- Emphasizing the effect size and answering the scientific hypothesis when reporting results.
- Statistical code used to produce results should be well annotated and raw data posted online to enhance reproducibility.

Approaches:

- Incorporate effective multiple modalities (i.e. didactic, demonstrative, hands on workshops, applications, and tools).
- Approach from "the drivers' seat" perspective, rather than strictly mathematical.
- Endorse flipped classroom approach

DISCUSSION/SIGNIFICANCE OF IMPACT: Like any complex discipline, biostatistical education can be approached from several dimensions, but it remains essential to focus on fundamental goals of science. We remind our trainees that the goal of science is to create knowledge, not to "find significance". Deep understanding of inferential methods and proper interpretation of results are key. CONFLICT OF INTEREST DESCRIPTION: None.

READ-TV: Research and Exploratory Analysis Driven Time-data Visualization

John Del Gaizo[1], Alexander Alekseyenko[1], and Kenneth Catchpole[1]
[1]Medical University of South Carolina

OBJECTIVES/GOALS:

- A web interface that allows for easy upload of CSV text data to time-based visualizations
- Implementation of change points analysis to identify and display points where event rates increased or decreased
- customizable plots where the user can change point shapes, color, etc.
- customizable and advanced filtering support
- support for plot comparisons and exports

METHODS/STUDY POPULATION: We used the R/Shiny framework to develop a web application for visualization of time stamped data. The Research and Exploratory Analysis Driven Time-data Visualization (READ-TV) application allows for user-friendly mining for longitudinal patterns in data. READ-TV is built specifically for FD analysis, but is easily adaptable to other clinical use cases, as we allow for the use of general metadata on events and cases. The building of a quantitative framework for event analysis starts with the application of homogeneous Poisson processes, which relate the times of occurrence of events in terms of an underlying rate. To understand the changes in this underlying rate, changepoint analysis is used to model the rate as a function of time using piecewise constant approximations. The changepoint analysis allows us to identify the specific periods of time where the rate of FD is increased relative to a baseline or a desired operating range. RESULTS/ANTICIPATED RESULTS: READ-TV application allows for import of time stamped event data from multiple cases. Event and case metadata are supported to facilitate filtering and mining of interesting subsets of data. Stem plots are used for visualization of selected event timelines in chosen cases. This visualization is accompanied with summary of the number and estimates of rates of occurrence of specific event types (e.g. types of FD). Change-point analysis is implemented using the 'changepoint'

R library. These analyses allow the users to quickly understand whether the rates of events (FD) is changing across the case timeline and where exactly these changes are occurring. DISCUSSION/SIGNIFICANCE OF IMPACT: We have demonstrated the READ-TV application to the team of the AHRQ-funded Human Factors and Systems Integration in High Technology Surgery (HF-SIgHTS) study. The ability to visualize and perform quantitative analysis of the study data was received with unanimous positive feedback and enthusiasm. We continue READ-TV development focusing on (1) increased user-friendliness using the HF-SIgHTS as our focus group, (2) increased functionality, and (3) use of more general localization terminology to allow for other applications.

Statistical Modeling for Predicting Correct Drug Dose in the Presence of Conflicting Dose Information Extracted from Electronic Health Records

Michael Lee Williams[1], Hannah L Weeks[1], Cole Beck[1], Elizabeth McNeer[1], and Leena Choi[1]
[1]Vanderbilt University Medical Center

OBJECTIVES/GOALS: Diverse medication-based studies require longitudinal drug dose information. EHRs can provide such data, but multiple mentions of a drug in the same clinical note can yield conflicting dose. We aimed to develop statistical methods which address this challenge by predicting the valid dose in the event that conflicting doses are extracted. METHODS/STUDY POPULATION: We extracted dose information for two test drugs, tacrolimus and lamotrigine, from Vanderbilt EHRs using a natural language processing system, medExtractR, which was developed by our team. A random forest classifier was used to estimate the probability of correctness for each extracted dose on the basis of subject longitudinal dosing patterns and extracted EHR note context. Using this feasibility measure and other features such as a summary of subject dosing history, we developed several statistical models to predict the dose on the basis of the extracted doses. The models developed based on supervised methods included a separate random forest regression, a transition model, and a boosting model. We also considered unsupervised methods and developed a Bayesian hierarchical model. RESULTS/ANTICIPATED RESULTS: We compared model-predicted doses to physician-validated doses to evaluate model performance. A random forest regression model outperformed all proposed models. As this model is a supervised model, its utility would depend on availability of validated dose. Our preliminary result from a Bayesian hierarchical model showed that it can be a promising alternative although performing less optimally. The Bayesian hierarchical model would be especially useful when validated dose data are not available, as it was developed in unsupervised modeling framework and hence does not require validated dose that can be difficult and time consuming to obtain. We evaluated the feasibility of each method for automatic implementation in our drug dosing extraction and processing system we have been developing. DISCUSSION/SIGNIFICANCE OF IMPACT: We will incorporate the developed methods as a part of our complete medication extraction system, which will allow to automatically prepare large longitudinal medication dose datasets for researchers. Availability of such data will enable diverse medication-based studies with drastically reduced barriers to data collection.