Iowa Summer Institute in Biostatistics

2016

7th Annual
Iowa Summer Research Symposium

University of Iowa
Iowa City, Iowa

July 21st, 2016

Room C217
College of Public Health

Table of Contents

ISIB Presentations

Mario Ochoa
University of Puerto Rico in Cayey
Michelle Weitz
Hamline University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Modeling the Potential Range of the Zika Virus Vector Aedes aegypti

The Zika virus is an ongoing public health crisis with the current outbreak primarily affecting Latin America, but with the potential to spread in the United States. The virus is primarily transmitted by bites of the Aedes aegypti and Aedes albopictus mosquitoes, both of which can be found in North America, and has been linked to serious neurological birth defects. Our research focuses on predicting the probability of the occurrence of the Aedes aegypti mosquito throughout the US. In order to assess the suitability of locations for mosquito survival, we use species distribution modeling, specifically logistic regression and spatial logistic regression, with geographic and climatic predictor variables. We perform this analysis both state-by-state and using individual mosquito observations. Based on our findings of statistically significant predictor variables including temperature, precipitation, anthropogenic biomes, and carbon dioxide emissions, we produce probability maps of the geographic distribution of Aedes aegypti.

Research Mentor
M. Kathryn Cowles, Professor, Dept. of Statistics and Actuarial Science, University of Iowa

Elliot Burghardt
University of Iowa
Daren Kuwaye
University of Hawai'i at Mānoa

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Medical Imaging for Lung Cancer Detection
Testing Statistical Models to Improve Screening

As the leading cause of cancer deaths, lung cancer is responsible for over 1 in 4 cancer deaths in the United States.[1] The National Lung Screening Trial (NLST) demonstrated that CT screening detects more early-stage cancers than chest radiography and reduces mortality from lung cancer.[2] As a result, the US Preventive Services Task Force has recommended annual CT screening for high risk individuals. However, 96.4% of the abnormal CT screens from the NLST were false positive results, requiring additional imaging and/or invasive procedures to resolve the diagnosis.[2] In an effort to decrease false positive rates of these screens, we studied the relationships between 144 radiomic features of nodule and surrounding tissue – size, intensity, shape, texture – and clinical information from a cohort of 199 individuals screened for lung cancer. We used supervised statistical learning methods to develop Stochastic gradient boosting, neural network, partial least squares and penalized logistic regression models to predict cancer status. These models utilized cross validation to establish predictive performance. We selected the model that maximized the area under the ROC curve as the best performing multivariate model to predict cancer status. This model suggests that we may decrease the false positive rate of lung cancer screening by using the radiomic data contained within those scans, reducing the number of patients who unnecessarily undergo further imaging and invasive procedures.

Research Mentor
Brian J. Smith, Associate Professor, Dept. of Biostatistics, University of Iowa

1. American Cancer Society. Cancer Facts & Figures 2016. Atlanta: American Cancer Society; 2016.
2. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. New England Journal of Medicine. 2011;365(5):395-409.

Ariana Cavazos
California State University, Fresno
Rebecca Rasnick
East Tennessee State University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Association Between Initial Treatments and Subsequent Primaries in Hodgkin's Lymphoma Patients

Hodgkin's Lymphoma (HL) is a type of cancer that affects the lymphatic system and compromises the human body's ability to fight infection. HL typically begins in white blood cells. It occurs when a specific type of cell, the Reed-Stenberg cell, is present in the host's system, causing the body's infection fighting cells to develop a mutation in their DNA. Each year, there are several thousand individuals in the United States and worldwide who develop HL. It has been hypothesized that initial treatment after diagnosis may be associated with subsequent recurrence—be it recurrence of HL or other types of cancer. We explore this association using the Oncology Registry at the University of Iowa. Using an adaptation of the Cox regression, we have found a significant effect of initial treatment on the hazard of recurrence, with Radiation Therapy most likely to significantly increase this hazard. Other prognostic factors such as Age and Gender have significant effect on the hazard of recurrence. In our findings, we adjusted for subsequent treatments following secondary primaries.

