Fall 2015 CO/WY ASA Meeting
Anschutz Medical Campus, ED2 South Building, Room 2305
Friday, October 16, 2015
Time: 12:00 - 3:20 pm (lunch provided)

Agenda:

12:00 - 12:20	**Welcome and lunch**
12:20 - 12:40	**Daniel Yorgov, UCD.** *Augmented Genome-Wide Association Studies of Autoimmune Vitiligo Disorder Identify 25 Novel Chromosome Locations*
12:40 - 1:00	**Michael Bronsert, UCD/AMC**. *A Simulation Study of Coefficient Bias, Stability of Prediction Variables and Individual Risk Estimates Using Stepwise Logistic Regression Analysis for Prediction Models in Very Large Samples*
1:00 - 1:20	**Mikaela Miller, UCD/AMC.** *Profiling the kinetics of transcriptomic responses with RNA-seq*
1:20 - 1:40	**Evan Colvin, UCD Undergraduate**. *The Doctor in the Machine: Machine Learning and Medical Diagnostics*
1:40 - 1:50	**Break**
1:50 - 2:10	**Business meeting**
2:10 - 2:20	**Amelia Greene, D. Young Award Winner.** *Happy, Sad, Curious, Scared: Which Are You? Children's Emotional Response to Musical Chords & Varied Images*
2:20 - 2:40	**James Crooks, NJHealth /AMC**. *Over-fitting in time-series models of air pollution and mortality*
2:40 - 3:00	**Kathryn Colborn, UCD/AMC.** *Statistical methods for handling time-to-response bias and cumulative effects of exposures in longitudinal HIV data*
3:00 – 3:20	**Douglas Lovell, Rogue Wave Software.** *International Competition Aerobatics: Statistical Scoring to Correct for Judge Bias*

Abstracts:

Daniel Yorgov, University of Colorado, Denver. *Augmented Genome-Wide Association Studies of Autoimmune Vitiligo Disorder Identify 25 Novel Chromosome Locations*.

Generalized vitiligo (GV) is an autoimmune disease in which white patches of skin and hair result from destruction of melanocytes. In previous genome-wide linkage and association studies (GWAS), our group identified 27 susceptibility chromosome locations for generalized vitiligo. To identify additional chromosome locations with smaller effects or with lower risk allele frequencies, we have carried out a third vitiligo GWAS, augmented the two previous GWASes with additional controls to enhance power, carried out genome-wide imputation and meta-analysis of all three vitiligo GWAS, and performed an independent replication study. The

combined analyses, with 4,680 vitiligo cases and 39,586 controls, identified 25 novel significant replicated chromosome locations. Many of these locations have also been associated with other autoimmune diseases, whereas others are specific to vitiligo. The new findings provide framework for understanding vitiligo pathobiology, and perhaps offer novel targets for vitiligo treatment. In this talk I will give details about the genotype imputation and the statistical analyses performed, present some of the results and outline current work on building a polygenic risk model. Our long-term goal is optimized prediction of vitiligo risks to facilitate clinical application of optimal therapies possibly based on genetic subtyping of the disease.

Michael Bronsert, University of Colorado, Denver, Anschutz Medical Campus. *A Simulation Study of Coefficient Bias, Stability of Prediction Variables and Individual Risk Estimates Using Stepwise Logistic Regression Analysis for Prediction Models in Very Large Samples*.

Prediction models are widely popular. In our motivation study, a prediction model for postoperative complications using the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) database, the particular challenge is to achieve a parsimonious predictor model. Stepwise selection methods are widely utilized to identify a parsimonious set of predictors, but have met criticism in the literature. Simulation studies have shown that automated selection methods can result in substantial amount of bias in estimation of regression coefficients and that random variability in a dataset can result in a different set of predictors being selected producing unstable models. The former bias is more apparent in smaller datasets with few events per variable (EPV). Little is known about the level of bias in very large datasets and the effect model stability has on individual risk predictions.

We investigated the performance of prediction models built through automated selection methods in very large databases. We examined four evaluation criteria: 1) relative bias and precision of estimated regression coefficients, 2) the proportion of times that a predictor is selected, 3) the accuracy of individual risk predictions and 4) relative bias of the area under the receiver operating characteristic curve (AUC). These evaluation criteria were calculated from multiple bootstrap samples selected with replacement from 2,275,240 surgical cases in the ACS NSQIP database. In addition, we simulated multiple samples with varying EPV and numbers of available predictors along with varying alpha levels for entering and exiting the models.

