2019 (Bio)Statistics Research and Career Day

September 27th, 2019
McGill University

McIntyre Medical Building, Room 210
3655 Promenade Sir William Osler

- 09:00–09:30 Registration and coffee
- 09:30–09:45 Welcome word
- 09:45–10:45 Keynote speech: Dr. Rui Wang
- 10:45–11:00 Coffee break I
- 11:00–12:00 Student presentation session I
- 12:00–13:45 Pizza lunch

McIntyre Medical Building, Room 521

- 13:45–15:00 Student presentation session II
- 15:00–15:30 Coffee break II
- 15:30–16:30 Keynote speech: Dr. Murray Clayton

Purvis Hall
1020 Avenue des Pins Ouest

- 17:30–18:30 Career panel
- 17:00–20:00 Wine and cheese

Student presentation session I

- 11:00–11:15 Zayd Omar
- 11:15–11:30 Daniel Rodriguez Duque
- 11:30–11:45 Md. Shaddam Hossain Bagmar
- 11:45–12:00 Armando Turchetta

Student presentation session II

- 13:45–14:00 Tyrel Stokes
- 14:00–14:15 Jiajun Mai
- 14:15–14:30 Deborah Chan
- 14:30–14:45 Fabien Baeriswyl
- 14:45–15:00 Brice Batomen

Career Panel

- Dr. Rui Wang, Harvard University
- Dr. Murray Clayton, University of Wisconsin-Madison
- Dr. Rebecca Burne, Groupe d'Analyse
- Dr. Kathryn Morrison, Precision Analytics
- Sarah Vahey, Innovaderm
- Alexandre Deschênes, TD Insurance

Members of the organizing committee:

Shomoita Alam
Janie Coulombe
Yi Lian
James McVittie

Thanks to our sponsors:

CANSSI
PGSS
EBOSS
Department of Mathematics and Statistics
Department of Epidemiology, Biostatistics and Occupational Health

Thanks to the people who helped in organizing the event:

Yi Yang, Raffaella Bruno
Department of Mathematics and Statistics

Nathalie Theoret, André Yves Gagnon
Department of Epidemiology, Biostatistics and Occupational Health

Guillermo Martinez-Zalce
Centre de Recherches Mathématiques, Université de Montréal

Keynote Presentations Abstracts

Design, Monitoring and Analysis of Cluster Randomized Trials

Dr. Rui Wang

Dr. Rui Wang is an Associate Professor of Population Medicine and Director of the Division of Biostatistics in the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Health Care Institute. Dr. Wang's interests include the design, monitoring, and analysis of parallel and stepped-wedge cluster randomized trials, where a group of subjects, as opposed to individuals, are randomized to each of the treatment arms in the trial. She has also been developing improved statistical techniques for a cross-sectional approach that, when combined with modern HIV screening methods, can substantially reduce the cost and increase the accuracy of HIV incidence estimation. Her research interests also include longitudinal modeling of non-linear trajectories and model selection, as well as addressing missing data issues in distributed data networks.

ABSTRACT

Cluster randomized trials are well-suited to evaluate intervention strategies against infectious diseases. In this talk, I will present a series of methods to address statistical challenges in the design, monitoring, and analysis of CRTs, motivated by a cluster-randomized HIV prevention trial in Botswana. To facilitate the design of the study, we develop an agent-based network/epidemic model to simulate the community-level impact of a combination prevention strategy that accounts for partnerships formed within and between clusters. Recognizing the commonly-faced challenge in obtaining accurate information on the coefficient of variation, a measure of the clustering effect and a major driver of the study power, we develop a random-effects model for cross-sectional HIV incidence estimation that provides one way to estimate this parameter based on information at baseline. During the course of the study, changes in treatment guidelines for HIV-infected individuals made the difference between the intervention and the control arms smaller and raised concerns about the study power. We propose a flexible approach for conditional power estimation when the outcomes are correlated and interval-censored. Finally, we develop semi-parametric doubly robust estimators for both the first-order intervention effect and second-order association parameters in the presence of informative missing outcome data.

