

---

## Poster Session

---

---

**OLU AWOSOGA**, University of Lethbridge

*Multi-site Pilot Randomized Control Trial of Congruence Couple Therapy for Problem Gamblers*

This study was conducted in Ontario and Alberta, Canada from 2009-2011 to compare the status of couples in treatment to control condition with usual treatment and non-treatment. Treatment couples received 12-week CCT while control couples received 3 brief check-ins over 12 weeks. Baseline mean DSM-IV gambling score = 8.7/10. Retention of the treatment couples was 89% at 2-month follow-up. Retention of control couples was 78%. Of interest is the non-intended positive effects of couple research participation on control participants. The sample (N=30; 15 couples) consisted of 66% male gamblers and 34% female. CCT treatment was found to be well accepted.

---

**ENTAO CHEN**, Acadia University

*A Beta Distribution Based Hierarchical Multinomial Model for Age Structure Analysis of Wildlife Animal*

Nested with a modified Beta distribution for estimating animal survivorship shape, a hierarchical multinomial model is developed to estimate the underlying yearly age distribution by sole usage of yearly age-specific harvest counts. Evaluation of model performance and robustness, by conducting extensive Monte Carlo simulation studies, indicates the accuracy and coverage of estimation under various simulation scenarios. A further analysis of performance will be provided by the real application on two focal age-at-harvest datasets.

---

**HAO (NELSON) CHEN**, University of British Columbia

*Tree-based Methods for Emulation of a Complex Computer Model*

Many complex phenomena are difficult to investigate through controlled physical experiments. Instead, computer models become important alternatives to provide insights into such phenomena. A Gaussian Process (GP) is commonly used as a statistical surrogate for the input-output relationship of a computer model. However, a GP has a strong assumption of stationarity of the output. This drawback can be overcome by tree-based methods, which split the output space and fit separate statistical surrogates within each subregion. In this poster, we first review several tree-based methods. A comparison between different tree-based methods is conducted via simulation and the optimal method is identified.

---

**JINGJIA CHU**, University of Western Ontario

*Modelling the Common Risk among Equities: A Multivariate Time Series Model with an Additive GARCH Structure*

The DCC GARCH models (Engle and Sheppard, 2001) have been well studied to describe the conditional covariance and correlation matrixes while the common risk among series cannot be captured intuitively by the existing multivariate GARCH models. A new class of multivariate time series model with an additive GARCH structure is proposed. The dynamic conditional covariances between series are aggregated by a common risk term which has been the key to characterize the conditional correlation. The model we proposed here can be implemented to apply to institutional portfolio management to determine the weights on equities and fixed income securities.

---

**HYUKJUN GWEON**, University of Waterloo

*Automated Occupational Coding Using Machine Learning*

Occupational coding refers to categorizing a survey respondent's text answer into one of hundreds of occupation codes. Instead of manual coding, we propose automated coding using machine learning. Automated coding is a challenging problem because answers usually consist of only a few words while there are hundreds of possible categories. Nevertheless, the use of machine

learning approaches may be helpful for a fraction of records. We include simulation results using occupational data from a German agency.

---

**FANG HE**, University of Western Ontario  
*Subscene design in R package rgl*

There are several packages in R (R Core Team, 2014) handling 3D plots, such as *rgl*, *ggplot2*, etc. The advantage of *rgl* is mouse control to rotate or zoom the scene. There were still some limitations in *rgl* prior to version 0.94. Our changes were inspired by the 2D plot package *grid* (Murrell, 2006). We have implemented “subscenes” which have properties similar to “viewports” which is a “rectangular region that provides a context for drawing” (Murrell, 2006). These new features could form the basis of interactive editing in the future.

---

**JAMES KIBERD**, Dalhousie University  
*The Good, the Bad, and the Cite-Able: Study Quality and Statistical Significance as Predictors of Study Citation Rates*

The frequency with which an academic article is cited is a marker of its contribution to the literature. Both higher quality studies and those reporting significant results tend to be cited more frequently in the literature compared to lower quality studies or those with null findings. We will evaluate the relative contributions of study quality, reported statistical significance, and agreement with field consensus (based on published meta-analyses) on citation frequency for studies cited in a sample of Cochrane meta-analyses. The impact of these factors on the frequency of citations may indicate potential biases in how authors evaluate the literature.

