

DATA 698: Final Project

Thomas Hill

2022-05-17

Abstract

Adverse drug events (ADE's) are a serious consequence of medication therapy. Previous research has associated adverse events with poorer outcomes, and suggested some patients are at higher risk for such events. Using hospital admission records from New York State in 2018, the most common adverse events were obtained, and the overall risk of an ADE were related to the number of diagnoses, length of stay, as well as some chronic illnesses. A model for predicting risk based on patient characteristics was developed, with a tree-based stochastic gradient model performing the best. However, all four models suffered from F1 and F2 scores less than 0.5.

Background

Adverse drug event (ADE) is the comprehensive term for the negative effects related to medication therapy. ADE's may be unforeseen or known and vary in severity, from untoward changes in laboratory values to injury and death. Some medications have therapeutic dosing close to their corresponding toxic dose, which is a characteristic described as narrow therapeutic index. In these situations, prospective monitoring may be warranted.

In the United States, the Food and Drug Administration manages the publicly available FDA Adverse Event Reporting System to collect voluntary reports from consumers and healthcare practitioners. In contrast, New York State mandates that adverse events occurring in hospitals and diagnostic centers are reported.

The US government has a program offering access to detailed hospital admissions data through its Agency for Healthcare Research and Quality (AHRQ). It combines hospital and admission characteristics as well as information about charges. The research done using these data is extensive, including applications to monitoring ADE's.

Dataset

The dataset is available through AHRQ's healthcare cost and utilization project (HCUP), which offers several options for obtaining admissions information. There are annual reports spanning the entire United States, as well as more detailed reports for each state. Datasets for each year are further broken down into pediatric and adult, as well as inpatient, emergency, ambulatory, and readmissions data. The most recent dataset available at the time is 2018, and while the variables reported are mostly uniform there are additional admission characteristics that are reported in some states.

Data are available in two separate files: CORE and CHARGE files. The CORE file represents patient characteristics, timing of admission and length of stay; it also includes diagnoses and procedures as described by the International Classification of Diseases (ICD-10) as well as diagnosis-related groups (DRG's) used extensively by Medicare. The CHARGE file contains dollar and unit amounts of services rendered during each individual admission. There is also a Healthcare Common Procedure Coding System (HCPCS) code as an additional descriptor of the hospital cost center where the charge originated. Each observation in both files has a unique admission number designated by the KEY attribute. In total, there were over two million unique admissions with over twenty million charges in 2018.

These data are also available for free via HCUPnet, a data query tool accessible on the AHRQ website. No API is available for repeated or programmable querying. HCUPnet returns counts and rates for selected diagnoses by year and region but remains useful to researchers to confirm the validity of their own statistics.

Methods

The two main files were first loaded into SPSS to be translated using HCUP's provided loader program. These were then converted to .csv format and loaded into a local SQL database. The CORE file for New York state has over 250 features, many of which are used to hold ICD diagnosis codes and dates for ICD procedure codes. Based on previous research, there are also over 500 codes located in described in ICD -10 are correlated with an ADE. The strength of associations is described in Table 1. Admissions were determined by filtering ICD codes where an ADE was likely present.

There were several additional databases used for labeling purposes. A .csv with all available ICD diagnosis codes and subcodes helped to interpret diagnoses. There was also a corresponding list of all ICD procedure codes. A ny.gov list of FIPS codes was used to translate patient location into county and then regions. Finally, for the CHARGES file, I reformatted a table available on New York's Department of Health website for descriptions of each revenue code.

Feature Engineering

The dataset required extensive transformation to better visualize and explore each feature. ICD diagnosis codes are located in 35 columns, with additional ones indicating whether each diagnosis was present upon admission. Similarly, each ICD procedure code filled one of 25 columns, plus 50 additional columns specifying the date and year. The columns containing the codes were converted to one-hot encoded columns for the most common codes. Diagnoses were sometimes specific, e.g., 'G40' for epileptic seizure, or more general by searching for all codes starting with 'G' indicating diseases of the nervous system. Procedures were handled identically, with the added step that procedures were counted more than once if repeated. For example, if the patient received three blood transfusions, the value would be '3' rather than '1'.

New features also included updating the hospital location by using hospital NPI and FIPS codes. These were converted to one of ten greater New York regions based on county. The month and time of admission were also refactored, but because of incomplete information about day of admission, they were not able to be considered as a date-time object.

Missing Values

For blood pressure and heart rate, values were available for a minority of observations. Instead of ignoring these data, each was refactored according to clinical categories, like Stage I Hypertension or bradycardia. In the case that patient location was unavailable via FIPS code, the hospital NPI was used as an approximate location by manually replacing with the closest zip code. There were also several observations missing both NPI and FIPS, as well as date of admission. These observations were omitted altogether, and only encompassed less than 5% of the total dataset.

Code Libraries and Algorithms

Programming was performed in R. Specific libraries include dplyr for preprocessing and descriptive statistics, and ggplot2 for data visualization. To correct for class imbalance, the ROSE library was used to oversample from observations with an ADE. Cross-validation and supervised learning techniques were automated using caret to predict the presence of an ADE. The machine learning models considered were logistic regression via generalized linear model (GLM), stochastic gradient boost (SGB), random forests (RF), and neural nets (NN).

```
## Warning in recode.numeric(.x, !!!values, .default = .default, .missing
## = .missing): NAs introduced by coercion

## Warning in recode.numeric(.x, !!!values, .default = .default, .missing
## = .missing): NAs introduced by coercion
```

