<table>
<thead>
<tr>
<th>Presenter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camille Hochheimer</td>
<td>2</td>
</tr>
<tr>
<td>Kingston Kang</td>
<td>3</td>
</tr>
<tr>
<td>Paul Hargartan</td>
<td>4</td>
</tr>
<tr>
<td>Jun He</td>
<td>5</td>
</tr>
<tr>
<td>Brian S. Di Pace</td>
<td>6</td>
</tr>
<tr>
<td>Hangcheng Liu</td>
<td>7</td>
</tr>
<tr>
<td>John Stansfield</td>
<td>8</td>
</tr>
<tr>
<td>Keith Zirkle</td>
<td>9</td>
</tr>
<tr>
<td>Narad Mishra</td>
<td>10</td>
</tr>
<tr>
<td>Victoria Garcia</td>
<td>11</td>
</tr>
<tr>
<td>Jin Liu</td>
<td>12</td>
</tr>
<tr>
<td>Anny-Claude Joseph</td>
<td>13</td>
</tr>
<tr>
<td>Amanda Gentry</td>
<td>14</td>
</tr>
<tr>
<td>Alicia Johns</td>
<td>15</td>
</tr>
<tr>
<td>Kellen Cresswell</td>
<td>16</td>
</tr>
<tr>
<td>Erin Donahue</td>
<td>17</td>
</tr>
<tr>
<td>Jonathan W. Yu</td>
<td>18</td>
</tr>
<tr>
<td>Kate Stromberg</td>
<td>19</td>
</tr>
<tr>
<td>Sprio Stilianoudakis</td>
<td>20</td>
</tr>
<tr>
<td>Viviana Alejandra Rodriguez</td>
<td>21</td>
</tr>
<tr>
<td>Christine Orndahl</td>
<td>22</td>
</tr>
<tr>
<td>Keighly Bradbrook</td>
<td>23</td>
</tr>
</tbody>
</table>
Title: Estimation of Survey Attrition Phases: Empirical and Change-Point Models
Presenter: Camille Hochheimer
Advisor: Roy Sabo
Abstract:

Attrition or dropout is a common phenomenon in surveys and questionnaires. One theory is that attrition occurs in three distinct types of phases: a curiosity plateau, a dropout phase, and then stable participation. Detection of these phases is especially important as it may suggest non-ignorable missingness. Our overarching goal is to find a method to test the null hypothesis of no distinct attrition phases (linear attrition) against the alternative hypothesis that distinct attrition phases exist. We demonstrate the shortcomings of empirical methods, specifically a user specified method, a generalized linear mixed model, and discrete time survival analysis, at detecting phase transitions through a simulation study. We then introduce both parametric and non-parametric hazard change-point models and explore their utility at detecting attrition phases. Future extensions are discussed.
Title: *Estimating the Respiratory Lung Motion Model Using Tensor Decomposition Derived from Multipatient 5-Dimensional Displacement Vector Field for Lung Cancer Radiation Therapy*

**Presenter:** Kingston Kang  
**Advisor:** Nitai Mukhopadhyay, Ph.D.

Radiation treatment is often delivered to parts of our anatomy under constant motion, lung tumors being the most common of all. Common courses of dealing with this motion include the delivery of gated radiation, breath holding, modification of the posture, or using image guidance in real time. However, even with all those interventions, uncertainty exists in the actual dose delivered, which is a function of the planned dose and the motion of the anatomy during the treatment. Several approaches to quantifying this uncertainty propose using a model to formulate the motion through a mathematical function over time. Li, Lewis et al (2011) use principal component analysis (PCA) to propose one such model using each image as a very long vector. However, the images come in a multidimensional array, and vectorizing them removes the spatial location information from the voxels. Population Value Decomposition (PVD) (Crainiceanu et al (2011)) is a new methodology proposed primarily to reduce the dimensionality of 2D images and combine multiple images to make population inference. We propose to generalize the PVD algorithm for 5D displacement vector field (DVF) to achieve dimension and noise reduction and build a respiratory lung motion model. Such a model can be used in estimating the actual dose delivered based on the planned dose and the organ motion. Our primary application to lung DVF shows better accuracy compared to the one-step principal component method. The total number of squared errors, calculated as the squared difference between the observed value and estimated value for each voxel, above 0.15cm achieved more than 70% reduction compared to the PCA method. Moreover, our algorithm can be extended to higher dimensional arrays and accomplish the model building within a reasonably short amount of CPU time.
Title: Assessing the Effect of Uncertainty from Single Imputation of Analytes below the Detection Limit on Chemical Mixture Analysis