Mikaela Miller, University of Colorado, Denver, Anschutz Medical Campus. *Profiling the kinetics of transcriptomic responses with RNA-seq.*

Analysis of gene expression time course data typically involves identifying differentially expressed genes, clustering those genes based on their temporal profiles, and inferring regulatory networks and interactions from those clusters. As next-generation sequencing technology becomes more widespread, time course RNA-seq experiments will become more prevalent. In RNA-seq experiments, sample DNA is fragmented into millions of short reads that can be sequenced accurately and in parallel, and then mapped to a reference genome. Experimental data consists of counts of reads mapped to a given gene in the genome, approximating the abundance

of the target transcript and thus gene expression. This presentation reviews approaches for performing differential gene expression with RNA-seq time course data, including generalized linear models that accommodate the overdispersed count data and non-parametric methods. The methods are applied to RNA-seq data collected from a mouse model of prolonged cigarette smoke (CS) exposure sufficient to cause emphysema.

Evan Colvin, University of Colorado, Denver, Undergraduate. *The Doctor in the Machine: Machine Learning and Medical Diagnostics*.

Machine learning algorithms seek to make accurate predictions. The ability to quickly understand the models comes second to their clinical or economic applications. I will present the Naïve Bayes machine learning algorithm and use it to build a model to classify tumors as either malignant or benign based on the interactions of the tumors' features. Naïve Bayes makes the unrealistic assumption all the variables in a data set are uncorrelated with each other—a concern for researchers—yet it still can make accurate predictions—to the satisfaction of diagnosticians. I will conclude my talk by briefly introducing other, more powerful and more complicated machine learning algorithms, and explore how computers may aid the diagnostician in the future.

Amelia Greene, David Young Award Winner, The Classical Academy. *Happy, Sad, Curious, Scared … Which Are You? Children's Emotional Response to Musical Chords & Varied Images.*

David Young Award winner Amelia Greene will talk about a few of statistical issues related to her science project. The following abstract describes her project.

The purpose of this project was to determine whether the interaction between musical chords and visual images can cause a change in children's emotions using a two-dimensional rating scale. I hypothesized that if musical chords (major, minor, augmented, diminished) and visual images (black and white, color) are varied, then children's emotions (as reflected on a 2D rating scale) will be 1) happiest when hearing a major chord and seeing a color image; 2) most sad when hearing a minor chord and seeing a black and white image; 3) most curious when hearing an augmented chord and seeing a colored image; and 4) most scared when hearing a diminished chord and seeing a black and white image.

Hypothesis t-tests were conducted to statistically confirm children's emotional response to color versus black and white images when hearing a musical chord (major, minor, augmented, and diminished). Ninety-three junior high students from my school participated in my experiment. The test subjects were given 34 randomly selected and randomly ordered options from four different musical chords (major, minor, augmented, and diminished) and twelve different images. There were six images (table, clouds, boys, baby, cat, and woods) presented in either color or black & white. The children filled out a 2D rating scale created specifically for this experiment about how each of the options of chords and images made them feel. The x-axis was a sad-happy rating scale and the y-axis was a scared-curious rating scale. The 2D rating scale was converted to numbers from -2 to 2 so that mathematical analysis could be accomplished. Data analysis previous to conducting the hypothesis t-tests included calculating the mean,

median, mode, range, random error, 90% confidence intervals, pie charts, F-tests, and two-way analysis of variance (ANOVA). The hypothesis t-tests were run in DDXL which is an Excel add-on which comes on a disk with the textbook Elementary Statistics Using Excel by Mario F. Triola. Even though the sample size n was greater than 40, hypothesis t-tests were conducted instead of hypothesis z-tests because the population standard deviation σ was unknown. DDXL computed the p-value for each hypothesis t-test. Conclusions were based on the p¬-value. This higher level of confidence is why the hypothesis t-tests were extremely valuable. For some hypothesis t-tests conducted, the level of confidence was greater than 99.99%! The hypothesis t-tests were more useful that F-tests and ANOVA because the t-tests were more easily interpreted for determining differences. F-tests and ANOVA would conclude that at least one item was different from all the others but not which item or items were indeed different.

