu	K01,M28
Sanger, Todd	Eli Lilly and Company	sanger@lilly.com	L6
Sanil, Ashish	Berry Consultants	ashish@berryconsultants.com	W11
Schaubel, E. Douglas	University of Michigan	deschau@umich.edu	M14
Seaman, John	Baylor University	john_seaman@baylor.edu	
See, Kyoungah	Eli Lilly and Company	seeky@lilly.com	
Sethuraman, Venkat	Novartis Oncology	venkat.sethuraman@novartis.com	T13
Shao, Jun	University of Wisconsin-Madison	shao@stat.wisc.edu	M13,L4
Shao, Qin	The University of Toledo	qin.shao@utoledo.edu	W26
Shen, Changyu	Indiana University School of Medicine	chashen@iupui.edu	T12
Shen, Gang	North Dakota State University	gang.shen@ndsu.edu	W26
Shen, J. Jeremy	Stanford University	jqshen@stanford.edu	T35
Shen, Lei	Eli Lilly and Company	shen_lei@lilly.com	M21,M22, T32
Shen, Shuyi	Genentech, Inc.	shuyis@gene.com	
Shen, Wei	Eli Lilly and Company	shen@lilly.com	Keynote Banquet
Shi, Peipei	Eli Lilly and Company	shi_peipei@lilly.com	
Skrivanek, Zachary	Eli Lilly and Company	zskrivan@yahoo.com	
Small, Dylan	University of Pennsylvania	dsmall@wharton.upenn.edu	M31
Smith, Brian	Amgen, Inc.	brismith@amgen.com	
Song, Yang	Merck Research Labs	yang_song@merck.com	W22
Sparks, JonDavid	Eli Lilly and Company	sparks_jondavid@lilly.com	
Stamey, James	Baylor University	james_stamey@baylor.edu	T32
Stein, Michael	University of Chicago, Dept of Statistic	stein@galton.uchicago.edu	T36
Stevens, R. John	Utah State University	john.r.stevens@usu.edu	W16
Stroud, Jonathan	George Washington University	stroud@gwu.edu	W15
Su, Haiyan	Montclair State University	suh@mail.montclair.edu	T33,W18
Sun, Shuxia	Wright State University	shuxia.sun@wright.edu	T24
Sun, Wei	University of North Carolina	weisun@email.unc.edu	T15
Sun, Wenguang	North Carolina State University	sun@stat.ncsu.edu	M27,T17
Sun, Yanqing	University of North Carolina at Charlott	yasun@uncc.edu	W25
Takeuchi, Masahiro	Kitasato University	takeuchim@platinum.pharm.kitasato-u.ac.j	
Takita, Yasushi	Eli Lilly and Company	takitaya@lilly.com	
Tamura, N. Roy	Eli Lilly and Company	tamura_roy_n@lilly.com	W23