Presenter: Paul Hargarten
Advisor: David Wheeler, PhD

Abstract: Many epidemiological studies attempt to find associations between chemical exposures and a disease of interest. In reality, individuals are exposed to many chemicals simultaneously that may increase risk of disease. Weighted quantile sum (WQS) regression is an approach for modeling simultaneous exposure to a large number of correlated chemicals and identifying the chemicals that are associated with disease. A complication in chemical mixture analysis is that experimental apparatus can only detect the levels of a chemical present in dust to the limit of detection. The current practice of mixture analysis with WQS is to impute concentrations below the detection limit (BDL) or to place BDL observations in the first quantile in the weighted index. However, the impact of the uncertainty associated with BDLs in WQS has not been previously studied. In this research, we developed a univariate Bayesian single imputation approach to estimate concentrations BDL and conducted a simulation study to examine the effect of various percentages of BDLs in a chemical mixture analysis using WQS. The mixture consisted of 14 chemicals with four assigned relatively high weight as important chemicals. Simulation study scenarios included BDL levels of 10%, 33%, 50%, and 80% per chemical and weak and strong health effects with a binary outcome. Future work will extend this approach to multiple chemicals and multiple draws so that the complex correlation between chemical concentrations will be used in developing a more informed imputation model of the BDLs.
Title: Misclassification Effect on SMART Design: Power of comparison between DTRs
Presenter: Jun He
Advisors: Donna McClish, Roy Sabo
Abstract:

Introduction: SMART designs are sequential multiple assignment randomized trials that provide the opportunity for patients to switch treatments during the trial. Misclassification based on patients’ initial treatment response results in patients being assigned to the wrong treatment. A set of decision rules used to choose between effective treatments for individual patients is called dynamic treatment regimens (DTRs). Our interest is to study the effect of misclassification on comparing DTRs.

Methods: Formulas for test statistics and power were derived that account for the possibility of misclassification, which were then numerically evaluated under various parameter templates. Our general assumptions for SMART designs were continuous response, equal variances for misclassified and non-misclassified groups, and 1:1 randomization. Medium and large effect sizes were studied under 80% power.

Results: Generally, the power was affected by sensitivity, specificity, true response rate, and effect sizes; in some cases, the relationship was not monotonic and non-intuitive. The effect of misclassification on power depended on which DTRs were being compared. Sensitivity had more impact on power than specificity if true response rate and the effect size of responders were high, while specificity had more impact than specificity on power if true non-response rate and the effect size of non-responders were high. Large effect size had a broader range of power than medium effect size.

Conclusion: Misclassification in SMART designs can lead to under-powered comparison of DTRs. Knowing the power of DTRs comparisons under misclassification allows us to calculate adjusted sample size to satisfy the power. Those findings also suggest we look for new methods to analyze data based on misclassification.
Title: *A Site-Adjusted Approach to the Covariate-Adjusted Response-Adaptive Allocation Design in Multi-Center Trials*

Presenter: Brian S. Di Pace

Advisor: Roy T. Sabo

Abstract:

Response-Adaptive (RA) designs, used to allocate patients in clinical trials, have been generalized to the Covariate-Adjusted Response-Adaptive (CARA) design, which balances treatment assignment among a set of covariates while maintaining features of the adaptive design. Challenges may arise in multi-center trials if differential treatment responses and/or effects among sites exist. We first show how allocation probabilities and treatment effectiveness can be adversely affected in balanced and RA designs when multiple sites are not accounted for. We then propose the Site-Adjusted Response-Adaptive (SARA) approach to account for inter-center variability in treatment effectiveness, including both site and site-by-treatment interaction random effects to calculate conditional probabilities. These success probabilities are used to update assignment probabilities for allocating patients between treatment groups as subjects are accrued. Both frequentist and Bayesian models are considered. We compare the balanced and RA cases with our proposed designs and show that the variability in treatment effectiveness is reduced when accounting for clustering during randomization.
Title: Examining the Vanishing Tetrad Number as an Index of the Complexity of SEM Models
Presenter: Hangcheng Liu
Advisor: Dr. Robert Perera
Abstract:

Structural Equation Modeling (SEM) is a series of statistical methods that allow complex relationships between one or more independent variables and one or more dependent variables, which is widely used in the behavioral sciences (e.g. psychology, psychobiology, sociology). Model complexity is defined as a model’s average ability to fit different data patterns and it plays an important criteria to do model selection. Like linear regression, the number of free model parameters (q) is often used in traditional SEM model fit indices as a measure of the model complexity. However, only using q to indicate SEM model complexity is crude since other contributing factors, such as types of constraints or functional form are ignored. To solve this problem, we propose a special technique, Confirmatory Tetrad Analysis (CTA) to be a complement of traditional methods to test the model fit. The purpose of this study is to examine whether SEM model Complexity is related to the number of vanishing tetrads implied in the models.
Title: HiCcompare: Methods for joint normalization and difference detection for Hi-C data
Presenter: John Stansfield
Advisor: Dr. Mikhail Dozmorov
Abstract:

The advent of Hi-C sequencing has allowed new insights into how the genome behaves in regards to its spatial organization. Changes in chromosomal organization are emerging as a mechanism of gene regulation. Additionally, disruptions in the 3D organization of the genome are an emerging hallmark of cancer. There are very few methods for detecting differences in 3D structure between datasets. We developed an R package, HiCcompare, which provides methods for the joint normalization and comparison of two Hi-C datasets. HiCcompare’s methods are based on what we term the MD plot (distance vs. difference plot) which allows for the visualization of distance-centric differences between interacting chromatin regions. The normalization methods use loess regression to adjust for biases between datasets. A permutation based method is then used to detect differential chromatin interactions. HiCcompare was tested on Hi-C data from the RWPE1 prostate epithelial cell line and a cancer derivative with overexpression of the ERG gene. Differentially interacting regions were detected and compared to regions detected in a different study of the same data using one of the few other packages for Hi-C comparison, diffHic. The regions detected by diffHic tended to have lower fold changes and occurred at shorter genomic distances while the regions detected by our methods had exhibited higher fold changes and covered the range of genomic distances. HiCcompare performs better than many other methods for normalization of a single Hi-C dataset and additionally provides one of the very few methods for comparative analysis of the 3D structure of different cellular conditions.
Title: Estimating Causal Effects of Air Quality on Non-Hodgkin Lymphoma When Interference Is Present
Presenter: Keith W. Zirkle
Advisor: David Wheeler
Abstract:

Non-Hodgkin lymphoma (NHL) is the seventh most common cancer in the U.S. Several studies have found evidence of an association between NHL and air pollution. Particulate matter (PM) are the microscopic solid and liquid materials found in air and are considered Group 1 carcinogens. In 2005, the U.S. Environmental Protection Agency (EPA) designated certain U.S. counties and tribes as “nonattainment” based on standards set for PM with aerodynamic diameter < 2.5μm (PM$_{2.5}$). In our study, we estimate the causal effects of nonattainment designation on NHL incidence in U.S. counties in California, Kentucky, and Georgia using Surveillance, Epidemiology, and End Results (SEER) data. Traditional causal inference assumes no interference, or a subject’s outcome is not affected by other subjects’ treatments. In air pollution studies, regulation at one location will affect downwind locations. We introduce a new assumption that allows causal inference under interference by identifying covariates X necessary to model the treatment assignment and spillover mechanism. We compare three models in a Bayesian framework. In the first model, we perform naïve outcome regression that adjusts fully for X and uses all data (N=337). In the second model, we identify comparable treated and control counties according to propensity score overlap and model using the pruned dataset (N=276). In the third model, we match counties by their propensity score (N=26). Across the models, we found a consistent protective effect from nonattainment designation, but non-significant spillover effects. The effect may be explained away when a spatial random effect is included.
Title: How Long Do Kidney Transplant Last?
Presenter: Narad Mishra
Advisor: Bob Carrico

