Derivation, Validation, and Testing of Novel Prediction Model to Identify Severe vs. Non-severe Epilepsy Patients

Chong H. Kim1, Ryan Hansen2, Sean Sullivan2, Jonathan D. Campbell1

*1 Department of Clinical Pharmacy, University of Colorado Anschutz Medical Campus*

*2 School of Pharmacy, University of Washington, Seattle, WA, USA.*

**ABSTRACT:**

*Objectives*: There is no known Internationally Classification of Disease version 9 (ICD-9) and ICD-10 codes that are used for certain forms of severe epilepsy such as Dravet Syndrome (DS). Research is needed in order to identify patients regarding the severity of epilepsy. This study aims to develop and validate a prediction model that can be used to predict severe and non-severe forms of epilepsy.

*Methods*: The study sample consist of clinically confirmed DS patients in addition to two comparator populations (n=189) that include a mild (i.e. childhood absence epilepsy) and severe (i.e. infantile spasm and Lennox Gastaut) form of epilepsy were included from Children’s Hospital Colorado that had at least one year of active retrospective follow-up in the electronic medical record. Data were pre-processed such that continuous variables that were included were centered and scaled, variables that had near-zero variances were removed, and Box-Cox transformation was applied. Four different models (i.e. Classification and Regression Tree, Random Forest, Support Vector Machine, and Logistic Regression) were used to classify the patients as Severe vs. Mild. The models were derived and validated using 124 of the sample and tested using 65 subjects.

*Results*: The Random Forest and logistic regression algorithm yielded the highest AUC (0.856 and 0.814) compared to that of the other classification algorithms using the test cohort. The sensitivity for Random Forest and logistic regression algorithms were 0.643 and 0.714 while the specificity 0.9412 and 0.784, respectively. Using Random Forest Algorithm, the top five variables with high variable importance scores were prescription count (100), laboratory count (57.04), number of chronic comorbidity conditions (54.33), frequency of clobazam prescribed (49.25) and insurance status (42.69).

*Conclusions*: The Random Forest algorithm identified and predicted severity of epilepsy better than other algorithms using medical and pharmacy claims data. Additional prospective validation of the prediction model is necessary before clinical implementation.