Name	Affiliation	E-mail	Sessions
Tan, Kean Ming	Purdue university	tank@purdue.edu	
Tan, Xin Lu	Purdue University, West Lafayette	tanx@purdue.edu	
Tanaka, Yoko	Eli Lilly and Company	yokot@lilly.com	T11
Tang, Szu-Yu	OSU	anitang0505@yahoo.com.tw	
Tarpey, Thaddeus	Wright State University	thaddeus.tarpey@wright.edu	W23
Taylor, MG. Jeremy	University of Michigan	jmgt@umich.edu	M12,M22,L7
Tchetgen Tchetgen, Eric	Harvard School of Public Health	etchetge@hsph.harvard.edu	T12
Thall, F. Peter	UT M.D. Anderson Cancer Center	rex@mdanderson.org	Short course #1,M11
Ting, Naitee	Boehringer-Ingelheim	naitee.ting@boehringer-ingelheim.com	
Tong, Christopher	USDA	christopher.m.tong@aphis.usda.gov	
Tong, Liping	Loyola University Chicago	ltong@luc.edu	W26
Trudell, Phyllis	Orange County Transportation Authority	pyphyllis@yahoo.com	
Tseng, C. George	University of Pittsburgh	ctseng@pitt.edu	T35
Tsong, Yi	FDA	Rolanda.Inniss@fda.hhs.gov	
Tu, Wanzhu	Indiana University School of Medicine	wtu1@iupui.edu	T34
Wahed, S. Abdus	University of Pittsburgh	wahed@pitt.edu	T22
Wang, Fujun	Eli Lilly and Company	wangfu@lilly.com	
Wang, Haonan	Colorado State University	wanghn@stat.colostate.edu	M37
Wang, Huixia	North Carolina State University	judy_wang@ncsu.edu	
Wang, Jane-Ling	Univ. of California, Davis	wang@wald.ucdavis.edu	M12
Wang, Jiantian	Kean University	jwang@kean.edu	W18
Wang, Jing	University of Illinois at Chicago	wangjing@math.uic.edu	M34
Wang, Junhui	University of Illinois at Chicago	junhui@uic.edu	T18
Wang, Libo	Purdue University	wang220@purdue.edu	
Wang, Lifeng	Michigan State University	wang@stt.msu.edu	M15
Wang, Lily	University of Georgia	lilywang@uag.edu	M37,T38
Wang, Lu	University of Michigan	luwang@umich.edu	M14,M24
Wang, Ming-Dauh	Eli Lilly and Company	md_wang@lilly.com	
Wang, Ping	Eli Lilly and Company	wangpi@lilly.com	
Wang, Sue-Jane	FDA	suejane.wang@fda.hhs.gov	Short course #6,M22
Wang, Tao	Pfizer	tao.wang7@pfizer.com	
Wang, Wei	Eli Lilly and Company	wang_wei_v@lilly.com	
Wang, Yanping	Eli Lilly and Company	wang_yanping@lilly.com	K02,T21
Wang, Zhifeng	Eli Lilly and Company	sandals@yahoo.com	
Wathen, Kyle	M. D. Anderson Cancer Center	ekribb@mdanderson.org	
Wehrenberg, Scott	Boston Scientific Corp.	scott.wehrenberg@bsci.com	W11
Wei, Hank	Eli Lilly and Company	hank_wei@lilly.com	
Wellner, A. Jon	Statistics,University of Washington	jaw@stat.washington.edu	M14
Whitlock, Kathryn	Indiana University	kburr@indiana.edu	
Woo, Rick	Bio Pharma Consultancy Company	rickwoo2005@yahoo.com.cn	

Name	Affiliation	E-mail	Sessions
Wu, Changbao	University of Waterloo	cbwu@uwaterloo.ca	W25
Wu, Dongfeng	University of Louisville	dongfeng.wu@louisville.edu	
Wu, S. Jeremy	U.S Bureau of the Census	jeremy.s.wu@census.gov	M17
Wu, J. Jiangming	Merck Research Laboratory	johnnywu1723@yahoo.com	
Wu, Jingwei	Indiana University, School of Medicine	wujw@iupui.edu	
Wu, Lang	University of British Columbia	lang@stat.ubc.ca	M12
Wu, Sutan	Roche Diagnostics	sutan.wu@roche.com	
Wu, Yichao	North Carolina State University	wu@stat.ncsu.edu	M27,T18, T27
Wu, Yujun	Sanofi-Aventis	yujun.wu@sanofi-aventis.com	M33,W24
Wu, Yuping	Cleveland State University	y.wu88@csuohio.edu	
Xi, Zhuangzhuang	Purdue University	zxi@purdue.edu	
Xia, Jin	Purdue University	jxia@purdue.edu	
Xie, Dawei	University of Pennsylvania	dxie@mail.med.upenn.edu	T16
Xie, Jun	Purdue University	junxie@purdue.edu	M33,T26
Xie, Li	Eli Lilly and Company	xie_li_lx@lilly.com	
Xie, Lingling	Eli Lilly and Company	xieln@lilly.com	
Xie, Tai	Brightech International	tomx@brightech-intl.com	L2
Xu, Chang	Eli Lilly and Company	xu_chang_g@lilly.com	
XU, HANGTAO	Eli Lilly and Company	xu_hangtao@lilly.com	
Xu, Jane	Sepracor	Marisa.Tellez@sepricor.com	
Xu, Lei	Eli Lilly and Company	xule@lilly.com	
Xu, Ronghui	UC San Diego	rxu@ucsd.edu	M31
Xu, Wei	University of Georgia	xuwei@uga.edu	
Xu, Wen	Eli Lilly and Company	xu_wen_isabel@lilly.com	
Xue, Hongqi	University of Rochester	Hongqi_Xue@urmc.rochester.edu	W13
Xue, Lan	Oregon State University	xuel@stat.oregonstate.edu	M34,M37
Yang, Lili	IUPUI	yanglili@umail.iu.edu	
Yang, Yuhong	University of Minnesota	yyang@stat.umn.edu	M37
Yeo, Kwee Poo	Eli Lilly and Company	kweepoo@lilly.com	
Yi, Grace	University of Waterloo	yyi@uwaterloo.ca	M12,M31
Yin, Helen	MacroStat	helen.yin@macrostat.com	L4
Yin, Xiangrong	University of Georgia	xryin@stat.uga.edu	M37,T28
Yu, Kai	DCEG, NCI	yuka@mail.nih.gov	M36
Yu, Lili	Georgia Southern University	lyu@georgiasouthern.edu	M23
Yu, Menggang	Indiana University School of Medicine	meyu@iupui.edu	T23
Yu, Zhangsheng	Indiana University School of Medicine	yuz@iupui.edu	
Yu, Zhinuan	Celgene Corporation	zyu@celgene.com	
Yuan, Ming	Georgia Tech	myuan@isye.gatech.edu	T27
Yuan, Ying	Univ. of Texas MD Anderson Cancer Center	yyuan@mdanderson.org	M11
Zeng, Donglin	University of North Carolina	dzeng@bios.unc.edu	M24
Zeng, Peng	Auburn University	zengpen@auburn.edu	W28
Zhang, Chunming	University of Wisconsin	cmzhang@stat.wisc.edu	
Zhang, Dabao	Purdue University	zhangdb@purdue.edu	T27
Zhang, Guangxiang	Stony Brook University	georgexzh@gmail.com	
Zhang, Heping	Yale University	heping.zhang@yale.edu	M22