##-Average LOS

```
ade_cost <- ade_flag %>% select(has_ade, LOS_X, TOTCHG_X)
ade_cost$LOS_X[ade_cost$LOS_X == 0] <- 1 #change 'zero' length of stay to '1'

ade_cost %>%
  summarize(AVG_LOS = mean(LOS_X, na.rm = TRUE))

ade_cost %>%
  filter(has_ade ==1) %>%
  summarize(AVG_LOS = mean(LOS_X, na.rm = TRUE))

ade_cost %>%
  summarize(med_los = median(LOS_X, na.rm = TRUE))

ade_cost %>%
  filter(has_ade ==1) %>%
  summarize(med_los = median(LOS_X, na.rm = TRUE))
```

Results

Baseline Characteristics

Demographic and admissions statistics are available in Table 2. Adverse drug events were present in 2.12% of the over two million inpatient admissions in 2018. Patients with ADE's differed from the total population in age and the route by which they were admitted to the hospital. Cases were on average 4 years older than the total population, while only 44% (versus 67%) were admitted via the emergency department (ED). There are a few ways to be admitted to the hospital other than the ED. Patients could have been more likely to undergo elective procedures, including labor and delivery. Another explanation is that these patients are transfers from nursing homes or skilled nursing facilities and were referred for observation and bypassed the ED altogether.

ADE cases also have much greater hospital charges associated with admission, approximately \$105,000 versus \$57,000. This pattern disappeared when considering charge per day, with the total cost almost wholly driven by their longer duration at the hospital (six versus three days).

2018 New York Hospital Admissions, Baseline Characteristics

Characteristics	Total Population	Target Population
n	2,018,259	42,829

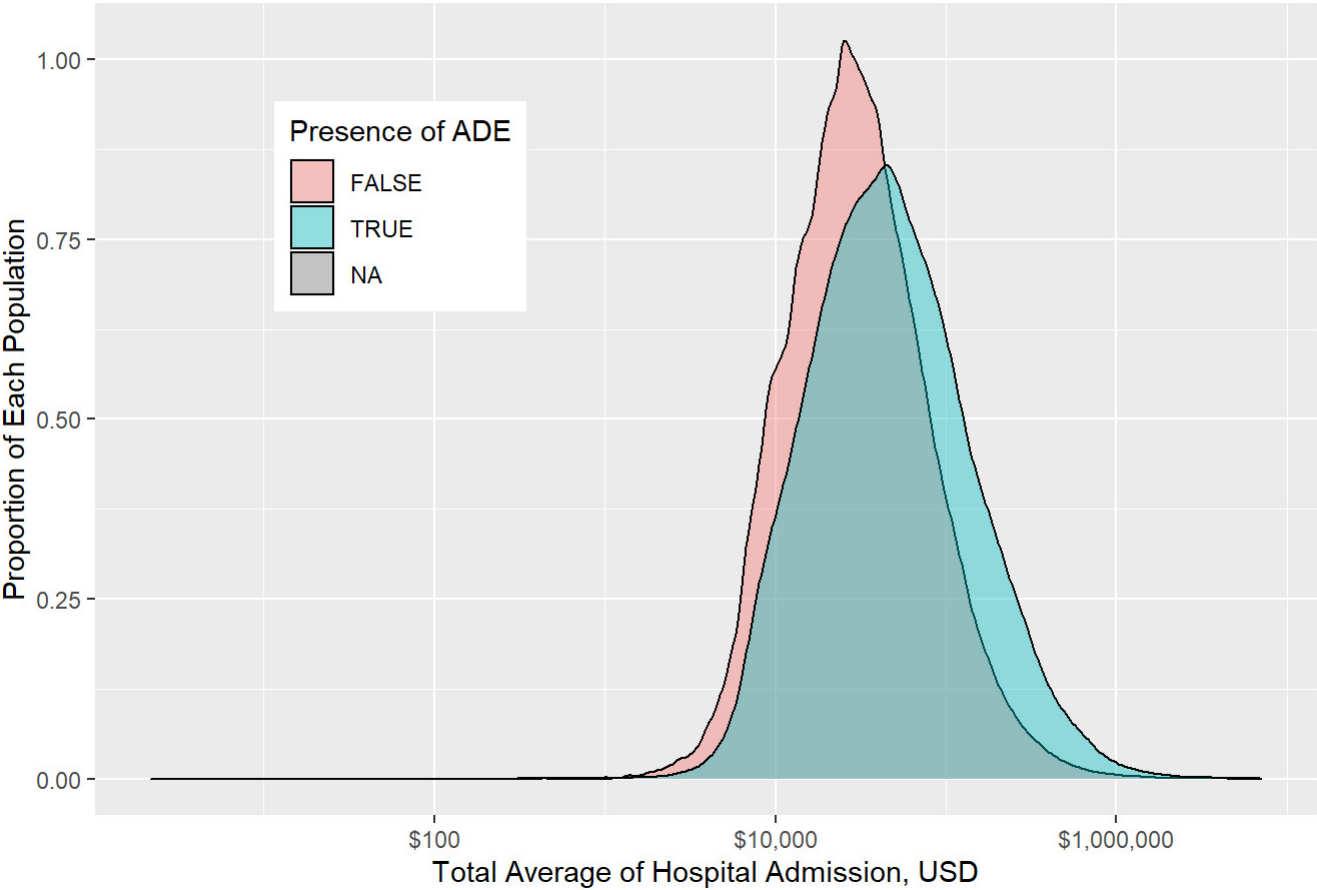
2018 New York Hospital Admissions, Baseline Characteristics

Age (years)	58.3	62.2
Female (%)	56.2	48.9
Race (%)		
White	54.6	59.7
Black	17.1	13
Hispanic	13.8	13.1
AAPI	4.6	4.1
Native American	0.2	0.7
Other/Missing	9.7	9.7
Hispanic (%)		
Not Hispanic	81.6	83
Hispanic, White	3.2	2.8
Hispanic, Black	0.7	0.6
Hispanic, Other Race	9.8	9.8
Missing	4.6	3.9
ED Admission (%)	67	43.8
Died (%)	2.4	4.8
Length of Stay (days)	5.7	9.2
Median Length of Stay (days)	3	6
Number of Diagnoses	12.1	18
Number of Procedures	2.1	2.9
Charge (\$1K)	56.8	105
Median Charge (\$1K)	32.1	51.9

2018 New York Hospital Admissions, Baseline Characteristics

Charge per Day (\$1K)	13.4	12.1
Median Charge per Day (\$1K)	9.4	9.9

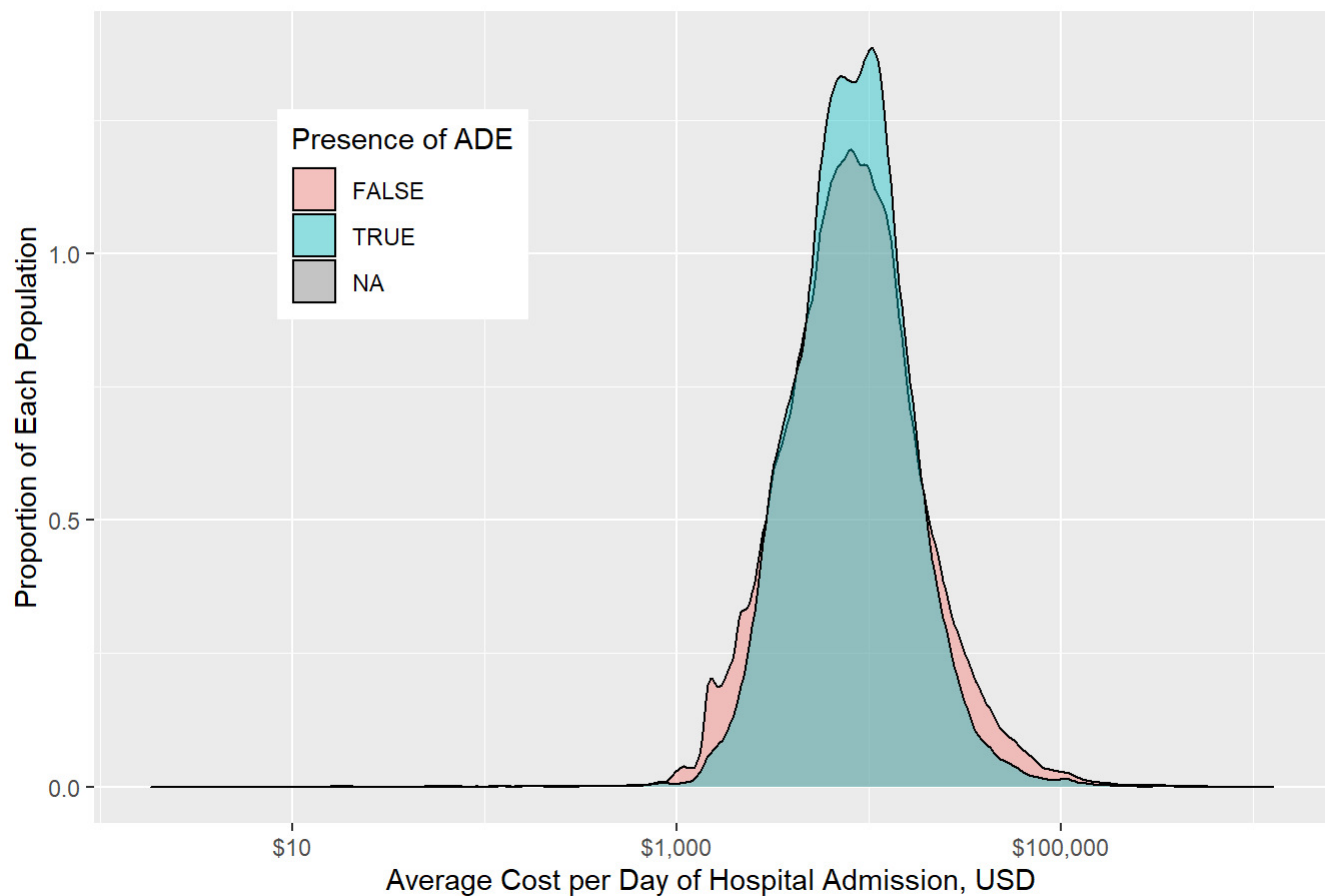
Average Cost Per Inpatient Visit



```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```

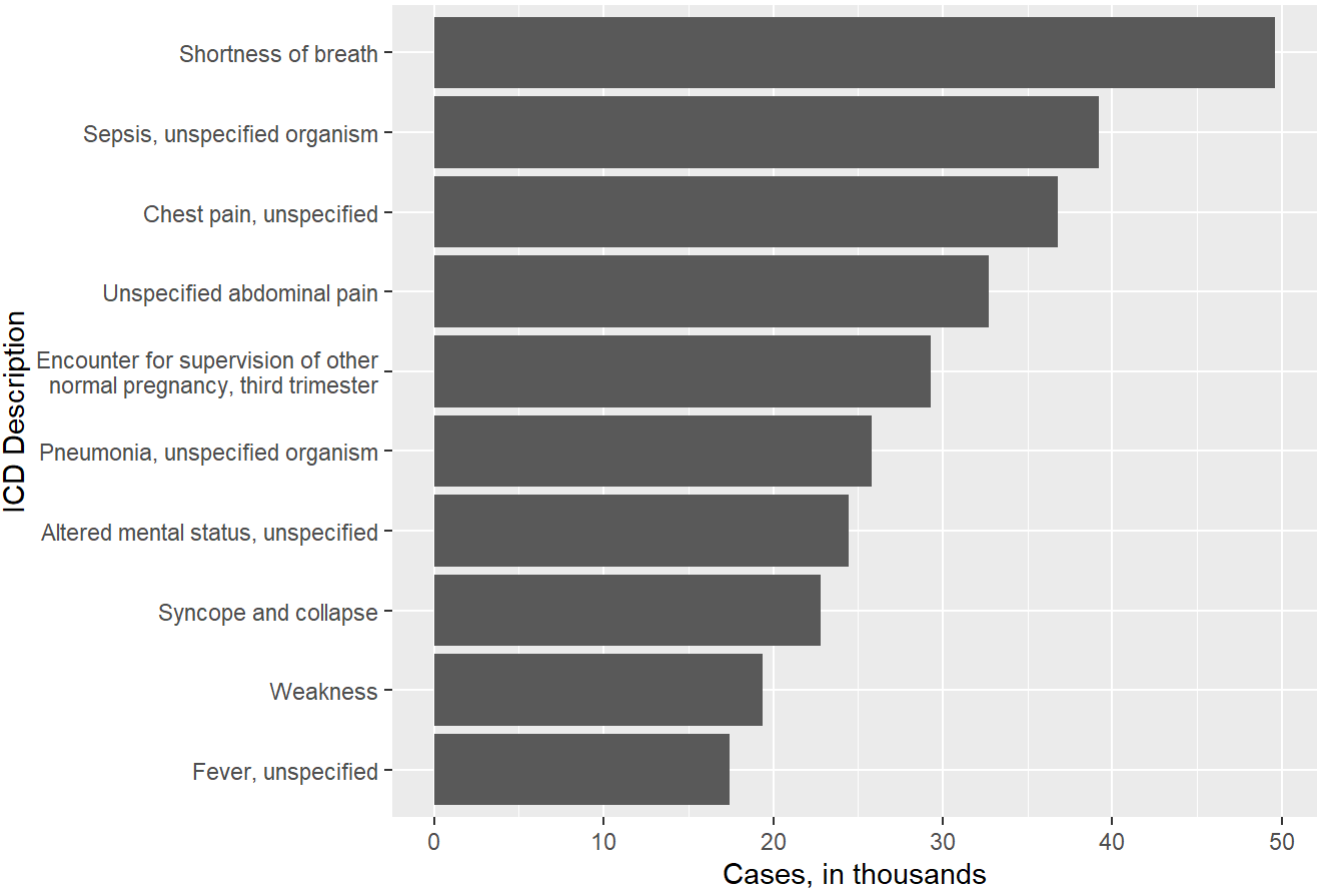
Average Cost per day of Inpatient Visit



ICD-10 Diagnoses and Procedures

