Building a Safer Health System

The report for Building a Safer Health System came with a lot of data to examine. There’s close to 32,992 reports in the dataset. A lot of reports for the ICU data errors. It outlines how an adverse event is defined as the result of an unintended harm to the patient by an act of commission or omission rather by the underlying diseases or condition of the patient and that’s what we see here. Our objective is determining the circumstances that would result in a patient receiving an improper dose/quantity of medication.

The report contains report numbers, facility ID, error categories, date of error, contributing factors, location of error, description of error, medication process nodes, staff type initiated errors, staff type perpetuated errors, staff type discovered errors, actions taken, actions taken details, and date records of when entered.

When we start off examining the data, we notice that there’s an empty column in the data, so we remove that column and make a new subset into a new variable. After using the summarize function on the new data, we can see that all classes and methods are of type character except the facility ID and report numbers. The facility ID and report numbers return descriptive statistics such as the minimum, first quantile, the median, mean, 3rd quantile, and the maximum value of our input data but those statistics don’t say much. The structure function does a little more than the summary function by giving us the data types but also with a few observations of the features.

To try and get some meaningful data we decided to factor a lot of the data features like dates, error categories, type of errors, cause of errors, contributing factors, medication process nodes, staff type initiated errors, actions taken, staff type discovered errors, and days of the week.

After factoring the data, we decide to summarize it to see our new findings. By summarizing, we’re now able to see the frequency of the factored data. We come across 7 categorical errors with them being distinguished as characters, A, B, C, D, E, F, and H. 2,218 were of type A, 6,536 were of type B, 16,458 were of type C, 5,660 were of type D, 1,353 were of type E, 147 were of type F, and 620 were of type H. Categorical error type F was the least type of categorical error and type C was the highest.

Dates of error start on January 1st and end on December 3rd of 2020 but we also get 19,079 reports that have no dates reported as NA’s. From the reports for day of the week Monday has 4,981, Tuesday has 5,968, Wednesday has 4,931, Thursday has 4,890, Friday has 5,423, Saturday has 3,094, and Sunday has 3,094. It seems like on Saturdays and Sundays the number of reports decreases and jumps back up on Monday to follow Tuesday.

Types of errors consists of omission errors, improper doses/quantities, unauthorized drugs, prescribing errors, extra doses, others, and NAs. Omission errors being cited with 7,404, improper dose/quantity with 6,319, unauthorized drug with 3,752, extra dose with 2,014, other marked as 8,864, and NA’s with 2,384. We know that the least type of error is the extra dose but the most is marked as other leaving us without a specific type of error.

To study why these types of errors occur, we can see the causes of these errors by seeing the frequencies as well. Performance (human) deficit marked 7,661 times, procedure/protocol not followed marked with 3,927 times, transcription inaccurate/omitted with 1,875, communication with 1,853, knowledge deficit with 1,801, other as 13,485, and NA’s as 2,390. Knowledge deficit by being the least cause of error but “other” being marked as the highest with 13,485.

Contributing factors to these reports start off with distractions marked as 5,383, 1,912 by workload increase, 1,604 by inexperience staff, 1,005 by insufficient staff, 971 because of shift changes, 5,661 as other, and NA’s having the highest number of reports with 16,456.

Medication processed nodes in the report consist of administering with 12,349, dispensing with 5,976, documenting with 7,614, does not apply with 2,218, monitoring with 447, and prescribing with 4,388.

Now we get to the interesting findings of staff type that initiated the errors. Registered nurses initiating the most with 13,053, pharmacists with 4,672, physicians with 3,048, pharmacy technicians with 1,727, unit secretaries or clerks with 1,510, other marked as 6,057, and NA’s as 2,295.

Staff type perpetuated errors were of registered nurses, pharmacists, pharmacy technicians, physicians, unit secretary clerks, other, and NA’s. Nurse registered with 8,297, pharmacists with 2,661, pharmacy technician with 841, physician with 805, unit secretary clerks with 743, other as 2,794, and NAs with 16,851.

In the staff type discovered errors, 16,789 registered nurses discovered errors, 4,686 were pharmacists, 1,968 were physicians, 1,048 were licensed practical nurses, 992 were pharmacy technicians, other as 2,659, and NAs as 4,850.

When finding errors, it’s important to take action on these errors. With actions taken for these reports, 9,766 informed staff who made the initial error, 3,975 informed staff who were also involved in errors, 2,937 were provided with education or training, 2,119 were marked as none, 1,812 had communication processes enhanced, 1,371 marked with “other” as type of action taken, and 11,012 marked as NA’s.

After these findings, we believe it’s time to start plotting and making charts to see these results a different but easier way. So, we begin to make use of the ggplot function with aesthetics, geometric bars, and geometric texts to better see these findings in graphs.

We’re able to plot a bar graph with x-values as days of the week to display the amount times of errors occurring in a particular day of the week. Next, we do the same kind of plotting but with errors of categories in the x-values making it easier to see just how much each error category contributed to the reports. Especially error category C which can be seen skyrocket compared to the other error categories.

Next, we make a new variable called improper doses ICU making it a subset of the new ICU data we had made from the original data so we wouldn’t change anything to the original. We do this to filter the type of error to when it only matches with improper dose/quantity in the improper dose ICU variable. Then again, we make a subset of the improper dose ICU data but to itself to update it with the matching filter plus the type of error, error category, day of the week, contributing factor, medication process node, and location of error. This way we have an accurate depiction of the data and details of the other features to when the type of error matches the improper dose/quantity.

Now we intend to use the Naive Bayes method to analyze the description of each report to determine the keywords that would identify improper dosage medicals errors. First, we create variable that holds a subset of our original data. The variable *ICU\_data* contains the column *Type-of-error* and *Description-of-error*. We then make a new column called *isImproperDose* which holds the text ‘Improper dose/quantity’ if in column *Type-of-error* matches the statement and other if a different type of error occurred. Now that we no longer have a need for the *Type-of-error* column and drop the selected column. The next step is to clean and analyze the error description using corpus function from the tm package. We create a classifier using the *naiveBayes* function. The *naiveBayes* function makes an assumption of conditional independence between every pair of features given the value of the class variable used to make predictions. Now we can make a cross table comparing predictions based on the keywords from corpus created from *Description-of-error* column. In the cross table we can see that our prediction was only able to accurately find 45.9% of improper dosage accurately based on 2148 total cases of improper dosage recorded. In contrast, we can correctly predict 82.6% of the 7129 confirmed *other* cases. Now we know that one can accurately predict type of errors that are not caused by improper dosage reliable we can use that to make estimations based on the circumstance of their origin.