

Iowa Summer Institute in Biostatistics

2014

Iowa Summer Research Symposium

University of Iowa

Iowa City, Iowa

July 17th, 2014

Room C217

CPHB

Table of Contents

ISIB Presentations

Predicting Lyme Disease Incidence in Humans and Dogs	1
Assessing the Effect of Practical Considerations when using..... the CRM in Dose Finding Studies	2
Assessing the Role of Patient Factors on Surgical Complication in..... Pancreatic Cancer Patients	3
Multivariate Missingness in the Identification of Prognostic Factors..... for Survival in Neuroendocrine Tumor Patients	4
Site-of-Origin Prediction for Gastroenteropancreatic Neuroendocrine Tumors...	5
An Analysis of Indicators for High Risk Behavior in Adolescents	6
Based on the Iowa Youth Survey	
Speech Intelligibility Index Model: A Key Aspect to a Child's Development..... of Speech and Language	7
Causes and Severity in Motor Vehicle Crashes in Iowa.....	8

Katherine A. Cartagena

Universidad de Puerto Rico at Humacao

Cody Hansen

Augustana College

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

Predicting Lyme Disease Incidence in Humans and Dogs

Lyme disease is the most common vector-borne illness in the United States. It is caused by the bacterium *Borrelia burgdorferi* and is transmitted to humans through the bite of infected blacklegged ticks (or deer ticks, *Ixodes scapularis*). Early treatment of Lyme disease is important to prevent complications like arthritis, pain, tingling in the hands or feet, and problems with short-term memory. According to the Centers for Disease Control and Prevention (CDC), there are up to 30,000 reported cases in the U.S each year and preliminary research from the CDC suggests that 300,000 cases of Lyme disease occur annually.

Our research goal is to predict new U.S. states where Lyme disease will occur. Dogs are routinely tested for Lyme disease in combination with their annual tests for heartworm. A second goal is to determine whether inclusion of data on incidence in dogs improves the predictive ability of statistical models for human Lyme disease incidence. We fit hierarchical generalized linear models to CDC human incidence data from 2007 to 2011 for each of the 50 states, with variables representing state climate, geography, and incidence rates in dogs as predictors. We confirmed that dog incidence rate is a highly significant predictor and improves the ability of the statistical model to predict human incidence rates in 2012.

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

Caroline Sprecher

University of North Carolina-Chapel Hill

Tyler Bonnett

Marshall University

Rolando Acosta Nuñez

University of Puerto Rico-Humacao

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

Assessing the Effect of Practical Considerations when using the CRM in Dose Finding Studies

The primary objective of most Phase I clinical trials is to determine the maximum tolerated dose (MTD) to carry forward in future stages of clinical research. The continual reassessment method (CRM; O'Quigley et. al. 1990) is a widely accepted and statistically sound technique for dose finding trials. Via simulation, our research investigates the performance of the CRM under various design constraints. Of particular interest are the practical considerations that arise due to the relationship between the number of doses selected and the maximum sample size allowed. Based on a dose finding study carried out by a large pharmaceutical company where it became obvious that the CRM might recommend a dose that had few-to-no subjects allocated to it, we analyze the performance of two modifications of the CRM: (1) removing untested doses from consideration at a predetermined late stage of the trial and (2) reducing the number of doses to be tested prior to the start of the study.

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

Mitchell A. Thomann, Graduate Student Mentor, Dept. of Biostatistics, University of Iowa

Jeffrey Antunes

Kean University

Brooke Bell

University of Southern California

Jolene Ranek

University of Scranton

Iowa Summer Institute in Biostatistics - University of Iowa

Department of Biostatistics

Assessing the Role of Patient Factors on Surgical Complication in Pancreatic Cancer Patients