---

**KAREN A KOPCIUK**, Cancer Epidemiology and Prevention Research, AHS/University of Calgary  
*Distribution-Based Imputation for Left-Censored Metabolomics Data*

Mass spectrometry instruments used to measure metabolomics features have left limits of detection that can result in substantial missing data. Imputing a constant value results in both bias and reduced variability in their distributions while dropping them altogether could distort relationships with other metabolites. Since estimation in projection-based methods such as principal component and partial least squares regression models is based on variation, misleading results are likely. A distribution-based imputation method is proposed to recover the missing data values and compared with imputed constant values using simulated data. Performance is evaluated using the area under the ROC curve.

---

**TAKUMA KUROSAWA**, Tokyo University of Science  
*On Asymptotic Normality of Maximum Partial Likelihood Estimators for Binary Time Series with Transition Models*

In this study, we consider a transition model to deal with binary time series, which have binary observations depended on some past observations and covariates. To estimate the parameters in the model, we can use the maximum partial likelihood estimate method. We prove that the estimators have asymptotic normality under some conditions. In addition, we study that asymptotical behavior of the estimators with finite sample size through simulation studies.

---

**WAN-CHEN LEE**, Health Canada  
*Statistical Strategies for Evaluating the Exposure of Chemical Mixtures during Pregnancy*

Scientists have discovered exposures to industrial chemicals occurring early in life, either in the womb or during early stages of childhood development, could result in a mild alteration, like how the brain develops resulting in changes in attention span, learning ability, behavioural changes, or other impacts like altering where fat cells are deposited in the body or modifying the development of an organ predisposing it to cancer later in life. We develop statistical criteria to evaluate the exposure of chemical mixtures during pregnancy, which are used in risk assessment of chemicals such as mercury, lead, phthalates and persistent organic pollutants.

---

**MICHAEL LI**, McMaster University

*Modeling High-Resolution Animal Telemetry Data: Hidden Markov Models and Extensions*

Clustering time-series data into discrete groups can improve prediction as well as providing insight into the nature of underlying, unobservable states of the system. However, temporal heterogeneity and autocorrelation (persistence) in group occupancy can obscure such signals. We use latent-state and hidden Markov models (HMMs), two standard clustering techniques, to model high-resolution hourly movement data from Florida panthers. Allowing for temporal heterogeneity in transition probabilities, a straightforward but rarely explored model extension, resolves previous HMM modeling issues and clarifies the behavioural patterns of panthers.

---

**JASON LOEPPKY**, University of British Columbia

*Parameterization of the Gaussian Process for Modelling a Blackbox Function*

In this poster we discuss two alternative parameterizations of the Gaussian Process model that are often used in the literature. We begin by discussing the invariance of the two models. We first discuss the lack of invariance and further discuss the lack of a numerical invariance between these two parameterizations. We then discuss the interpretation of parameterizations and show what implications this has in terms of modelling a complex computer code. Additionally we show through a series of examples that one parametrization is typically much better in terms of prediction quality for a set of test data.

---

**BIN LUO**, University of Western Ontario

*Spatial Statistical Tools for Genome-Wide Mutation Thunderhower Detection under a Microarray Probe Sampling System*

In genetics, the study of mutation showers can help better understand mutagenic mechanisms. A cost-effective method is to use an organism specific genotyping array that is designed to detect mutations at defined sites on probes across the entire genome. Mutations at non-probe sites are unobserved. To establish formal statistical tools for genome-wide mutation detection, several test statistics are proposed and are based on the probe array spatial properties. Power performance of the test statistics are evaluated under Neyman-Scott clustering processes via Monte Carlo simulation. Statistics with good performance are recommended as screening tools for geneticists.

---

**ADAM MUIR**, Dalhousie University

*Did a Systematic Review Just Eat Your Citation? The Impact of Reviews on the Citation Frequency of Primary Studies*

Systematic reviews occupy a unique position atop the hierarchy of evidence, and are among the most influential contributions to the literature. Research has established that systematic reviews are frequently cited. However, it is as yet unclear whether reviews create citations within a given field by drawing attention to the primary literature, or merely siphon future citations away from the primary studies they cite. In the current investigation, we examine primary studies cited by Cochrane systematic reviews and explore the rates at which the primary studies are cited before and after the publication of the review.

---

**DREW NEISH**, University of Guelph

*Clustering Using Mixtures of Dirichlet-multinomial Regression Models*

Compositional analysis of the human microbiome is made possible through advanced sequencing techniques, where the output consists of abundance of different bacterial taxa in each microbiome sample. Previously, a Dirichlet-multinomial mixture model has been used for modelling such microbial metagenomic data, where each mixture component represents distinct meta-communities that show similar biota compositions. However, identifying the association of environmental/biological covariates with abundance in different meta-communities remains an important problem. Here, a mixture of Dirichlet-multinomial regression models is proposed and illustrated. These models allow for a probabilistic investigation of the relationship between bacterial abundance and biological/environmental covariates within each inferred meta-community.