Abstract:

End-stage renal disease is treated with conventional medical therapy that renders patients to lower quality of life as they experience dialysis regimens for their disease. Kidney transplantation allows the patient to not be restrained to that therapy and to experience a longer survival after diagnosis. The longevity of kidney transplant depends on various factors that are associated with patients and donors’ characteristics. The Purpose of this study is to calculate the graft and patient survival of living vs deceased donors, and to develop a model to predict how long kidney transplants last and to show survival patterns in historical data for important characteristics to transplant recipients. Kaplan Meier estimates were calculated to display the difference of graft survival between deceased and living donors, and a model was built using cox hazard regression model building focusing on only transplants using kidneys from deceased donors. The analysis showed that the survival of patients and graft survival with living donor was much better than the survival of patients with deceased donor (P-value <0.001). The twenty-year patient survival estimate for living donor transplants was 39.4% and for deceased donor transplants it was 20.9%, and the graft survival estimate was 24.1% and 14.5% for living donor and deceased donor respectively. Modeling the deceased donor transplant survival using Cox Regression, it was found that age of both donor and recipient, if Dialysis was performed within first week of transplant (delayed graft function), and recipient BMI were some of the most significant factors in determining the longevity of the allograft.
Title: The effects of misspecifying the optimal response-adaptive allocation ratio in adaptive randomized clinical trials
Presenter: Victoria Garcia
Advisor: Adam Sima
Abstract:

Existing methods for determining an optimal response-adaptive allocation (ORAA) ratio based on some observed response data seem to labor under the implicit assumption that the derived allocation ratio is the correct one for those observed response data. Absent from the literature, however, are details regarding performance when a given design is not appropriate for the observed response data at any point throughout the duration of the trial. Therefore, a simulation of 100 hypothetical two-arm trials using an ORAA according to Biswas, Bhattacharya, and Zhang (2007) was conducted, where the target response distributions were normal (N=220) or gamma (N=150). Powered for an effect size of 0.3, the mean response to treatment A was 13.05, 15 for treatment B, and the standard deviation (SD) for both groups was 6.5. Smaller responses were deemed more desirable and the target allocation to treatment A (the more favorable treatment) was 0.75. Along with the target distributions, four additional distributions (double exponential, uniform, two-parameter exponential, and non-central t) were examined under an anticipated and a null scenario. Plots of mean allocation probabilities to treatment A per each scenario per each target distribution were generated, along with plots of their SDs. For the normal target, in the anticipated scenario, only the two-parameter exponential distribution over-allocated subjects to treatment A, while, in the null scenario, only the gamma distribution under-allocated. For the gamma target, in the anticipated scenario, all comparison distributions under-allocated subjects to treatment A. In its null scenario counterpart, only the two-parameter exponential distribution over-allocated subjects. Allocating subjects according to a misspecified allocation ratio leads to different study characteristics than anticipated.
Title: A frailty model for non-susceptible clustered current status data with informative cluster size
Presenter: Jin Liu
Advisor: Dipankar Bandyopadhyay
Abstract:

Current status (survival) data, a form of interval-censored data, abounds in the fields of epidemiology and public health, where a subject (or a study unit) at risk for an event of interest is only monitored at an inspection time, with an indicator denoting whether the event occurred. Analyzing these type of data poses various interesting statistical challenges. Our motivating dataset in this project arises from a periodontal disease (PD) study, where the current-status event times (corresponding to the event of extreme PD) that are recorded for tooth-sites are clustered within subjects. In addition, this type of dataset exhibits excessively heavy levels of censoring, given that a proportion of tooth-sites may never experience the event of interest, and can be hypothesized as non-susceptible, or cured. Furthermore, the number of available tooth-sites for a subject can be informative of the overall periodontal health of that subject, and ignoring that may lead to imprecise parameter estimates. To mitigate this under a unified framework, we develop a joint (shared parameter) model for the site-level current status response and the subject-level count response of available tooth-sites. The current status model follows a semi-parametric generalized odds-rate (GOR) mixture cure regression model, while the count responses follows a Negative Binomial regression, with the two connected by a random intercept. Conditional on this random term, the two responses are independent (the theoretical basis of the shared parameter framework), and the estimate of the random term throws light on the positive/negative association between the latent disease status and the cluster size. Our semi-parametric GOR regression model with piece-wise constant baseline hazards is a relatively underexplored class of survival models, having the popular proportional hazards and proportional odds models as sub-classes. Our approach is classical, and the estimation technique follows full maximum likelihood using adaptive Gaussian quadrature, powered by the SAS NLMIXED tool (SAS Institute, Inc., Cary, NC). Post model building, we illustrate our methodology through our motivating data by picking up the best model via standard AIC/BIC techniques, followed by relevant interpretation of covariate effects. Future work will involve model diagnosis and validation via extensive simulation studies.
Title: *The Timing of Geographic Power*

Presenter: Anny-Claude Joseph  
Advisor: David C. Wheeler Ph.D

Abstract:

In many studies on the spatial risk of disease, investigators search for a spatial signal of elevated risk using geographic locations at the time of disease diagnosis in hopes of identifying new risk factors. However, studies often fail to find a significant spatial signal. This is likely due in part to not looking in the right place at the appropriate time. Environmental exposures related to cancer risk are intrinsically temporal and many cancers have a long latency. When these factors are considered in conjunction with a mobile population, it is likely that the spatial signal related to relevant historic environmental exposures is obscured.

To investigate this hypothesis, we conducted simulation studies to characterize the effect of population mobility on the ability of generalized additive models to detect areas of significantly elevated historic environmental exposure. We generated data based on the residential histories of participants in the National Cancer Institute Surveillance, Epidemiology, and End Results non-Hodgkin lymphoma study, and varied the duration and intensity of the environmental exposure. Results showed that the probability of detection, mean spatial sensitivity, and mean spatial specificity of models decreased steadily as the time prior to study enrollment increased. This suggests that spatial areas of high-intensity exposure of relatively short duration will be difficult to detect over time when using residential locations at the time of diagnosis in mobile study populations.
Title: Mixed models for ordinal outcomes in twin and sibling studies with high-dimensional covariate spaces  
Presenter: Amanda Elswick Gentry  
Advisor: Kellie Archer  
Abstract:

The Brisbane Longitudinal Twin Study (BLTS) was being conducted in Australia and was funded by the US National Institute on Drug Abuse (NIDA). Adolescent twins and their non-twin siblings were sampled as a part of this study. We are analyzing a subset of this data that includes demographics, cannabis use metrics, personality measures, and imputed genotypes for 8,572,909 single nucleotide polymorphisms (SNPs) for 1,307 patients. Our primary goal is to determine what combination of SNPs and additional covariates may predict cannabis use, measured on an ordinal scale as: “never tried,” “used moderately,” or “used frequently”. To conduct this analysis, we have extended the ordinal Generalized Monotone Incremental Forward Stagewise (GMIFS) method for mixed models. Our mixed model includes a random intercept term for each family with correlations between family members assigned according to a pre-specified kinship matrix. The variance of the random intercept term includes both the specified kinship matrix and an additional term estimated by the model; this additional term estimates variability due to environment. Since the data are high-dimensional, (the number of covariates is much greater than the sample size) the GMIFS procedure is used to achieve a parsimonious model. The penalization implemented by GMIFS allows only those SNPs and additional covariates that contribute significantly to the outcome to enter the model. This modeling scheme accounts for genetic and environmental components of variance and efficiently chooses a subset of covariates from a high-dimensional space that most accurately predict the ordinal outcome.
Title: Clinical Outcomes From a Pharmacist-Physician Team-Based Care Model  
Presenter: Alicia Johns  
Advisor: Robert Perera  
Abstract:

The Carilion Clinic Improving Health of at Risk Rural Patients (IHARP) study implemented a pharmacist-physician team-based care model to deliver a comprehensive medication management plan for patients in the southwest region of Virginia. The primary goal of the program was to provide better disease control for patients with multiple chronic conditions. Clinical outcomes including hemoglobin A1c, systolic and diastolic blood pressures, LDL, and total cholesterol were collected from patient electronic medical records for both the group that underwent the care model and a propensity score matched comparator group. A generalized estimating equation (GEE) model was used to assess the differences between the two groups from the pre- to post-care period adjusting for covariates. Covariates included sex, age, race, primary insurance, and number of chronic conditions. The GEE model was also applied to individuals whose clinical values were above goal values to determine if the care model was effective in that subset of individuals. Results from this analysis will allow medical providers to identify how a care-based model can potentially affect their patients’ outcomes.
Title: A Proposal on Computing Equivalent Dose in Radiation Therapy with Accommodation for Spatial Heterogeneity
Presenter: Kellen Cresswell
Advisor: Nitai Mukhopadhyay
Abstract:

Radiation therapy is prescribed with multiple parameters to allow maximal dose to the tumor and minimal dose to normal tissue. However, the final delivery plan is a very heterogeneous distribution of radiation dose across the entire treatment volume. This makes it difficult to study the effect of dose on outcomes. Traditionally, the dose is summarized into a single number, named Biologically-effective dose (BED) before any further inference on dose and outcome is done. It is currently the most widely used method and is said to reflect the true biological dose delivered to a patient. Studies have found that negative side-effects can still arise in patients who have received a BED deemed safe. One prominent issue with current methods is the failure to account for complex spatial relationships between dose intensities. This project seeks to develop a new method of determining safe dosing which takes these spatial relationships into account. A range of possible solutions were developed and applied to simulated 2-dimensional datasets and, when possible, real 3D datasets. Preliminary work has shown that while certain algorithms can successfully differentiate between different 2D dosage concentrations, issues arise when scaling these algorithms to analyze 3D data. These issues, and their potential solutions, will be discussed in detail.
Title: Natural lead-in approach to covariate-adjusted response-adaptive allocation
Presenter: Erin Donahue
Advisor: Dr. Roy Sabo
Abstract:

Response-adaptive (RA) clinical trial designs aim to maximize treatment successes compared to equal allocation designs, while covariate-adjusted response-adaptive (CARA) schemes aim to maximize treatment successes conditional on a set of patient characteristics. For both designs adaptation is often delayed until sufficient subjects are accrued to overcome small-sample irregularities, which reduces the realized benefits from adaptation. Alternatively, natural lead-in periods can be implemented through the use of decreasingly informative prior elicitation in the Bayesian framework. The decreasingly informative prior is a function of the sample size such that allocation weights are constrained early in the trial, but where adaptation increases as more subjects are accrued. While already implemented for RA allocation, we extend this approach to CARA allocation assuming a continuous response, continuous covariate, and two treatment groups. Simulation studies are used to estimate the behavior of this approach in several scenarios in which the effect size, covariate effect, and interaction presence and direction are varied. The results show that as effect size increases, allocation weights increase toward the better treatment. Further, the presence of a covariate can have a synergistic, antagonistic or diffusing effect on allocation weights, depending on the nature of the main and interaction effects.
Title: Propensity-adjusted semi-parametric survival analysis under heavy censoring for clustered kidney transplant data
Presenter: Jonathan W. Yu
Advisors: Dr. Dipankar Bandyopadhyay and Dr. Le Kang
Abstract:

In situations where randomized control trials (RCT) cannot be conceptualized (for example, in large observational studies with survival endpoints), various propensity score (PS) techniques are popularly used to control for pretreatment confoundings in baseline characteristics. However, this inferential setup of time-to-event outcomes can be further plagued with a myriad of complexities, such as heavy censoring (sometimes due to non-susceptibility of study units to a particular disease), high imbalance between comparator (treatment) groups, and clustered nature of the data (survival outcomes appearing as groups). Furthermore, there exists a plethora of PS techniques without any clear consensus on the choice of the covariables for the propensity adjustment. The motivating data is derived from the United Network of Organ Sharing (UNOS) database, which monitors survival behavior of patients undergoing kidney (and various other organ) transplantations. Our goal in this project is to develop a unified modeling approach to address the aforementioned complexities that are observed in the UNOS database and derive precise parameter estimates. We address this via the following steps. First, we estimate the propensity scores for the heavily imbalanced multiple comparator groups via the generalized boosted regression model. Next, these estimated scores were fed into a semi-parametric Cox-proportional hazards modeling framework, adjusted for multi-center clustering (via a parametric frailty specification) and excess censoring (via cure-rates), via the popular inverse probability of treatment weighting (IPTW) approach. This approach provides the necessary balance between the groups, converts the observational database into a RCT, thereby providing appropriate risk estimates. Future goals include robustifying the current setup via introducing efficient IPTWs, adjusting for informative cluster sizes (which is an additional complexity of this type of databases), and simulating extensive studies under various model misspecifications.
Title: Adjusting for Dropout in Wait List Control Clinical Trials
Presenter: Kate Stromberg
Advisor: Adam Sima

Abstract:

In waitlist control (WLC) clinical trials, the treatment group receives the intervention upon being admitted into the study and the waitlist control group receives the intervention after a certain amount of time. In an unpublished waitlist control design, Sima et al. use the distribution of the treatment duration of the treatment group to determine the wait time for the WLC group. However, there is a disconnect in the distributions between the treatment and WLC groups due to dropout in the treatment group skewing the distribution to appear to have longer treatment times. The proposed method uses the rate that patients are receiving the intervention to estimate the probability that they complete the intervention. This value is used as a sampling weight for each patient for the calculation of the distribution. A simulation study was run using varying amounts of dropout and the percentage of WLC participants having a shorter wait time was calculated. The results show that, by weighting the distribution, the WLC group received intervention earlier than if dropout was not considered.
Title: Developing a methylation-based survival analysis for patients with glioblastoma
Presenter: Spiro Stilianoudakis
Advisor: Mikhail Dozmorov

Abstract:

DNA methylation is an epigenetic mechanism that plays a critical role in the regulation of gene expression. Methylation levels are measured using CpG (cytosine-phosphate guanine) sites that are typically present in the promoter regions of genes. Increased methylation (known as hypermethylation) is generally considered to suppress gene expression and is a prevalent feature of abnormally silenced genes in cancers. In contrast, global hypomethylation plays a role in inducing genomic instability by resulting in an increase in transcription and an elevated mutation rate due to mitotic recombination. However, little is known as to whether these changes in methylation are associated with survival from various cancers. Clinical and methylation data has been obtained from the Infinium HumanMethylation27k platform for 16,502 CpG sites in 280 patients with glioblastoma. A Cox proportional hazards model was performed in order to determine those CpG sites that were significantly associated with survival, while also adjusting for clinical information such as age, gender, and race. Survival-associated CpG sites were characterized by their epigenetic environment and relation to gene-centric annotations using different machine learning techniques. Out of the 16,502 CpG sites, a total of 875 were found to be significantly associated with survival. It was found that out of the 92 original genome annotation features, only a handful were important. The location of CpG sites with respect to the transcription start site of the closest associated gene was proven to be the most significant attribute of survival-associated CpGs.
Title: Longitudinal Analysis of Energy Expenditure in Cold-induced Thermogenesis Via a Piecewise Polynomial Curve
Presenter: Viviana Alejandra Rodríguez
Advisors: Dr. Shanshan Chen, Dr. Yongyun Shin
Abstract:

Background & Aim: Energy expenditure is a vital component in body weight regulation. Resting EE accounts for two-thirds of total EE in humans. Colder environmental temperature has shown to increase resting EE in lean adults. However, the effect of cold-induced thermogenesis remains unclear in the presence of other influential factors such as gender, ethnicity and obesity. In this study, we aim to characterize the impact of cold temperature on EE controlling for individual characteristics as well as time-varying factors.