Research Mentor
Gideon K.D. Zamba, Associate Professor, Dept. of Biostatistics, University of Iowa

Alyssa Bellomo
Kean University
Dorcas Washington
Transylvania University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Justifying Modifications to a CRM Dose-Finding Study Because of a Large Change in the Recruitment Rate

The Continual Reassessment Method (CRM) is an adaptive, dose-finding design commonly used in Phase 1 clinical trials. The CRM chooses a maximum tolerated dose (MTD) by continually updating the dose-toxicity curve after each determination of dose-limiting toxicities (DLTs) within each cohort of subjects. Similar to other dose-finding designs, a drawback for the traditional CRM is that the rate of recruitment can outpace the DLT observation period, thereby creating delays in subject accrual that can negatively impact recruitment. Instead of pausing subject recruitment to wait for the DLT assessment period to close, an alternative solution would be to implement a lagged approach during recruitment where only data from the previously completed cohorts is used while the most recent cohort is still being processed. In this study, simulations of a modified CRM were constructed using R programming language to examine the effectiveness of a current dose-finding trial when, due to a large change in the recruitment rate, a lagged approach is adopted halfway through subject accrual. One and two cohort lags were compared to the original simulation used to design the trial. It is shown that the implementation of a lagged approach in the modified CRM does not hinder the study's ability to determine the correct MTD.

Research Mentor
Eric D. Foster, Assistant Professor, Dept. of Biostatistics, University of Iowa

Jose Rueben Bautista
The University of Arizona

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Predictive Modeling for Student Retention

For years, universities have been collecting large amounts of data on current and prospective students. With advances in analytical techniques, there is an increasing demand for productive use of this information. For the purpose of this project, we will focus on data regarding incoming freshmen students at the University of Iowa. Using this information, we will investigate the degree to which second year student retention and first year GPA can be predicted at the time of admission.

In order to predict second year student retention, many variables must be considered. Statistical learning allows an investigator to flexibly explore the relationship between a wide array of predictor variables and a specific outcome. While traditional statistical techniques such as linear regression can be used to make predictions, the nature of the outcome in question may not follow a simple linear pattern. Decision trees, one of the many tools of statistical learning, divide observations into groups based on predictor variables and can be used to construct procedures with good predictive power. In R, both the randomForest and gbm packages use decision trees to make predictions, though they work in very different ways. We find that second year enrollment may be predicted by: high school GPA, high school class size, home state, and the number of contacts a student has had with a university. First year GPA is predicted by similar variables as enrollment; however, ACT score is also useful predictor variable.

Research Mentor
Grant D. Brown, Assistant Professor, Dept. of Biostatistics, University of Iowa

Joseph Alamo
Kean University
Senan Agblonon
Norwalk Community College

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Predicting Crop Damage by Vervet Monkeys on St. Kitts

The island of St. Kitts in the Caribbean has hosted a population of vervet monkeys for over 350 years. These monkeys are notorious for raiding farms and inflicting crop damage. This project develops a predictive model to evaluate the probability of crop damage occurrence due to vervet monkeys.   Data has been compiled on 6115 distinct units of farmland on the island. General linearized models are constructed from 80 percent of the data with the following covariates: season, the amount of neighboring farms, the extent of guarding, the crop preference of vervet monkeys, and proximity to water, forest, and roadways. Candidate models include the saturated model as well as reduced models obtained from a simple model selection procedure. The remaining 20% of the data serves as a testing set to assess the predictive capabilities of the candidate models. Different computational techniques are applied to obtain optimal specificity and sensitivity.

Research Mentor
Daniel K. Sewell, Assistant Professor, Dept. of Biostatistics, University of Iowa

Justin DeMonte
The Evergreen State College
Elise Northrop
Clarke University

Iowa Summer Institute in Biostatistics – University of Iowa
Department of Biostatistics

An Analysis of Risk Factors for Teen Dating Violence

The Youth Risk Behavior Surveillance System (YRBSS) consists of surveys given biannually since 1991 at the national, state, territorial, tribal, and local levels, based on representative samples of students in grades 9 through 12. This project investigates potential risk factors for physical and sexual teen dating violence (TDV) among a subset of 47,720 students surveyed in the 2013 YRBSS. We use univariable and multivariable logistic regression to estimate associations between risk factors and each form of TDV. To assess the relative strength of each risk factor, we consider the values of the Bayesian Information Criterion. Relative to the baseline prevalence of physical TDV of 10.2% and sexual TDV of 10.8%, the results show that students are more likely to report either form of TDV if they also report bullying victimization, greater levels of alcohol or illegal drug use, multiple recent sex partners, or school absences due to feeling unsafe. Lastly, we construct an R Shiny risk assessment tool encompassing our models to predict the likelihood of an individual experiencing TDV.