Regression Models for Spatial Images

Dr. Murray Clayton

Dr. Clayton is a Professor Emeritus in the Departments of Plant Pathology and Statistics at the University of Wisconsin-Madison. His research interests include the development of theoretical statistics as well as the development of statistical tools to address complex problems in the agricultural, environmental and biological sciences. For example, he has focused on the detection and description of patterns of plant and human diseases across large geographical regions. He has applied Bayesian and non-Bayesian methods, coupled with Markov Chain Monte Carlo techniques, to address spatial clustering problems in epidemiology such as identifying locations where the rates of cancer might be enhanced. He has also worked on association studies of variables measured across regions in which there may be spatial correlation. He has collaborated with many scientists on a diverse array of problems, including modeling bee movements to better predict gene flow from transgenic to nontransgenic crops as well as designing clinical trials to compare diets for persons with certain metabolic disorders, to name a few.

ABSTRACT

This work is motivated by a problem in describing forest nitrogen cycling, and a consequent goal of constructing regression models for spatial images. Specifically, I present a functional concurrent linear model (FLCM) with varying coefficients for two-dimensional spatial images. To address overparameterization issues, the parameter surfaces in this model are transformed into the wavelet domain and then sparse representations are found using two different methods: LASSO and Bayesian variable selection. I will briefly discuss extensions to address missing data problems for colocated spatial images and the modeling of tree species in landscape ecology. In addition I will discuss the use of the sextant in marine navigation.

Student Presentations Abstracts

Estimating ICU Heart Rate Data Using a Bayesian State-Space Model with GARCH(1,1) Errors

Zayd Omar, Alexandra Schmidt, David A. Stephens

Heart rate data from the ICU often resembles a non-stationary time series and displays time varying volatility. The non-stationarity shown in the data, makes modelling heart rate difficult using standard techniques used for time series data. Gaussian State space models are a better approach since they allow us to model the non-stationarity observed using latent state vectors. Taking a Bayesian approach, we extend the univariate and the multivariate Gaussian state-space model by adding a GARCH(1,1) component at the observation level, which allows us to model the conditional heteroscedasticity. We also study these model under structural breaks. We propose a Markov chain Monte Carlo (MCMC) algorithm that allows us to estimate the GARCH(1,1) parameters, the latent state vectors and the variance of the state vectors. We compare the estimates from our model to the estimates from a standard Bayesian state space model.

Bayesian Inference for Dynamic Treatment Regimes via Marginal Structural Models

Daniel Rodriguez Duque, Erica E.M. Moodie, David A. Stephens

In precision medicine, researchers are often concerned with studying sequences of decision rules that allow them to tailor patient care through time. These decision rules are termed dynamic treatment regimes (DTRS), and interest lies in identifying optimal DTRs, for example ones that maximize quality of life. In a frequentist setting, inference for DTRs often suffers from non-regularity of estimators which hinders asymptotic inference. Furthermore, the frequentist approach does not always provide coherent methods of estimation as plug-in estimators are used without the necessary propagation of uncertainty. Recently, a Bayesian semi-parametric approach was developed for estimating the causal effect of time-varying treatments via marginal structural models. Importantly, the inverse weighting used to fit these models is motivated by an importance sampling argument that allows us to draw inference in a world in which patients are sequentially randomized to treatments by considering data from a non-experimental world. In this presentation, we outline how these methods may be extended to estimate the causal effects of DTRs by considering a world in which patients are randomized to DTRs at baseline. In particular, we show how to draw singly and doubly robust inference via the maximization of a posterior predictive utility. Simulation studies are presented demonstrating our approach unbiasedly estimates the expect utility under a given DTR as well as correctly identifies the optimal DTR within a family of regimes.

Causal Inference with Missingness in Confounders

Md. Shaddam Hossain Bagmar, Hua Shen

Causal inference is the process of uncovering causal connection between the effect variable and disease outcome in epidemiologic research. Confounders that influence both the effect variable and outcome, need to be accounted for when obtaining the causal effect in observational studies. In addition, missing data often arise in the data collection procedure, working with complete cases often results in biased estimates. We consider the estimation of causal effect in the presence of missingness in the confounders under the missing at random assumption. We propose an expectation-maximization (EM) algorithm to estimate the expected values of the missing confounder and utilize weighting approach in the effect estimation to obtain doubly robust estimators. Both simulation and real-life data application has been considered to see whether there is any gain in estimation efficiency under the proposed method than complete case analysis and multiple imputation. The analysis identified EM method as most efficient and accurate method for dealing missingness in confounder.

