Data-Analytics Modeling to Detect Peripheral Neuropathy: Augmenting Big Data with Google Trends

Tomala, Neil (School: Parkway West High School)

I used Big Data (> 2.3 million records) to identify predictors of peripheral neuropathy (PN). PN costs the U.S. over 10 billion annually, so predicting factors may lead to large cost savings. Using patient data from the CMS, I used logistic regressions to identify factors that may detect PN. My model includes the variables Age, Gender, use of Dexamethasone, NSAIDs, Opioids, and the patient's location by state. Age is a significant predictor; increased age leads higher risk of PN. Gender is also significant; women are more likely to develop PN. Dexamethasone (steroid) significantly increases the risk of PN . NSAIDs (Non-Steroidal Anti-Inflammatory Drugs) are also significant predictors. My models show that the states with highest risk of PN are Illinois, Missouri, and Mississippi. I then used data from Google searches to examine if socially generated data have additional predictive power. While Age, Gender, drug use are information generated after a doctor visit, Google search data could show if patients who will later develop PN use certain search terms that can predict later manifestation of the condition. I use "Pain", "Tingling", "Numbness", as possible search terms. For each state, I model the association of these terms to later rates of PN, after controlling for all predictive factors. I find the correlations between "State Betas" and each search term to be positive, which shows that after eliminating the effects of the predictive factors, the location of the patient has additional information to predict PN. I validate that results in some counterfactual settings as well. Finally I show how cheap cloud computing can help facilitate such big data modeling. My estimates show that this exercise could be done in about $40, using Amazon cloud computing.

Awards Won: