

PH1976 Fundamentals of Data Analytics and Predictions

Spring 2019 - Class Project

Predicting Mortality for Subarachnoid Hemorrhage Patients

The course project can be done in groups of 2 or 3 students. The project will consist of an in-class oral presentation by two or more of the group members. The day of the presentation, the instructor will randomly choose which of the members will present the project so all members should be prepared to present.

- Oral presentation. 15 minutes (10 minutes for presentation, 5 minutes for questions).
- Detailed report of your work with introduction, methods, results (with illustrations), and discussions with a maximum of 5 pages.
- Program source code
 - Make sure to include enough documentations and comments in your code. An R Markdown file is preferable.
- Test set predictions, an R dataset that has one column with the patient ID numbers and the other column with their predicted outcome of mortality.

The report, code, and predictions are due by midnight on April 18.

The project will be graded as

- 10% internal group evaluation,
- 40% Written report as graded by the professors,
- 30% oral presentation as graded by the professors,
- 10% oral presentation as graded by fellow students,
- 10% based on algorithm performance on the left-out test set based on the ranking among other groups (best team take full points (10/10) and worst performed team take (0/10)).

Every student will fill out an evaluation form of every other presentation as well as an internal evaluation of your group members. Attendance of other group's presentations is mandatory. Evaluations will be anonymous, but each student will be checked for submission/attendance by TAs before submitting the form. If a student does not submit attendance, that student's final project grade will be deducted by 3 points.

Once you form your group, send all names in the group to the TAs who will assign a group number.

The talk will be split up into a background/methods section presented by one randomly chosen person and the results/discussion section presented by a different randomly chosen person from the group.

The goal of this project is to predict in-hospital 'death' of Subarachnoid Hemorrhage (SAH) patients using all patients' records in the first 3-5 days from the start time of the hospitalization. Inputs of prediction includes patients' encounter data, medications, clinical events, and lab tests. A test set will be given but the outcome will not be included in the dataset provided to students.

Data

A link will be sent out on Canvas.

Definition

Subarachnoid hemorrhage (SAH) is sudden bleeding into the subarachnoid space. The most common cause of spontaneous bleeding is a ruptured aneurysm. Symptoms include sudden, severe headache, usually with loss or impairment of consciousness. Secondary vasospasm (causing focal brain ischemia), meningismus, and hydrocephalus (causing persistent headache and obtundation) are common. Diagnosis is by CT or MRI; if neuroimaging is normal, diagnosis is by CSF analysis. SAH is a potentially devastating cerebrovascular condition due not only to the effect of initial hemorrhage but also the complicated treatment techniques required to manage such patients.

The main treatment for SAH patients, beside IV fluid, are three common Vasopressors (dopamine, phenylephrine, or norepinephrine), helping to elevate blood pressure. Patients undergoing these three treatments are considered more severe cases and comprise almost 10% of SAH patients. There are no global guidelines for distinguishing these patients.

Data Source

In this study, we query the Cerner Health Facts® EMR dataset for SAH patients. Cerner Health Facts® EMR database comprises de-identified EHR data from over 700 participating Cerner client hospitals and clinics in the United States including patients demographics, encounters, diagnoses, procedures, lab results, medication orders, vital signs, and other clinical observations. This database is useful for performing in depth retrospective studies on relatively rare disorders with statistically significant population samples.

Study Design

The effective population queried included 4,838 total patients admitted to any hospital using Cerner as their EMR from the years 2000 to 2015. Only adult patients with a new diagnosis of aneurysmal (non-traumatic) SAH based on ICD-9 code 430 for Subarachnoid Hemorrhage, with no pre-existing subarachnoid hemorrhage diagnosis were included; furthermore, only patients who received either dopamine, phenylephrine, or norepinephrine infusions were included.

The patients were clustered to one of three subsets based on the choice of vasopressor or a combination of multiple vasopressors where 1=dopamine, 2=phenylephrine, and 3=norepinephrine. 67.8% of all patients have been treated with only one vasopressor choice, 27% and 5.2% for two and three vasopressors, respectively. The mortality rate increases with the number of vasopressor, from 35.4% for one vasopressor, to 47.5 and 57% for two and three vasopressor, respectively. It seems that patients treatment is started with a choice of one vasopressor, then if the results (blood pressure level, or patients' condition) is not satisfactory, the vasopressor is re-administered or changed to another one.

Study Population

All patients over the age of 17 were included, with comorbid conditions including anemia, edema, heart failure, myocardial infarction, pneumonia, acute renal failure, and sepsis. Patients who had a diagnosis which included a traumatic cause of SAH will be excluded from the study population (ICD 9 codes: 800.2x ,800.7x ,801.2x ,801.7x ,803.2x ,803.7x ,804.2x , 804.7x, 852.x)

Outcome

The primary outcome was mortality, which was defined as death or discharge to hospice care.

Task

Build a machine-learning model to predict mortality in subarachnoid hemorrhage patients.