will investigate dietary intervention as a potential disease-modifying factor. CONFLICT OF INTEREST DESCRIPTION: NA, nothing to disclose.

4183

Micro-consults: An effective tool for meeting statistical support needs in an academic medical research center

Sandra Taylor[1], and Susan L. Stewart[1]
[1]University of California, Davis

OBJECTIVES/GOALS: Access to biostatistics expertise is essential for a successful clinical and translational research program. However, demand for statistical support at academic research centers can strain the capacity of biostatistics units. Our objective was to efficiently increase access to statistical expertise. METHODS/ STUDY POPULATION: In cooperation with the Cancer Center Biostatistics Shared Resource, we replaced an informal 1-hour drop-in consultation program with structured office hours to provide statistical support to clinical and translational researchers at the University of California, Davis Medical Center. We doubled office hours to 2 hours per week and established six 20-minute appointments. Two Ph.D. level statisticians staff office hours. Researchers schedule appointments through Acuity Scheduling, a free on-line resource. Availability of the service is advertised monthly by sending an informational flyer to various university listservs. RESULTS/ANTICIPATED RESULTS: Prior to implementing the program in 2014, we averaged 91 office hour consults per year. Subsequently, consultations jumped to 171 in 2014 and have averaged 150 per year since then. Office hours attract students, residents, staff and faculty from a wide range of disciplines including the Schools of Medicine, Nursing, Veterinary Medicine and basic science departments. Project types span the clinical and translational spectrum covering lab, animal, clinical and population-level studies. Most consults related to data analysis and interpretation (57%) followed by sample size calculations/study design (29%) and response to reviewers (4%), with general statistical advice as the remainder. DISCUSSION/SIGNIFICANCE OF IMPACT: With 6 micro-consults per week, we can meet with many investigators and triage their statistical support needs. This program has proved very popular and was highly rated in a recent user survey, with several investigators noting that the consults facilitated successful publications and proposals.

4494

Predictors of Reintubation After Cardiac Surgery

Robert Edward Freundlich[1], Gen Li[1], Jonathan P Wanderer[1], Frederic T Billings[1], Henry Domenico[1], Daniel Byrne[1], and Pratik Pandharipande[1]
[1]Vanderbilt University Medical Center

OBJECTIVES/GOALS: We modeled risk of reintubation within 48 hours of cardiac surgery using variables available in the electronic health record (EHR). This model will guide recruitment for a prospective, pragmatic clinical trial entirely embedded within the EHR among those at high risk of reintubation. METHODS/ STUDY POPULATION: All adult patients admitted to the cardiac intensive care unit following cardiac surgery involving thoracotomy or sternotomy were eligible for inclusion. Data were obtained from operational and analytical databases integrated into the Epic EHR, as well as institutional and departmental-derived data warehouses, using structured query language. Variables were screened for inclusion in the model based on clinical relevance, availability in the EHR as structured data, and likelihood of timely documentation during routine clinical care, in the hopes of obtaining a maximally-pragmatic model. RESULTS/ANTICIPATED RESULTS: A total of 2325 patients met inclusion criteria between November 2, 2017 and November 2, 2019. Of these patients, 68.4% were male. Median age was 63.0. The primary outcome of reintubation occurred in 112/2325 (4.8%) of patients within 48 hours and 177/2325 (7.6%) at any point in the subsequent hospital encounter. Univariate screening and iterative model development revealed numerous strong candidate predictors (ANOVA plot, figure 1), resulting in a model with acceptable calibration (calibration plot, figure 2), $c = 0.666$. DISCUSSION/SIGNIFICANCE OF IMPACT: Reintubation is common after cardiac surgery. Risk factors are available in the EHR. We are integrating this model into the EHR to support real-time risk estimation and to recruit and randomize high-risk patients into a clinical trial comparing post-extubation high flow nasal cannula with usual care. CONFLICT OF INTEREST DESCRIPTION: REF has received grant funding and consulting fees from Medtronic for research on inpatient monitoring.

4099

Principles of Statistical Education for Translational Scientists in the Age of Rigor, Reproducibility, and Reporting

Emilia Bagiella, PhD[1], Paul Christos[1], Mimi Kim, ScD[2], Shing Lee[1], Roger Vaughan[3], and Judy Zhong[4]
[1]Mount Sinai School of Medicine; [2]Albert Einstein College of Medicine; [3]Rockefeller University; [4]New York University General Clinical Research Center

OBJECTIVES/GOALS: To describe principles, best practices, and techniques recommended to instill deep understanding of the application and interpretation of statistical techniques and statistical inference among translational scientists and trainees, that best support the concepts of scientific Rigor, Reproducibility and Reporting. METHODS/ STUDY POPULATION: Each of the six New York City Area Biostatistics, Epidemiology and Research Design (BERD) resources have strong educational programs, novel curricular components, and creative strategies, implemented by award winning educators. To capitalize on shared knowledge, innovation, and resources, the six teams formed the **New York City Area BERD Collaborative (NYC-ABC)** comprised of BERD resources from Mt. Sinai, Cornell, Einstein, Columbia, Rockefeller, and NYU. The collaborative suggests principles, concepts, tools and approaches to support the concepts of scientific Rigor, Reproducibility and Reporting in translational science. RESULTS/ ANTICIPATED RESULTS: Principles:

- Value of team science approach and including biostatisticians early and often.
- Carefully designing experiments to reduce bias and increase precision.
- Trainees' focus is often on "statistical significance" and the p-value. Consequences of data dredging/p-hacking, and the impact of sample size and other factors on statistical significance.

- 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.

4474

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.

4501

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.