The data collected supports the hypothesis for happiness and sadness, does not support the hypothesis for curiosity, and partially supports the hypothesis for scariness. The data suggest that the children could differentiate the major and minor chords with emotions, but had a hard time associating diminished and augmented chords with emotions. Thank you for selecting "Happy, Sad, Curious, Scared – Which Are You?" for the David Young Award.

James Crooks, National Jewish Health /Anschutz Medical Campus. *Over-fitting in time-series models of air pollution and mortality.*

Time series studies of air pollution and mortality have long been the workhorse of air pollution epidemiology, providing estimates of excess mortalities due to short-term changes in exposure. While there is consensus on the necessity of controlling for temperature and for medium-to-long-term mortality trends, there remain important differences regarding how these confounders should be represented and which other confounders (if any) to include. Because some formulations require many more parameters than others, much of the debate focuses on the issue of over- or under-fitting. Standard information criteria (AIC, BIC) do not adequately address this issue. We present preliminary results from a large-scale study comparing models on the basis of empirical predictive ability, which penalizes both under- and over-fit using the data itself rather than heuristics or asymptotic results. Specifically, we performed 10-fold cross-validation on a set of single-pollutant time-series models encompassing 11 model formulations, 6 pollutants, and 4 mortality endpoints for each of 90+ American cities.

Kathryn Colborn, University of Colorado, Denver, Anschutz Medical Campus. *Statistical methods for handling time-to-response bias and cumulative effects of exposures in longitudinal HIV data.*

This study utilized routinely collected health care facility data from over 140,000 HIV positive patients over 10 years in Mozambique. Treatment interruptions (TI), drop-out and death were common. A TI is defined as a lapse in a patient's adherence to treatment. The primary goal of this study was to determine the effects of the number and duration of TIs experienced over a person's follow-up on their risk of drop-out or death.

These data posed many challenges. First, the longer a patient survived the more chances he/she had to experience a TI. Second, for patients that either dropped out or died fairly quickly after starting ART, they did not have enough survival time to experience a TI – commonly referred to as time-to-response bias. Finally, the longer patients interrupted their treatment, the fewer chances they had to re-interrupt. To correct for these biases, we first limited our sample to those patients that were on ART for at least one year, giving them enough time to experience a TI. Second, by calculating annualized rates, we accounted for differences in total TIs and TI durations that were due to differences in total survival times. Additionally, we fit a time-dependent Cox proportional hazards model and calculated the cumulative effects of TI annualized rate and TI duration, which permitted us to account for compounding risk due to multiple TIs and longer time spent not taking drugs.

At the database closing date, 67% of patients were alive, 3% had died and 30% were LTFU. Patients that accumulated at least one TI per year on ART increased their rate of permanent LTFU or death by more than 60% compared to those with less than one TI per year. Accumulation of days of treatment interruption up to 50% of total time on ART was associated with a 20% increased rate of LTFU or death over those that did not interrupt their treatment. For those with more than 50% of their ART time interrupted, their rate of LTFU or death was almost 30% higher than those that did not interrupt their treatment.

Douglas Lovell, Rogue Wave Software. *International Competition Aerobatics: Statistical Scoring to Correct for Judge Bias*.

Every year, sixty or so of the world's top aerobatic pilots face off in competition for European, American, and world titles in aerobatics. This year in France, 58 pilots from eight nations competed for the world championship title. Estimated direct costs of participation for those pilots would be about $1.2 million. Including training and equipment, the investment on the part of pilots and sponsoring nations is upward from $30 million.

Competition aerobatics is a judged sport, much like figure skating. CIVA, the world organization sponsoring and regulating the sport, uses an unusual statistical system for converting judge marks into results. There is nothing like it in any other sport. This presentation will unwrap the CIVA results system, labeled "Fair Play," and expose it to your critical analysis.

Douglas Lovell is an aerobatic competitor and judge, member of the Board of Directors of the International Aerobatic Club (IAC), keeper of records and statistics for the IAC, and a software engineer with sixteen years at IBM Research Division, now at Rogue Wave Software in Louisville, Colorado.