SMART Design and Sample Size Considerations

Armando Turchetta, Erica E.M. Moodie, David A. Stephens

In the management of most chronic conditions characterized by the lack of a widely effective treatment, adaptive treatment strategies (ATS) have recently become popular as they can be tailored to the patients evolving conditions. Their goal is the detection of the optimal sequence of interventions that leads to the best outcome for the patient concerned. Over the last decade, sequential multiple assignment randomized trials (SMARTs) have been gaining popularity as the most suitable clinical trial design to formalize the study of these strategies. SMARTs are based on multiple stages each representing a clinical decision point: at each step, the patients are randomized accounting for their characteristics and responses to previous interventions. While the number of SMARTs has been increasing during the last years, and it is clear they have the potential for playing an important role in the management of chronic conditions, their theoretical features are yet to be fully discovered. In particular, power and type I error are strongly affected by deviations from the assumed rate responses in the sample size calculations. In this presentation, we show the main features of the SMART design, an example of its application in a study on children affected by ADHD, and we focus on some problems arising from sample size calculations through a small simulation study.

Simulating Bias Amplification

Tyrel Stokes, Russell Steele, Ian Shrier

In ignorability approaches to identifying Average Causal Effects, we seek to condition on a set of variables such that the potential outcomes are conditionally independent of the treatment. In non-experimental settings, however, the entire set of variables required to satisfy the ignorability condition is often not available and confounding pathways remain. Recent theoretical work has explored an important class of variables which can further

amplify existing unmeasured confounding bias when conditioned on. This is contrary to intuition developed in causal variable selection procedures, that conditioning on additional observables never increases unmeasured confounding and only increases variance. In this talk I will talk about the geometric origins of Bias Amplification and the problem of developing causal variable selection techniques for combating it. Finally, I will consider the challenge of simulating bias amplification in a system of equations such that the replications can be properly viewed as a counter-factual experiment.

Semiparametric Estimation in Targeted Learning by Numerical Construction of the Efficient Influence Function

Jiajun Mai, David A. Stephens

Semiparametric estimation is now widely used in many applications because it is potentially more flexible and robust than the parametric estimation due to the reduction of parametric assumptions. However, the general semiparametric estimator can be difficult to derive analytically and some of the assumptions required are difficult to satisfy. In this presentation, we will study a recent widely used estimation technique that is based on an implementation of semiparametric theory named targeted learning. We study its statistical properties theoretically and numerically, and study a recently introduced alternative semiparametric estimation procedure that is based on computing the numerical Gâteaux derivative. We give some theoretical and numerical examples to show that this method does not always produce consistent and asymptotically efficient estimators.

Sex-specific association of human milk hormones and asthma in the CHILD cohort

Deborah Chan, Allan B. Becker, Theo J. Moraes, Piushkumar J. Mandhane, Malcolm R. Sears, Stuart E. Turvey, Padmaja Subbarao, Catherine J. Field and Meghan B. Azad

Background: Asthma is the most common chronic childhood disease, affecting 15-20% of children worldwide. It is well established that asthma affects more boys than girls in early childhood, but this sex difference is not fully understood. The impact of human milk on asthma development is unclear and potential sex-specific effects of human milk composition may play a contributing factor, though little is known about this phenomena. Human milk contains adiponectin, leptin, and insulin which primarily regulate appetite but may influence immune development.
Methods: We studied a representative subset of breastfed children from the CHILD birth cohort (220 males and 210 females). Mothers provided a sample of human milk at 3-4 months postpartum. Milk leptin, insulin and adiponectin were measured using the MesoScale Discovery System. Diagnosis of possible or probable asthma at three years of age was determined by standardized medical history and physical exam. Logistic regression was used to explore the relationship between human milk hormones and asthma while adjusting for relevant covariates. An interaction term was used to evaluate sex differences.

Results: Milk hormone concentrations did not differ by whether mothers had a girl or boy. Overall, 12% of children were diagnosed with possible or probable asthma at 3 years of age. In stratified analysis, higher concentration of milk leptin (aOR 2.62 95%CI 1.21, 6.13) and milk insulin (aOR 2.95 95%CI 1.60, 5.96) were associated with higher odds of possible and probable asthma in girls only but not boys after adjusting for relevant covariates (study site, maternal BMI, maternal self-report of physician-diagnosed asthma, breastfeeding exclusivity and exact infant age at sample collection). These associations persisted after adjusting for child BMI and subscapular skinfolds at 3 years. Higher concentrations of milk insulin remain associated with higher odds of possible and probable asthma after adjusting for all three hormones in girls (aOR 2.51 95%CI 1.31, 5.24) only.
Conclusion: Breastfed girls consuming milk with high concentrations of leptin and insulin were more likely to develop possible or probable asthma by 3 years of age. This association was not mediated by body composition and was not seen in boys.