ICD-10 diagnoses were limited to 35 and procedures to 25; however, there appeared to be a local mode at 25 diagnoses. This may be attributed to a limit on claims to large insurers like Medicare. The most common admitting diagnoses coordinate well with clinical practice; many of them are admission for childbirth (encounter for supervision of other normal pregnancy, encounter for full-term uncomplicated delivery), infection (sepsis, pneumonia, fever), or one of several chronic illness exacerbations (shortness of breath, chest pain).

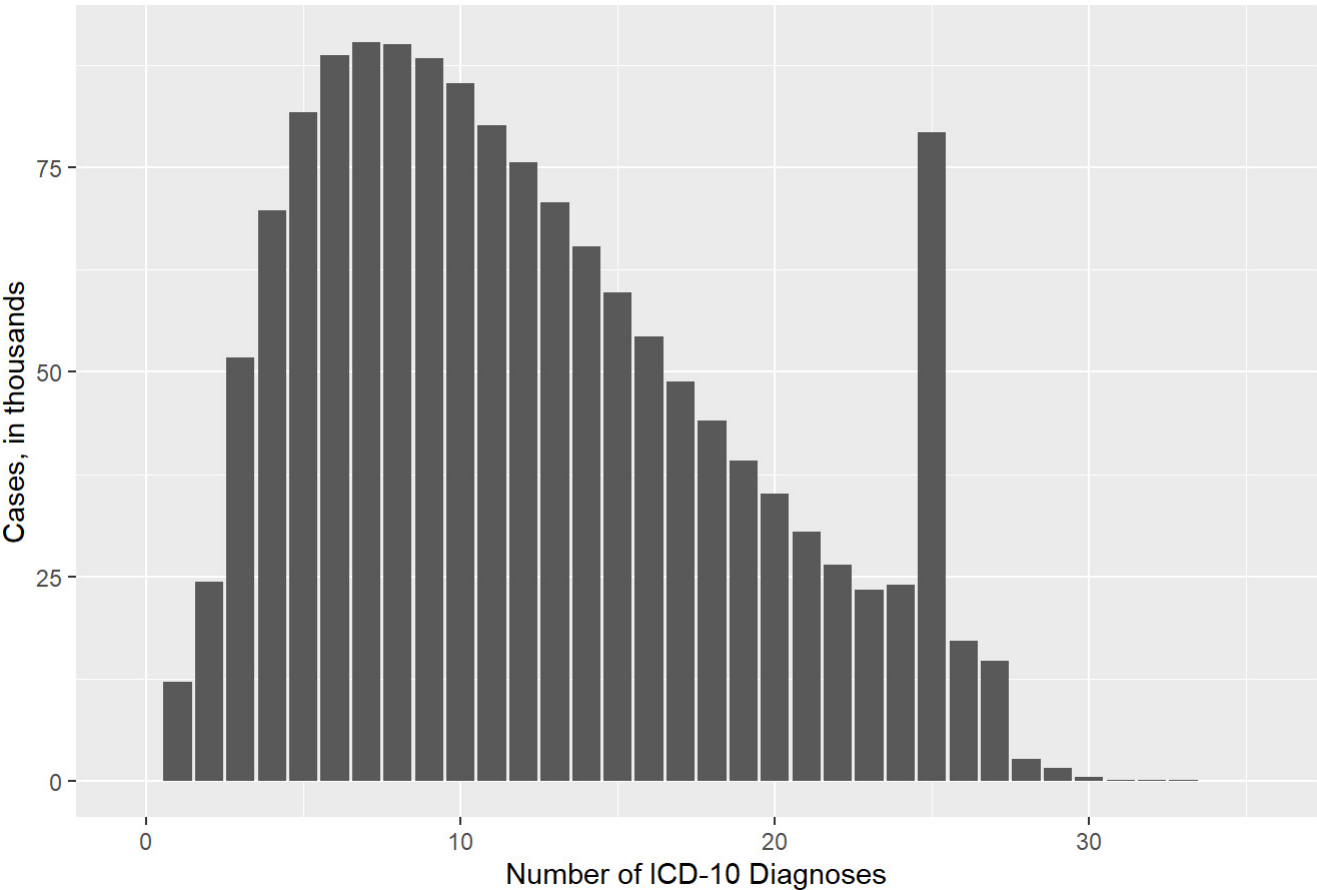
Top 10 Admitting Diagnoses, New York Hospitals 2018



Total Number of Diagnoses

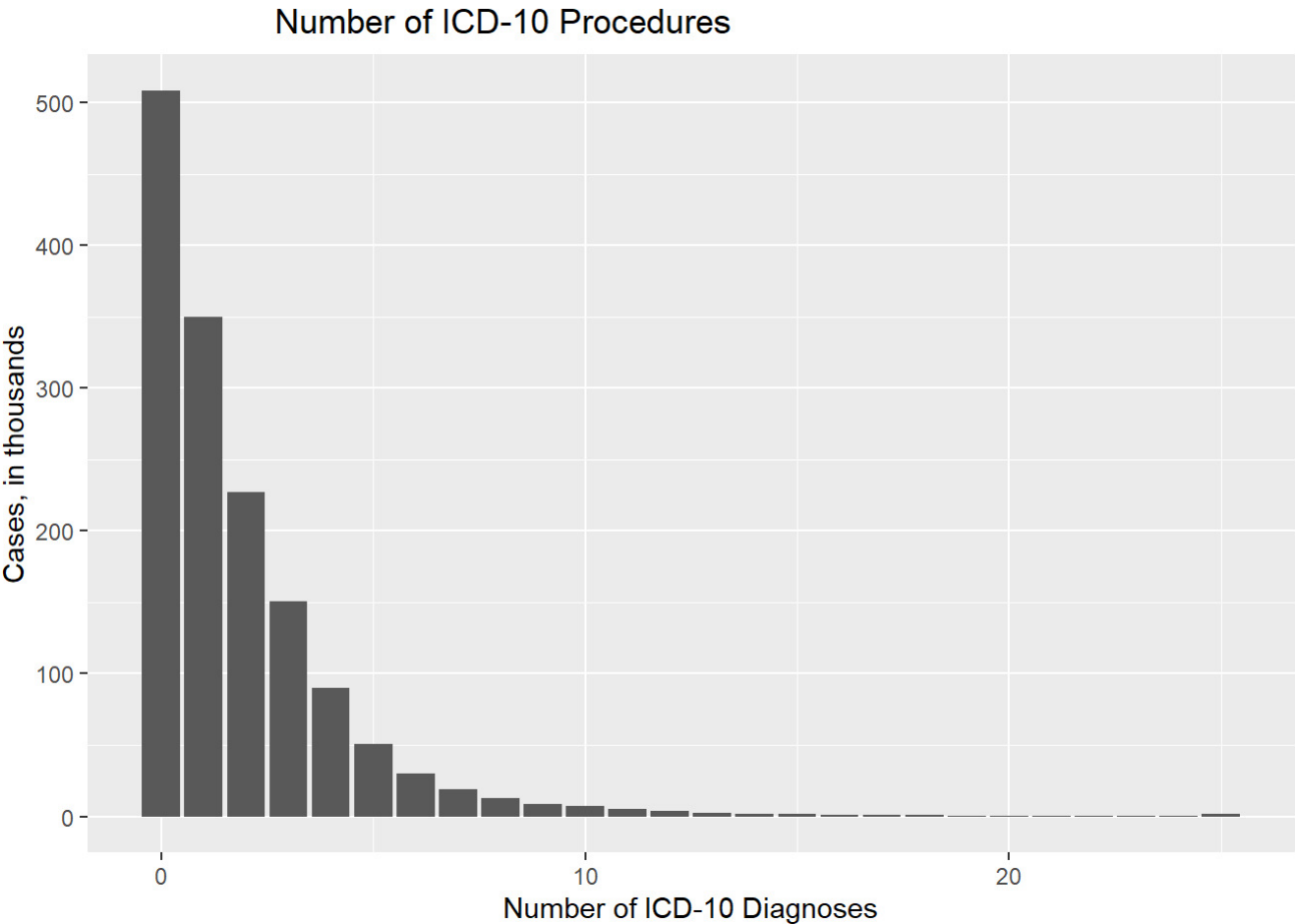
```
## Warning: Removed 1 rows containing non-finite values (stat_count).
```

Anomaly in the Number of Diagnoses

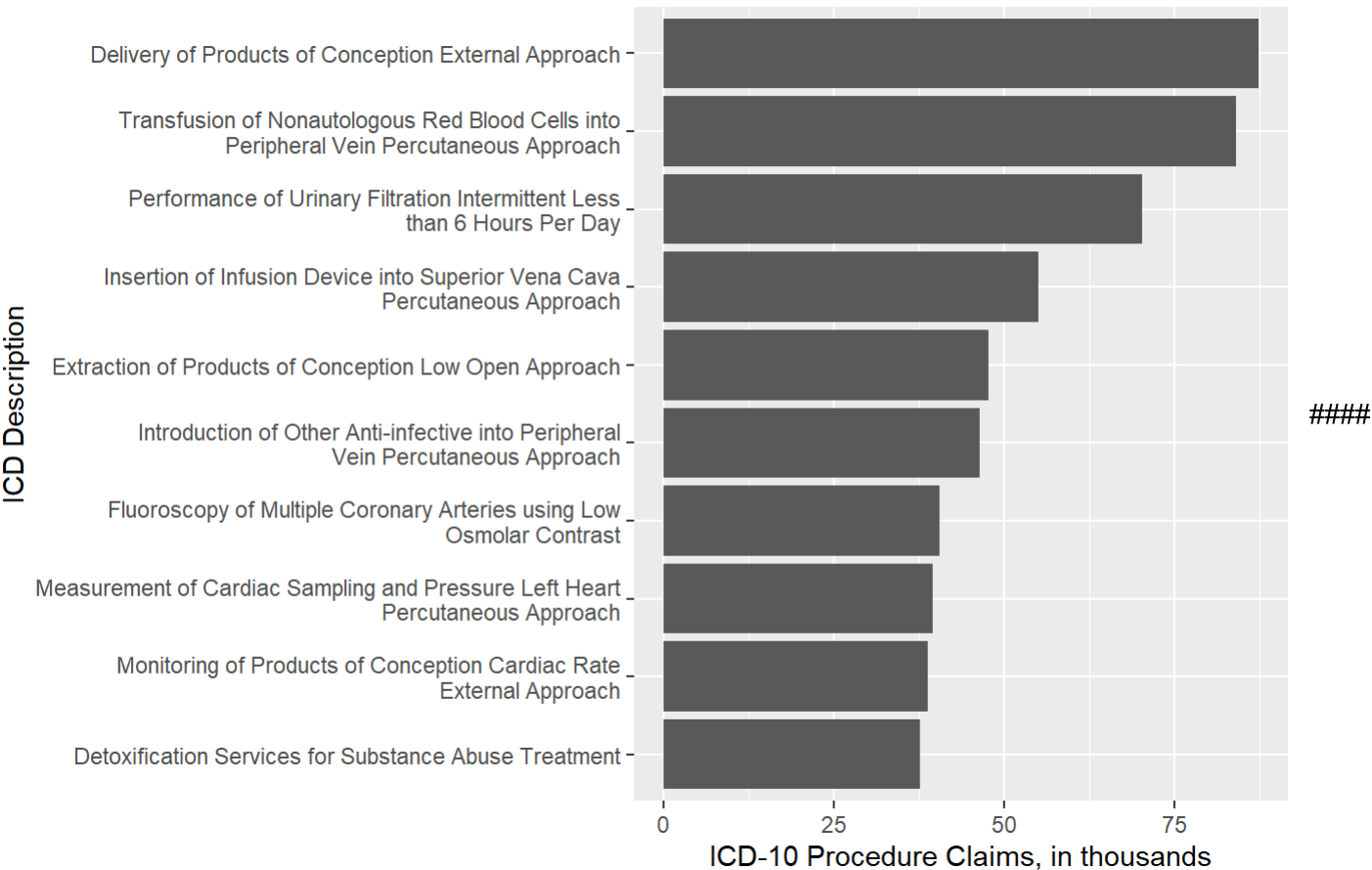


Most Common Procedures

```
## Warning: Removed 1 rows containing non-finite values (stat_count).
```

Top 10 Most Common Procedures



Most Common Adverse Drug Events

Overall, the most common adverse drug events were split between categories A and C, caused by a drug or ADE likely. The most common code was for secondary thrombocytopenia, with 15,000 reported cases. This is one of the ICD’s many ‘other’ categories, so it is possible that thrombocytopenia is truly unknown or misattributed to the administration of a drug. Otherwise, thrombocytopenia is a common side effect for antineoplastic agents, and also seen when giving heparin-based anticoagulants.

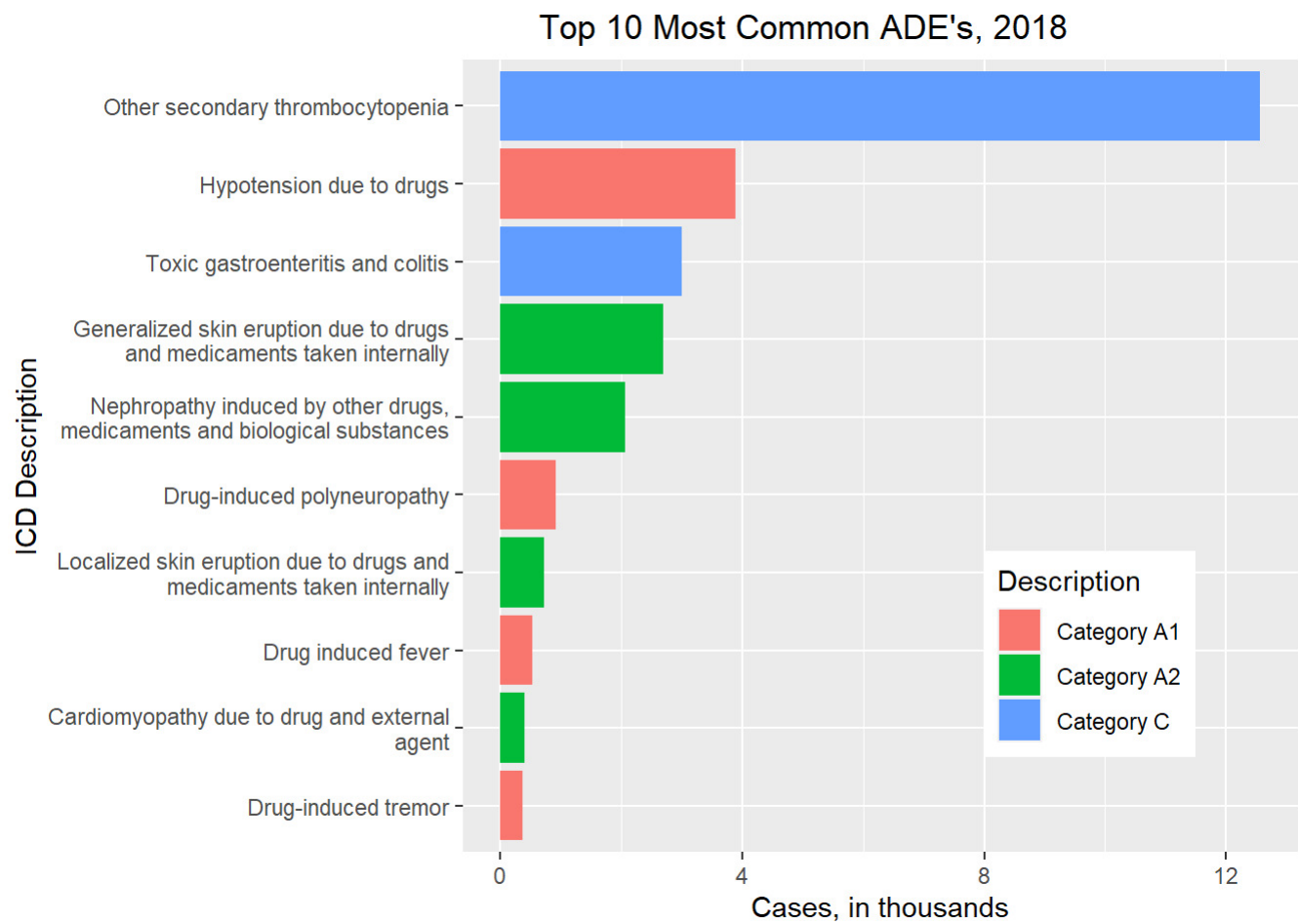
The top Category A codes vary between 500 and 5,000 cases, with the most common diagnosis being hypotension due to drugs. There are several drug classes that lower blood pressure, which are known to cause falls and hospitalization in geriatric patients. In this case, it’s possible that the dose was too high or postural/orthostatic hypotension was an issue.

Next, generalized skin eruption due to drugs may be a combination of a few underlying issues: some medications do increase sensitivity to sunlight or cause rare life-threatening conditions like Stevens-Johnson Syndrome or toxic epidermal necrolysis. More likely is that this code is used for patients