Cicely Krebill
Northeastern University
Maia Adar
Claremont McKenna College

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Genetic Risk Factors for Complications due to Arrhythmia

Implantation of an internal cardioverter-defibrillator (ICD) can be lifesaving for people with arrhythmias, but it is also a costly and invasive procedure. This research focuses on predicting which patients will benefit from ICD placement by looking at both clinical and genetic factors. Using survival analysis of time until death and time until first shock from ICD among a population of ICD users, we created Kaplan-Meier curves and used Cox regression to determine which clinical factors should be accounted for in the genetic analysis. We then performed a genome-wide and a targeted analysis to locate single nucleotide polymorphisms (SNPs) that are associated with risk of either shock or death. This resulted in the identification of 28 SNPs that are potentially indicative of whether a patient will benefit from an ICD.

Mentors of Research Group

Patrick J. Breheny, Assistant Professor, Dept. of Biostatistics, University of Iowa
Lauren N. Sager, Graduate Student Mentor, Dept. of Biostatistics, University of Iowa

Charimhel Rarallo
University of Hawaiʻi at West Oʻahu
Helin Hernandez
University of California, Los Angeles

Iowa Summer Institute in Biostatistics – University of Iowa
Department of Biostatistics

An Evaluation of Elevated Blood Lead Levels in Iowa Newborns

High lead levels can lead to serious health problems such as headaches, stomach pain, behavioral problems, neurological development issues and anemia. Exposure to lead is particularly dangerous for newborns and young children because their small growing bodies are at a higher risk of absorbing and retaining lead compared to a developed body. This study aims to analyze the socioeconomic factors and geographical areas that contribute to the risk of high lead levels in Iowan newborns. Our study evaluated 2,741 newborn dried blood spot samples over a 2-month period in 2006 which were analyzed at the State of Iowa Hygienic Laboratory. The dried blood spot samples with elevated blood lead levels collected and evaluated were identified based on the mother's ZIP code of residence. We implemented a Bayesian generalized linear mixed model to evaluate the relationship between the number of babies with elevated blood lead levels to the averages of five ZIP code level covariates of interest drawn from the U.S. Census Bureau's 2007-2011 American Community Surveys. We also conducted a Poisson regression model that accounted for spatial correlation by using a spatial random effect which incorporated a neighborhood structured conditional autoregressive (CAR) model. The results of our analysis can be used to inform the population of Iowa where there may be a higher potential of lead exposure that would cause harm to newborns.

Mentors of Research Group
Jacob J. Oleson, Associate Professor, Dept. of Biostatistics, University of Iowa
David Zahrieh, Graduate Student Mentor, Dept. of Biostatistics, University of Iowa

Benjamin Jacobs
University of Iowa
Kelvin Ausby
Savannah State University

Iowa Summer Institute in Biostatistics - University of Iowa
Department of Biostatistics

Does the CPAP Machine Improve the Amount of Sleep for People with Obstructive Sleep Apnea?

Obstructive Sleep Apnea (OSA) is a condition which negatively affects individuals' ability to sleep at night, and is commonly treated with a Continuous Positive Airway Pressure (CPAP) machine. Using nightly sleep in minutes and CPAP machine usage data, originally gathered to study the effects of OSA on driving ability, we tested to see how the sleep patterns of individuals with OSA differ from those without. Using mixed model regressions, we test to see which other covariates may have an influence on the efficacy of CPAP machines in improving the sleep of individuals with OSA. Although preliminary findings showed no difference in the average amount of sleep per night between control and OSA groups at baseline, there appears to be a positive relationship between the amount of CPAP machine usage and the amount of sleep obtained by individuals with OSA. Also, the patterns of CPAP machine usage appears to differ between male and female subjects.