Name	Affiliation	E-mail	Sessions
Zhang, Lu	Eli Lilly and Company	lu.zhang@lilly.com	W23
Zhang, Min	Purdue University	minzhang@stat.purdue.edu	M18,M38
Zhang, Qi	Eli Lilly and Company	zhangqix@lilly.com	
Zhang, Wei	Eli Lilly and Company	zhang_wei_wz@lilly.com	
Zhang, Xiaoxi	Pfizer Inc.	xiaoxi.zhang@pfizer.com	M25
Zhao, Luping	Eli Lilly and Company	zhao_luping@lilly.com	
Zhao, Sihai	Harvard Univ. Dept. of Biostatistics	szhao@hsph.harvard.edu	W21
Zhao, Daniel Yan	UT Southwestern Medical Center	Daniel.Zhao@utsouthwestern.edu	W12
Zheng, Hanzhe	Merck Co., Inc	hanzhe.zheng@spcorp.com	
Zheng, Tian	Columbia University	tz33@columbia.edu	
Zheng, Xianming	Eli Lilly and Company	zhengxi@lilly.com	
Zhong, Wenxuan	Univer. of Illinois at Urbana-Champaign	wenxuan@illinois.edu	
Zhou, Chunmei	Eli Lilly and Company	cmzhou30@lilly.com	
Zhou, Harrison	Yale University	huibin.zhou@yale.edu	T14
Zhou, Lan	Texas A&M University	lzhou@stat.tamu.edu	T14
Zhou, Yinmei	St. Jude Children's Research Hospital	yinmei.zhou@stjude.org	
Zhu, Jun	Colorado State University	jzhu@stat.colostate.edu	T36
Zhu, Ling	Covidien	ling.zhu@covidien.com	
Zhu, Yanni	Eli Lilly and Company	zhu_yanni@lilly.com	T28,W24
Zhu, Zhengyuan	Iowa State University	zhuz@iastate.edu	
Zou, Hui	University of Minnesota	hzou@stat.umn.edu	M27

Jennifer *- Research Scientist for Lilly*

Every action has a reaction. A thought that sparks curiosity. A step that leads to discovery. And an opportunity to improve the lives of millions. Every decision you make can change the future. Make Lilly your next big move and fully realize the power of choice.

At Lilly, you will have opportunity. The chance to research, to work with promising new methods and develop new drugs, and to move within a company with over 130 years of history. Choose Lilly and make your future, and that of others, a better one.

Be connected. Be inspired. **Be a catalyst.**
Visit lilly.com/careers

Lilly

Join the Conversation

It is now easier than ever to become engaged in the ASA and its members through social media.