suffering from anaphylactic/ allergic reactions from medications. This is made more likely by considering these are hospital admissions, and mild skin conditions could be handled outpatient by discontinuing the offending agent. Polyneuropathy is yet another condition that could be caused by certain antineoplastic agents, or also associated with poorly controlled diabetes mellitus.

Category B is the smallest category represented in the total population, with only several dozen associated reactions. The official definition of a Category B reaction is an event caused by ‘poisoning’; however, the descriptions illustrate something else happening. The most common category B code is laxative abuse, which is an intentional overuse of laxatives potentially related to an attempt at weight loss. Category B codes detected in this study are all ICD codes starting with ‘F’, or mental/behavioral disorders. These codes may not be widely used because of their high specificity and the indirect causal relationship they may have to hospitalization. For instance,

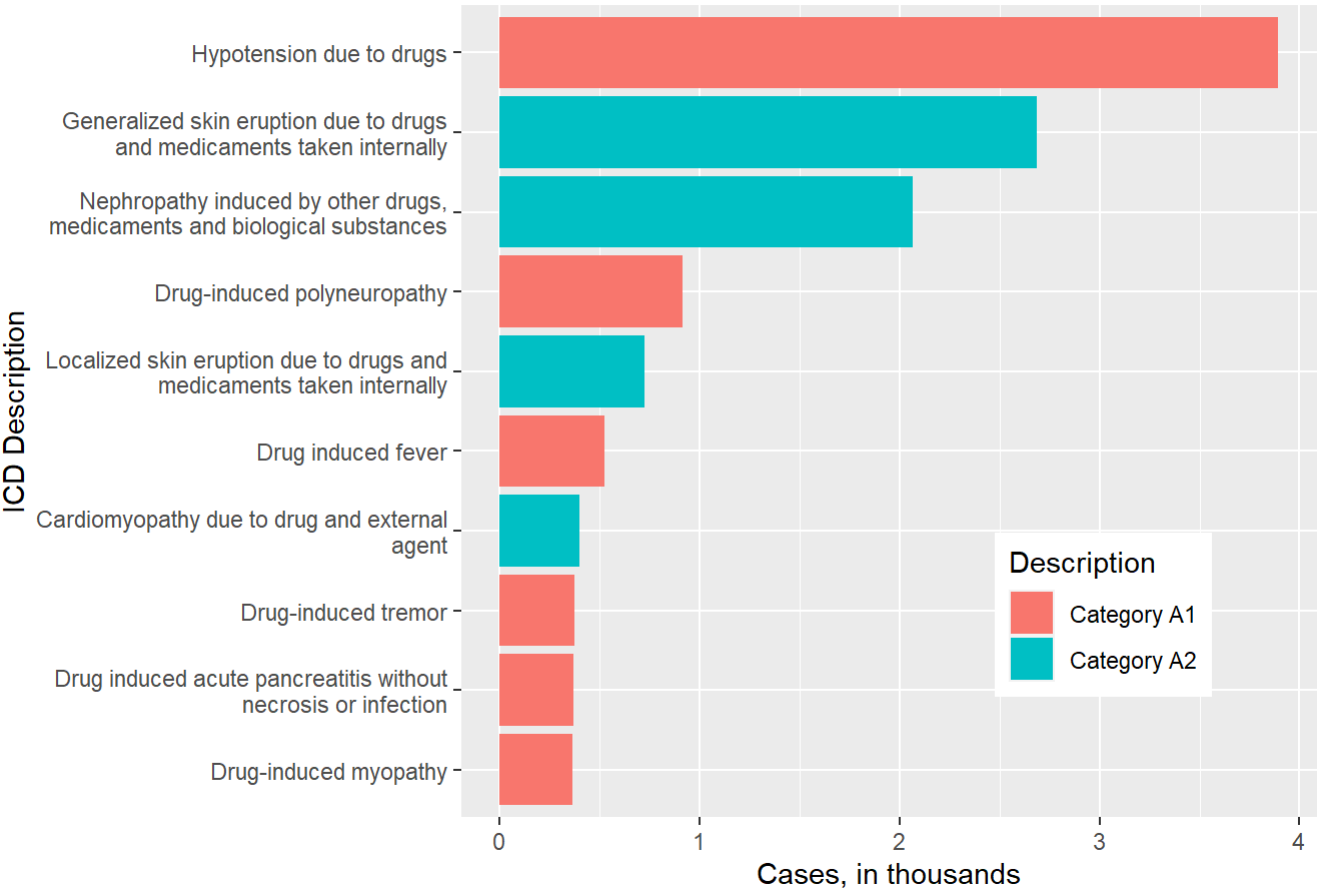
abuse of laxatives may present in the ED as a chief complaint of diarrhea, electrolyte or acid-base imbalance, or a more general mental health diagnosis. It is not until a patient is interviewed that the underlying cause is established.