Self-exciting process goes extreme*

Fabien Baeriswyl, Maximilian Aigner, Valérie Chavez-Demoulin

Self-exciting point process models arise in situations where the history of the process, and specifically occurrence of past points, has a positive impact on the probability of observing new points. These models are particularly used in earthquakes modelling and epidemic forecasting. On the other hand, the classical GPD distribution arise naturally to model tails of distribution. In this presentation, we consider a marked point process in which the subprocess for the time and location of an event is a self-exciting process and the marks are GPD. We apply this methodology to earthquake data and underline the advantages of such a model.

* Title kindly adapted from Opitz et. al. (2018), "INLA goes extreme (...)"

Estimating the impact of a hospital intervention on length of stay in a competing risks framework

Brice Batomen, Lynne Moore, Erin Strumpf, Arijit Nandi

Background: Hospital length of stay is an important outcome when assessing the impact of hospital interventions such as accreditation. In this context, in-hospital mortality acts as a competing risk, given that it makes it impossible for a patient to be discharged. However, the competing risk of death is rarely appropriately handled. In the presence of competing risks one-to-one correspondence between rate and risk is lost. In addition, for measuring the public health impact of interventions, risk differences or ratios are preferred to the cause-specific and subdistribution hazards. We aim to present an approach for estimating the effect of trauma center accreditation, on length of stay that considers in-hospital mortality as a competing risk in both a pre-post and interrupted time series design.
Methods: Data are from admissions to a level I adult trauma center in Quebec, Canada between 2008 and 2017. We estimated the standardized risk of being discharged following

accreditation by combining inverse probability weighting and the Aalen-Johansen estimator of cumulative incidence function.
Results: 5,300 severely injured patients were admitted during the study period. Results suggest that accreditation was associated with standardized risk differences of discharge at 14-day of 8.3% (95% CI: 6%, 10.6%) and 8.7% (95% CI: 2.1%, 15.3%), respectively, in pre-post and interrupted time series designs.
Conclusion: After accounting for measured confounders and the competing risk of in-hospital mortality, accreditation seems to reduce the length of stay by increasing the risk of discharge following a patient admission.

Career Panel

Dr. Rui Wang

Dr. Murray Clayton

Dr. Rebecca Burne

Rebecca Burne joins us from Groupe d'Analyse where she has worked as an Associate since 2017. Prior to starting at Groupe d'Analyse, Rebecca completed her Ph.D. in Biostatistics at McGill, during which time she also worked as a Research Assistant and as Co-Course Lecturer for the course, 'Analysis of Correlated Data.'. At Groupe d'Analyse, Rebecca has worked on studies of clinical and economic outcomes in a broad range of disease areas, including oncology and cardiovascular diseases.

Dr. Kathryn Morrison

Dr. Kathryn Morrison is an accredited statistician with over twenty peer-reviewed publications in journals such as Statistics in Medicine and The Lancet. Her training was a blend of statistics, epidemiology, and geographic information science. As founder and chief technology officer of Precision Analytics, she oversees the team of data scientists and software engineers, interface with clients to answer statistical and scientific questions, and guides the development of statistical technology at the company. Kathryn still participates in research as an Adjunct Professor at McGill University, and in the open-source `R` community as a co-organizer of `R` Ladies Montreal. In her free time she loves knitting, running, podcasts, and dogs.

Sarah Vahey

Sarah's current position is as a Biostatistician for a CRO (contract research organization) called Innovaderm. She started working in the Pharma Industry in 2008 after a spell as a research assistant for Samy Suissa. She started out working on bioequivalence trials for generic drugs with MDS Pharma Services. After that she moved to Quintiles for almost 5 years where she mainly worked on Oncology studies, firstly as a SAS programmer and then as a Biostatistician. She worked for a short while with Algorithme Pharma who specialize in bioequivalence trials also. She has been with Innovaderm for over a year now, their specialty being Dermatology.

Alexandre Deschênes

Alexandre Deschênes has been a data scientist at TD Insurance for over 7 years now. Alexandre works with TD Insurance business owners to identify key analytical opportunities. He provides technical knowledge and expertise on predictive modeling, machine learning and AI capabilities for claim management process improvement. Alexandre has a Master's degree in Statistic from Université de Montréal.