ASA Community—Join this online setting for ASA members to communicate, collaborate, and share.

Twitter—Follow *Amstat News* on Twitter @AmstatNews and keep up to date on the profession.

Facebook—Check out the ASA's page on Facebook to keep up with the latest deadlines, news, and activities and to share your tips and comments with colleagues worldwide.

Engage your fellow statisticians and enhance your mind, education, and career at ***www.amstat.org***.

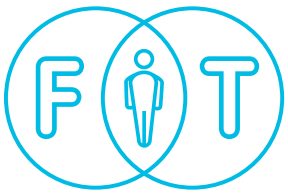
inVentiv Clinical

Built to FIT

Early in the history of inVentiv Clinical, we made a decision that would have a profound impact on our company. We decided to build our business through a customer-centric approach. This has allowed us to nimbly meet customer requirements as they change, provide flexible services and frameworks, and solve problems.

It is a simple formula that explains how we're "Built to **FIT**".

Flexibility around customer needs, with
Infrastructure, Intelligence, and Integrity to
Tailor Solutions that accelerate results

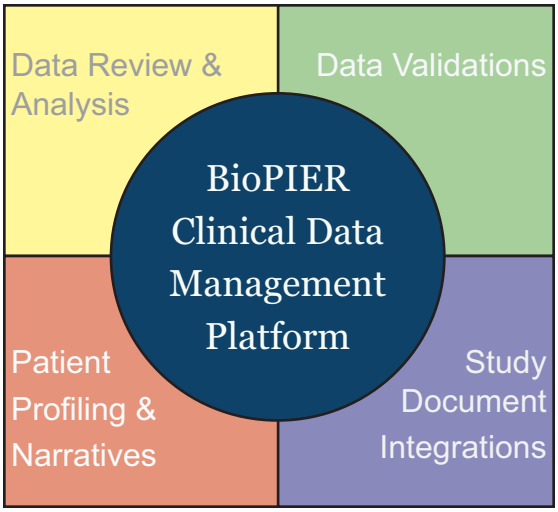


So whether you need comprehensive **clinical development solutions**, a **dedicated team** of professionals working without the constraints of co-employment, or a few specialized **contractors** for an upcoming project, we can help effectively manage your resources, drive costs down, and accelerate results.

866.707.9567 info@inventivclinical.com www.inventivclinical.com

an inVentiv health company

Better Tools for Clinical Data Browsing & Reporting



If you are looking for a system for data review, validation, and patient profiling, BioPIER Clinical Workbench™ is the one! One system works for all and it is very easy to implement!

Contact us for a demo or free trial

How can you benefit?

You are a ...	You can benefit ...
Member of a Clinical Study Team	This is a friendly easy-to-use "DIY" data review tool. One simple interface to access all studies, and you can be on the top of study data all the time.
Clinical Data Manager (CDM)	CDMs are empowered by this real-time data review tool. Easily configure data listings and reports; Proactively run data validation (QC) and research ongoing data. All work are done without requiring SAS or programming.
Biostatistician / Clinician/Safety / Medical Director	View ongoing key-indicators, play what-if scenarios, and study patient profiles for vital efficacy outcomes and safety events. When you review a summary, numbers on the report are hyperlinked to the original data.
Statistical Programmer	SAS coding is automated from one-click" to produce paper-trail CSR report.
Clinical Project Manager/CRA	Be informed with "real time" clinical trial status.
IT Manager	Web-based enterprise architecture. SaaS service is available. Very easy to deploy the system on the customer site, one company just needs one installation.

Business cases to resolve

Data Review and Patient Profiling

"Our clinical team needs a tool to review data from clinical clinicals. Such needs can be anywhere from a quick serious adverse event(AE) listing, to a summary of AEs by age group, or patient profiles for these serious AE patients. Since it is ad-hoc and on-going, it would be a great deal if such a tool exists."

Outsourced Clinical Data Validation

"Our clinical trials have been outsourced to CROs. Some CROs did a great job, and some didn't. Our CDM group needs a tool to review and validate data. Our CDM staffs can define and execute post hoc validation checks on received data to find issues in the early stage. We are short of CDM staffs, and most don't have SAS® or SQL knowledge, a tool that is easy to use is vital."

