Peter E. DeWitt

Worldwide, more than 3.3 million pediatric deaths per annum can be attributed to sepsis and septic shock. The standard for diagnosis of sepsis was defined by the International Pediatric Sepsis Consensus Conference (IPSCC) in 2005. However, the IPSCC criteria has become outdated, has low specificity, does not allow for risk stratification by resource setting, and may be discordant with clinician-based diagnoses. Diagnostic criteria for adult sepsis was updated by the Sepsis-3 Task Force in 2016 to shift the diagnostic focus from inflammatory response to a diagnosis based on life-threatening organ system dysfunction. In 2019, the Society of Critical Care Medicine created the Pediatric Sepsis Definition Task Force to update the diagnostic criteria for pediatric sepsis based on the conceptual framework of sepsis as suspected infection with life-threatening organ dysfunction and that would generalize across differently resourced settings. The result is the Phoenix criteria.

This presentation will focus on the data driven Delphi process used to define the Phoenix criteria. Data came from the electronic health records (EHR) of six high- resourced, and four lower-resourced hospital systems from North America, Asia, South America, and Africa; nearly 3.7 million emergency department and intensive care unit encounters between 2010 and 2019. By splitting the data into multiple disjoint subsets, we were able to 1) identify established organ dysfunctions scores which were predictive of hospital mortality and could be applicable regardless of resource setting. 2) Assessing multiple combinations of the preferable organ dysfunction scores to find a composite model. 3) Define an easy to use integer valued scoring system and threshold, and finally, 4) test the integer valued model on both within and holdout testing sets. Each step was a combination of data driven and Delphi process.

Lastly, we conclude this presentation by introducing tools for applying the Phoenix criteria to other EHR data sets.

Jason Bernstein

The Space Domain Awareness (SDA) Tools, Applications, and Processing (TAP) Lab is a Colorado-based initiative of the U.S. Space Force to develop and transfer technology from industry, academia, and federally funded research and development centers to Space Force operators. Tasks of interest include detecting rocket launches from seismic data, determining if two objects in orbit around Earth are on a collision course, and monitoring the light reflecting off a satellite over time to assess whether the satellite is experiencing a change in stability. In this talk, I will discuss the statistical nature of these problems. For example, estimating orbits from range or angle data can be viewed as a maximum likelihood estimation problem, and computing collision probabilities requires propagating probability distributions through non-linear physics models. I will also highlight opportunities for academic engagement with the SDA TAP Lab and show how publicly-available data on orbits and collision probabilities can be obtained and analyzed. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. IM-ABS-861168

Matthew Brunner and Michael Kumsher

We analyzed several variables from an NFL data set, pertaining to all the NFL teams and covering 21seasons (2003-2023). We explored the probability distributions of these variables, trends over time, and the correlations between them. We also developed a regression model to see how the win percentage of a team can be predicted from factors such as total yards, turnovers, number of offensive plays, penalties, etc. Additionally, we compared the two conferences, and the four divisions within each conference. The overall purpose is to show that the use of statistical methods can help gain valuable insight into past and present performances in the NFL, and possibly facilitate data-driven decision making and strategy.

Mathias Boissevain

We model student progress towards graduation by using a Markov Chain methodology, where the states are the class standings of the students, as well as absent status. Graduation will serve as the absorbing state. The model leads to predicting graduation rates, mean times it takes students to graduate, as well as mean times spent in each class level. We also examine trends over time from 2009 through 2023.

Max Silver

Recent wildfires in Colorado raise the question of whether rising global temperatures have caused an increase in fire weather in Colorado. We use two datasets to address the question: "How has the occurrence of fire weather changed in Colorado?" Using 21 years of observed weather conditions from a meteorological tower at the National Renewable Energy Laboratory and 66 years of ERA5 reanalysis data, we assess changing trends in Colorado fire weather. The observational data are limited in temporal extent, but they capture exact real-world conditions at a location in complex terrain. The reanalysis data are available for an extended period of time and for the entire state, but the data are of relatively coarse spatial and temporal resolution and may fail to capture extremes.

To quantify fire risk, we calculate the hot-dry-windy index (HDWI), which relies on wind speed and vapor pressure deficit. No statistically significant trend in the HDWI appears in the observational dataset. However, according to the reanalysis data, strong increasing trends in HDWI values emerge across all of Colorado. This apparent conflict between observational and reanalysis data suggests that reanalysis data may not be representative and more long-term observational datasets are required to assess fire risk.

Zhixin Lun

Dealing with missing data problems for skewed data is a difficult task especially since most of the imputation and data augmentation methodologies assume multivariate normality. The performance of imputation and hence the accuracy of inference on parameters become questionable when the violation of the above assumption occurs. One approach to solve the normality violation is to apply normalizing transformation prior to the imputation phase. However, this approach may introduce new problems such as altering the dependence structure among random variables. We present a general purpose multiple imputation approach based on Copula transformation. The approach is used to effectively transform any continuous multivariate non-normal data to multivariate normal, thereby allowing the imputation using standard normality-based techniques. The method then allows to conveniently back transform the data into original space. Real data sets are used to illustrate the techniques. We then compare the performance of our Copula-based method with other traditional normality based multiple imputation approaches through extensive simulated and real non-normal multivariate datasets. We demonstrate that this method significantly mitigates the problem and hence the practice of making the blind assumption of multivariate normality for non-normal multivariate data under the assumption that data are missing at different mechanisms.

Ying Jin

Dynamic prediction, which typically refers to the problem of predicting future outcomes using the historical records from the same subject, is often of interest in biomedical research. For generalized function data, dynamic prediction turns out challenging due to the high density and complex correlation structure of repeated measures, which imposes heavy computational burden on traditional methods such as mixed models. Moreover, out-of-sample estimation of individual random effects is not feasible when the outcome value is not continuous. To address these issues, we developed a novel, fast, scalable method that combines the principals of mixed models and functional data analysis. This method, based on the fast Generalized Functional principla Analysis (fGFPCA) model, is able to handle large-scale, high-density repeated measures much more efficiently. Its model structure can accommodate great flexibility and estimate the out-of-sample individual-specific random effects without refitting the model. We designed a simulation study and a case study on the National Health and Nutrition Examination study (NHANES) data to demonstrate the predictive performance and computational efficiency of the proposed method, compared to existing methods that can achieve similar purposes. The results show that the proposed method can fit much more complex models much faster, also with greater overall performance than the other methods.