Methods: Each of fifty subjects underwent a random sequence of two 12-hour sessions in a cross-over trial. Each subject's EE was recorded every minute in temperature-controlled whole-room indirect calorimeters at 24°C and 19°C, respectively. Counts of spontaneous movements in wrist, ankle, and hip were recorded every 15 seconds by accelerometer-based activity monitors. Investigators recorded gender, age, race, body compositional measurements, and hormone levels from plasma and urine tests at baseline. We analyze EE by a hierarchical random coefficient model, we control for activity counts and mealtime. At individual level, we control by the baseline. We employed forward variable selection by the likelihood ratio test and tested random effects by an equally-weighted mixture of chi-squared distributions. AIC was used to compared non-nested models.

Results: Thirty-five subjects had complete information on all covariates. Within-person variability represented 54% of the total variability. Overall, cold environmental temperature increased EE. However, this effect is lower among the overweight than the normal, and decreases with age.
Title: Development of the Healthy Places Index (HPI)
Presenter: Christine Orndahl
Advisor: Roy Sabo
Abstract:

Due to finite resources and increased budgetary constraints many states and localities are interested in tools and approaches to more strategically allocate resources and aid. Health Disadvantage Index 1.1 (HDI 1.1) was created in an effort to help determine how California’s resources should be distributed among the census tracts. While HDI 1.1 was able to combine necessary components, such as economic and environmental policy action areas, into a single accessible score, it was underutilized by health policy officials. One possible reason was that the determination of appropriate weights for the different policy action areas was based on prevailing literature and expert judgement instead of statistically-based evidence. To remedy this, a hybrid approach was taken to reanalyze and revamp HDI 1.1 into the Healthy Places Index (HPI). This new index takes into consideration past literature and expert judgment for selecting individual components while introducing weighted quantile sum regression to statistically determine the proper weights for policy action areas. In doing so, the HPI outperformed HDI 1.1 in its ability to explain variation in life expectancy, a health outcome chosen to determine the effectiveness of these indices, by increasing the value of $R^2$. The combination of expertise and statistics was successful in improving index performance while maintaining the ability to help identify policy action areas most in need of assistance.
Title: Sequential Assignment of Responder Status in the First Stage of SMART-Like Trials
Presenter: Keighly Bradbrook
Advisor: Roy Sabo
Abstract:

Sequential, multiple assignment, randomized trials (SMARTs) are a personalized approach to study randomization in which, individuals are adaptively and successively randomized to different stages of treatment, based on their responsiveness to the current stage of treatment. Successful implementation of SMARTs depends on the ability of investigators to distinguish which patients are responsive to the current treatment and which patients are not. In some clinical situations, there may not exist an acceptable mechanism to make the classification between responder and non-responder. In such cases where response designation is not clear, the sequential, multiple assignment, randomization method is difficult to use, as there is no known basis for next-stage randomization. The goal of the current work is to determine whether it is possible to develop a method to sequentially assign responder statuses to subjects completing the first stage of a SMART-like trial based on information provided through pilot data. The proposed method performs a cluster analysis on pilot data to determine which individuals could likely be classified as responders for each of two first-stage treatments. The responder/non-responder designations resulting from the cluster analysis are then used in a discriminant analysis to estimate the probability that the latest individual with an observed stage-one outcome is a responder. Lastly, this probability of response to the treatment is used to allocate that subject to responder or non-responder status. A simulation study is used to investigate the efficacy of this approach.