Pancreatic cancer is a particularly aggressive disease that ranks as the fourth most common cause of cancer-related deaths in the United States. The 5-year survival rate is 23% for patients with localized disease and 2% for those with advanced disease. Surgery has the potential to increase survival, but also carries the risk of surgical complications. There is interest nationally in identifying patients and practices that may be associated with complications so that steps can be taken to reduce their occurrences. Using data obtained from the American College of Surgeons National Surgical Quality Improvement Program (NSQIP), we examined three possible surgical complications that occur within 30 days after the operation; these infections are categorized by their involvement with the skin or subcutaneous tissue, deep soft tissues, or any part of the anatomy. We employed statistical methods such as logistic regression and stepwise regression model selection to find statistically significant variables. We used these variables to build predictive models for the risk of infection in surgically treated pancreatic cancer patients. Accordingly, we tested the predictive performance of our models on an independent set of data. We distinguished the most significant characteristics of patients that maximized the predictive property of our models, which can be used to predict the risk of infection in pancreatic cancer patients who undergo surgery.

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

Sarah L. Bell, Biostatistician, Holden Comprehensive Cancer Center, University of Iowa

Evan Walser-Kuntz

Haverford College

Philip Stallworth

Reed College

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

Identification of Prognostic Factors for Survival in Neuroendocrine Tumor Patients in the Presence of Multivariate Missingness

The objective of this project is to identify prognostic factors associated with higher survival in patients diagnosed with malignant neuroendocrine tumors (NET) in the small bowel from 1988-2010. It has been hypothesized that various demographics (age, gender, racial/ethnic background, and marital status) and cancer related factors (surgery status, tumor size, cancer stage, grade, and lymph node ratio (LNR)) may be associated with survival. Physicians are particularly interested in determining whether an association between tumor size and survival exists in the presence of previously known prognostics. In the de-identified data from the Surveillance Epidemiology and End Result (SEER) database, some of the factors of interest were not fully measured--resulting in missingness within the data. Missingness has rendered complete case analyses unviable. We use multiple imputation (MI) from the MICE algorithm and imputation models to address a potential loss of information induced by missingness. The data analysis is stratified by year of diagnosis and separate analyses are conducted per location (duodenal, jejunoileal, or small bowel NOS). We use the Cox proportional hazard model to identify factors associated with the hazard of death within the population under study. Our analyses reiterate that age, surgery status, grade, and stage have a strong association with survival and provide evidence that LNR and tumor size also have a significant association with survival in NET patients. Using these factors, simplified survival models could be constructed for each individual site.

Gideon KD Zamba, Associate Professor, Dept. of Biostatistics, University of Iowa

Sarah L. Bell, Biostatistician, University of Iowa Holden Comprehensive Cancer Center

Thomas Nemmers

Creighton University

Adriana Granados

California State Polytechnic University, Pomona

Saniya Khullar

Georgetown University

Iowa Summer Institute in Biostatistics – University of Iowa

Department of Biostatistics

Site-of-Origin Prediction for Gastroenteropancreatic Neuroendocrine Tumors

In the past thirty years, the number of cases of Neuroendocrine Tumors (NETs) has increased five-fold. The most common NET is the Gastroenteropancreatic NET (GEPNET); the two most common GEPNETs are Pancreatic NETs (PNETs) and Small Bowel NETs (SBNETs). Both of these types of tumors have very different prognoses and treatment procedures. However, adequate practical methods have not been developed for distinguishing between the two GEPNETs at the biopsy level, especially after metastasis. In lieu of this problem, we analyze data from a microarray of 22011 genes from five PNETs and eleven SBNETs. From this data, we use a two-sample t-test to find 4635 genes that are differentially expressed between SBNETs and PNETs at a one-percent false discovery rate. We also develop logistic regression models using primary tumor microarray, primary tumor qPCR, and metastatic tumor qPCR data to predict site-of-origin for metastatic tumors and validate these predictions against an external testing set.

Patrick J. Breheny, Assistant Professor, Dept. of Biostatistics, University of Iowa

Javier Flores

University of Texas at Brownsville

Elizabeth Sigworth

St. Olaf College

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

An Analysis of Indicators for High Risk Behavior in Adolescents Based on the Iowa Youth Survey

The Iowa Youth Survey (IYS) is a tool used to measure the experiences of students in the 6th, 8th, and 11th grades regarding various forms of violence, substance abuse, and other high risk behaviors. The survey also provides information regarding students' views of themselves, their peers, and their surrounding environment. Based on IYS data collected in 2005 and 2008, this research analyzes the relationship between select indicators and high risk behavior outcomes using simple and multiple logistic regression. Graphically, we present the changes in the odds ratios of each outcome with respect to the levels of each indicator. We gauge the relative strength of each association by referencing the differences in the Bayesian Information Criterion (BIC) between models with and without given indicators. Additionally, we use generalized estimating equations to demonstrate the effect of school district level clustering on each association.

Joseph E. Cavanaugh, Professor, Dept. of Biostatistics, University of Iowa

Katelyn Zumpf
North Central College

Melissa Jay
Colorado College

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

Speech Intelligibility Index Model: A Key Aspect to a Child's Development of Speech and Language

The development of speech and language in children is critically impacted by the child's ability to hear. It has been found that children with hearing loss, who fail to seek proper help, have a delay in their speech development. Therefore, properly identifying and taking appropriate corrective action for their hearing is vital to the future development of these children. The overall goal of this study is to determine when hearing loss is identified, which hearing corrective action approach would provide the child with the most long term advantages: Hearing aids or Cochlear implants. To do this, we propose a method to use a child's Speech Intelligibility Index (SII), the standard estimate of an individual's understanding of speech, as a measuring tool to discriminate between providing the child with hearing aids or cochlear implants. However, in children with cochlear implants, there is no means of acquiring an SII, so other measures of hearing such as Better Ear Pure-Tone Average (PTA), and reading ability measures such as Word Attack, and Passage scores are used to aid in the discrimination. In this research, we use multiple imputation and multiple linear regression to find a model that predicts "functional" SII values for children with cochlear implants. We find that a model of three explanatory variables, Word Attack, Passage Score, and Mother's Education, best predicts the SII scores for children wearing cochlear implants and we provide a tool for discriminating between corrective approaches for children identified with hearing loss.

Jacob J. Oleson, Associate Professor, Dept. of Biostatistics, University of Iowa

Tiffany Bailey
Central State University

Amy Hahn
University of Iowa

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

Causes and Severity in Motor Vehicle Crashes in Iowa

The Injury Registry of the University of Iowa Injury Prevention Research Center (IPRC) has worked with the Iowa Department of Transportation for many years, obtaining details of motor vehicle crashes. The collected information includes hundreds of variables grouped in 17 annual datasets pertaining to crash location, environment and roadway, vehicle, driver and passenger characteristics, and contributing factors.

In our study, we obtained crash data from 765,894 crashes recorded for the years 2001 to 2013. The data set did not include personal information that could be used to identify the drivers or passengers involved. We focused our analyses on the causes, costs, and injury severity of the crashes. We used tables, stacked bar charts, boxplots, and line plots to illustrate associations among these variables, and to look for changes across time. We used the Kruskal-Wallis test and ordinal logistic regression to perform formal statistical tests to show the significance of the relationships we saw. For example, we found that the causes of crashes that tended to have the most severe injuries were a) disregard of road signs and b) road departures ($p < 0.0001$), while the ones with the highest repair costs were a) disregard of road signs and b) reckless or inadequate driving ($p < 0.0001$). We also found that, given the level of crash severity, all types of crashes decreased over the 13 year period we studied.

Our findings could be used in educational campaigns by driving schools, insurance companies, and government agencies to target high risk behavior in an effort to decrease the incidence of crashes with the most severe injuries and costs.

Jeffrey D. Dawson, Professor, Dept. of Biostatistics, University of Iowa