Title: Predictive modeling approaches for predicting cognitive impairment in veterans and service members with a history of mild traumatic brain injury using longitudinal, observational data

Authors: Gayle S. Bieler, Barry Eggleston, Tracy L. Nolen, Nancy Temkin, Rick L. Williams

Background: The Chronic Effects of Neurotrauma Consortium (CENC) is conducting a longitudinal, observational study in veterans and service members with an aim of identifying predictors of neurocognition; occupational and social function; and a battery of motor, sensory, neurologic and psychological outcomes in the mTBI population with a particular focus on developing a predictive tool for chronic traumatic encephalopathy (CTE). A joint aim is to assess the impact of mild traumatic brain injury (mTBI) on these outcome measures. These aims present the challenge of identifying appropriate analytic techniques for prediction using longitudinal observational data while also making causal inferences about mTBI and CTE. This presentation provides an overview of analytic techniques currently published for prediction in the neurotrauma field as well as other viable techniques using longitudinal, observational data.

Methods: A literature review of high impact neurotrauma journals was conducted to identify predictive modeling and data mining techniques in current use in this field, assess the limitations of employed approaches, determine if there are other existing approaches to introduce into this field, and investigate modifications to existing approaches to meet the needs of the data being collected in the CENC. A programmatic search of article titles and abstracts on PubMed, http://www.ncbi.nlm.nih.gov/pubmed, was conducted and identified abstracts were vetted by the authors to select a final set for a more in-depth review.

Results: The analytic approaches employed for prediction in the published articles range from traditional statistical techniques including regression modeling to more sophisticated machine learning and feature selection techniques, including support vector machines, neural networks, principal component analysis, graph analysis, hierarchical modeling approaches, decision tree analysis, random forests, and singular value decomposition. Limitations of the approaches or their presentation include finding few examples employing predictive modeling approaches using longitudinal data, minimal discussion of handling informatively missing data, limited comparison of different approaches to develop the best model, and lack of exploration of how a predictive tool can subsequently be developed. Additionally, while some approaches employed have the ability for making causal inference, jointly addressing both causality and prediction and the issues that must be considered with using observational data for making causal inference has not been fully addressed in existing neurotrauma publications. Overall, our review of prediction in neurotrauma research show that predictive modeling is not used very often especially methods such as Ensemble methods like Random Forests, bagged classification/regression trees, or support vector machines which might be very useful in creating highly complex and accurate predictive models. Such models could be used as guides to identify acceptable simpler predictive models or used themselves as the numerical engines for diagnostic tools programmable into portable medical devices or a web-based application.

Conclusion: Prediction of cognitive performance is an important topic in neurotrauma research and as such, there are an increasing number of publications each year. This presentation provides an overview of available analytic options, particularly with respect to analysis of the observational CENC study; however, a comprehensive summary of available approaches as well as their benefits and limitations would be beneficial.