---

**AURÉLIEN NICOSIA**, Laval University

*A General Directional Random Walk Model: Application to Animal Movement*

We propose a general directional random walk model to describe the movement of an animal that takes into account features of the environment. A circular-linear process models the direction and distance between two consecutive localizations of the animal. A hidden process structure enables modeling situations where the animal exhibits various movement behaviors. The main originality of the proposed approach is that many environmental targets can be simultaneously included in the directional model. The model is fitted using the EM algorithm. We illustrate its use by modeling the movement of an animal in the Canadian boreal forest.

---

**ANDREW PORTER**, University of Guelph

*Estimating an Experimental Error Variance for Fractional Factorial Designs*

In analyzing fractional factorial designs it can be uncertain which higher order interactions are inactive and thus can be pooled to estimate the experimental error variance. It is even less clear how pooling mean squares that are contaminated by an active effect may affect bias or the Type I error. We present simulation results on the performance of four methods for constructing the estimated experimental error variance when one of the mean squares may be active. Replacing the largest contribution to the error MS with the expectation of the maximum order statistic from a chi-square distribution demonstrates good bias reduction.

---

**SHI QIU**, University of Saskatchewan

*Cross-Validatory Model Comparison and Divergent Region Detection using  $iIS$  and  $iWAIC$  for Disease Mapping*

Two statistical problems arise in using Bayesian hierarchical models for disease mapping. The first is to compare goodness of fit of various models, which can be used to test different hypotheses. The second problem is to identify outlier/divergent regions with unusually high or low residual risk of disease. Leave-one-out cross-validatory (LOOCV) model assessment has been used for these two problems. However, actual LOOCV is time-consuming. This paper introduces two methods, namely  $iIS$  and  $iWAIC$ , for approximating LOOCV with only Markov chain samples simulated from a posterior based on a full data set.

---

**RHONDA ROSYCHUK**, University of Alberta

*An Examination of Spatial Scan Statistics Based on Time to Event Data*

The spatial scan statistic (SSS) is used to identify the geographical clusters of higher number of cases of a disease. It can also identify geographic areas with longer time to events using appropriate distributions. Other authors have proposed Exponential and Weibull distributions for the event times. We propose log-Weibull as an alternative distribution for the SSS and compare the three distributions through simulation to investigate Type I censoring. Methods are also illustrated on time to specialist visit (cardiology or internal medicine) data for discharged patients presenting to Emergency Departments for atrial fibrillation and flutter in Alberta during 1999-2011.

---

**ASANAO SHIMOKAWA**, Tokyo University of Science

*Construction of Tree-Structured Prediction Model Based on Interval-Valued Covariates*

The survival tree, which is constructed using the CART is considered in this study. In our proposed model, covariates can be assumed as interval-valued symbolic variables. The model allows that a concept can be included in several terminal nodes in a tree-structured model. Then, the prediction of a new concept will be constructed by using all terminal nodes based on observed frequency of the concept included in each terminal node. We present the application results of the proposed approach based on covariates obtained by MRI to patients with brain metastases from breast cancer to show the utility of the model.

---

**ANJALI SILVA**, University of Guelph

*Comparative Analysis of Clustering Techniques for RNA-seq Data*

RNA sequencing (RNA-seq) is a deep sequencing-based approach for transcriptome profiling. RNA-seq provides counts of transcripts, offering a method to quantify gene expression. Despite the vast availability of RNA-seq data, interpreting these data in their biological context remains a challenge. Using clustering algorithms, a systematic investigation of relationships between genes can be carried out to identify genes sharing similar expression patterns. A comparative study of three clustering techniques is presented using RNA-seq data obtained from a gene expression study looking at the response of maize to nitrogen limitation. Clusters of genes identified by each method are analyzed for biological significance.

---

**GABRIELLE SIMONEAU**, McGill University

*An Empirical Comparison of Methods to Meta-Analyze Individual Patient Data of Diagnostic Accuracy*

Individual patients data (IPD) meta-analysis has many benefits. In the context of diagnostic accuracy studies, pooled sensitivity and specificity are traditionally reported for a given threshold and meta-analyses are conducted via a bivariate approach. With IPD, it is possible to obtain pooled sensitivity and specificity for each possible threshold. One way to analyze these data is to apply the BREM model at every threshold. Another idea is to analyze results for all thresholds simultaneously, and thus accounting for the within-study correlation. Our aim is to compare two multivariate approaches to the BREM when IPD are available, empirically.

---

**YUBIN SUNG**, University of Guelph

*A Two-Step Method for Genetic Association Analysis with Multiple Longitudinal Traits of Samples of Related Subjects*

We propose a two-step procedure to identify pleiotropic effects on multiple longitudinal traits from a family-based data set. The first step analyzes each longitudinal trait via a three-level mixed-effects model. Random effects at the subject-level and at the family-level measure the subject-specific genetic effects and between-subjects intraclass correlations within families, respectively. The second step performs a simultaneous association test between a single nucleotide polymorphism (SNP) and all subject-specific effects for multiple longitudinal traits. This is performed using a quasi-likelihood scoring method, in which the correlations structure among related subjects is adjusted.

---

**FODE TOUNKARA**, Université Laval

*Archimedean Copulas for Clustered Binary Data*

This presentation shows that Archimedean copulas provide several models to accommodate an extra-binomial variation in Bernoulli experiments. These models feature parameters for the marginal probability of success and a copula dependency parameter. Two applications are presented. First, we construct profile likelihood confidence intervals for the intra-cluster correlation. The second is concerned with the estimation of the closed population size from a mark-recapture study. Unit level covariates are recorded on the units that are captured and copulas are used to model a residual heterogeneity that is not accounted for by covariates. A particular copula model can be selected using the AIC.

---

**DOUGLAS WHITAKER**, University of Florida

*Students' Understanding of Histograms and Bar Graphs: Results from the LOCUS Assessments*

Histograms and bar graphs are standard data displays, and fluency with their use is a requisite for understanding many important statistical topics. These data displays are frequently included in K-12 mathematics curriculum documents in addition to being taught in college-level introductory statistics courses. Including histograms and bar graphs in standards documents does not necessarily imply that students understand them at the expected level. Using data from the LOCUS assessments (NSF DRL-1118168) administered to over 3500 students ages 12-18 in the United States, this poster illustrates the current level of student understanding and common misconceptions about histograms and bar graphs.

---

**DOUGLAS WHITAKER**, University of Florida

*Transition and Collaboration: The Identity of an Advanced Placement Statistics Teacher*

Statistics has a visible and increasing presence in the K-12 curriculum. Despite widespread recognition of statistics as a field independent of mathematics at the tertiary level and beyond, at the K-12 level statistics instruction largely occurs in mathematics classrooms. Without appropriate preparation and professional development, this arrangement can result in an incomplete treatment of statistics. Mathematics teachers charged with teaching statistical content may need to develop a new identity as a statistics teacher to be effective. This poster examines the identity of an in-service secondary level mathematics teacher engaged in collaborative teaching of an Advanced Placement (AP) Statistics course.

---

**KEVIN WILSON**, Dalhousie University

*Out of Cite, Out of Mind? The Effects of Citation by Review Articles on Frequency of Citation in the Epidemiological Literature*

The influence of a piece of academic writing is commonly gauged by the frequency with which it is cited in the literature. Systematic reviews are, unsurprisingly, some of the most frequently cited contributions, and occupy a place atop the hierarchy of evidence. In the present study, we evaluate the effect of review papers (general, systematic, and meta-analyses) on the frequency of citations among primary articles in a cohort of epidemiological studies published in 2005. Using Poisson regression with general estimating equations, we will assess whether reviews affect how often researchers cite primary literature.

---

**LINGYUN YE**, Canadian Centre for Vaccinology

*The Validity of Test-Negative Case-Control Design*

We study a test-negative case-control design (TNCC), which is an extension of the traditional case-control design, in which the study population consists of subjects seeking health care services due to acute respiratory illness. Subsequent laboratory testing is then performed to confirm disease outcome. Due to its simplicity, the TNCC has become the “gold standard” for estimating vaccine effectiveness (VE) for influenza and rotavirus. By modeling the case and control series as independent Poisson processes, we show that the TNCC provides a consistent estimate of VE. The rationale, interpretation and several methodological issues of TNCC designs are also discussed.

---

**BOYKO ZLATEV**, University of Alberta

*Studying Selections of Poems by Statistical Methods*

A sample of selections of P.B. Shelley’s poems is studied by applying various statistical methods. The selections, produced by different editors in the time range from 1840 to 2014, were obtained from published selected editions and anthologies. Multidimensional scaling is applied to both selections and poems, with distances obtained computed by transforming appropriate correlation coefficients. Then cluster and classification analysis is performed on the scaled data. Conclusions about evolution and some non-obvious features of the reception of Shelley’s poetry can be made from the study.