Clinical Data Repository

"Our studies are from hybrid sources, mostly non-CDISC. We are looking for a tool that can funnel data from different studies and provide a centralized view. The team can use the same tool the same interface to run summary statistics, data listings, and patient profiles and without requiring SAS® programming."

BioPIER
the drug information platform

About BioPIER Inc.

BioPIER provides web-based biometrics and data management solutions for clinical trials and drug safety.

Contact: Lixin (Simon) Gao, President
394 Lowell St, Suite 17
Lexington, MA 02420
Tel: (888) 9-BIOPIER
Email: info@biopier.com

http://www.biopier.com

Doing Well By Doing Good®

- **Want to Make a Difference?**
- **Want to Join a Growing Company?**

...Here You Can!

At Celgene, our innovative therapies for the treatment of cancer, immunology and inflammatory disorders and other life-threatening diseases are part of our commitment to improving the lives of patients around the world. This is our culture!

We're an organization that will value your ideals, reward your achievements and encourage your pioneering spirit. Listed #7 on Forbes "Fastest Growing Tech Companies" list, Celgene boasts the structure and organization of a global powerhouse, the urgency and entrepreneurship of a start-up company and the warmth of a family business.

To view current opportunities, we encourage you to visit www.celgene.com

Celgene

Change the course of medicine with us!

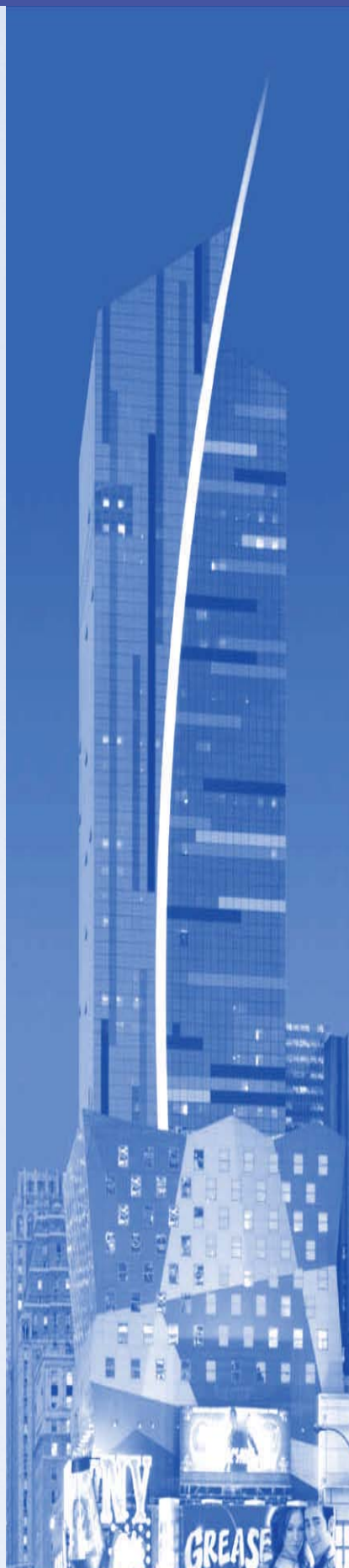
International Chinese Statistical Association

泛華統計協會

ICSA 2011
Applied Statistics Symposium

The Westin New York
At Time Square
June 26-29, 2011

We cordially invite you to join us
In New York City
for
the Celebration of
the 20th Anniversary of
ICSA Applied Statistics Symposium!



Keynote Speakers:

David Donoho *(Stanford University)*
Danyu Lin *(UNC-Chapel Hill)*
Ji Zhang *(sanofi-aventis)*

Banquet Speaker:

James Fu *(University of Manitoba)*
"The history of the International Chinese Statistical Association and ICSA Applied Statistics Symposium"

Executive Committee:

Chair: *Zhezhen Jin (zj7@columbia.edu)*
Members: *Gang Li , Mark Chunming Li, Mengling Liu, Hui Quan, Ying Wei, Wei Zhang and Tian Zheng.*

Hyatt Regency Indianapolis

at State Capitol

DIRECTIONS

From Indianapolis International Airport (8 miles): Take I-70 East to downtown.
Exit at Illinois St. North (79B). Proceed to Washington St. Turn left.
Go one block to Capitol Ave. Turn left. Hotel is on left.