```
## `summarise()` has grouped output by 'Code'. You can override using the
## `.groups` argument.
```

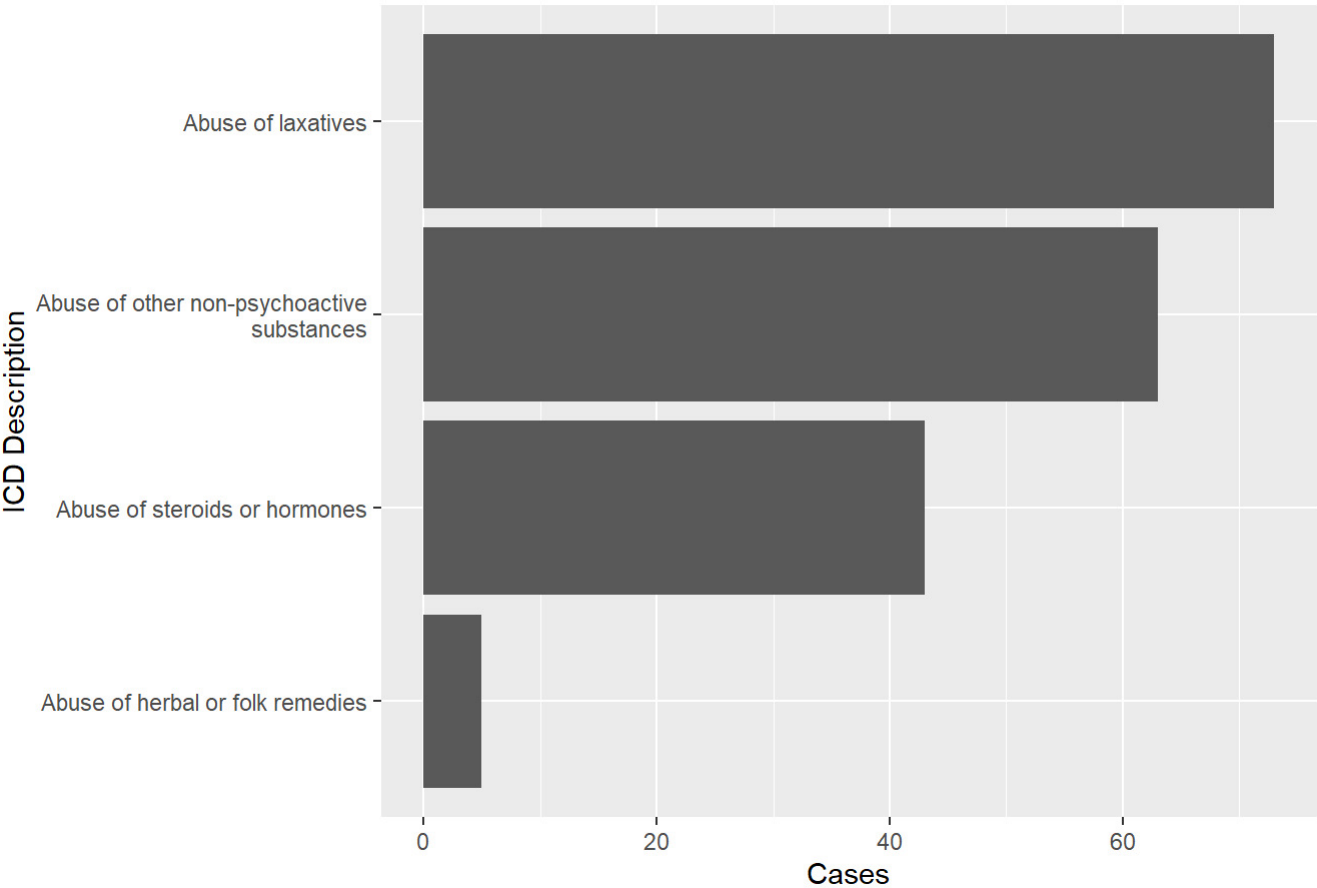


```
## `summarise()` has grouped output by 'Code'. You can override using the
## `.groups` argument.
```

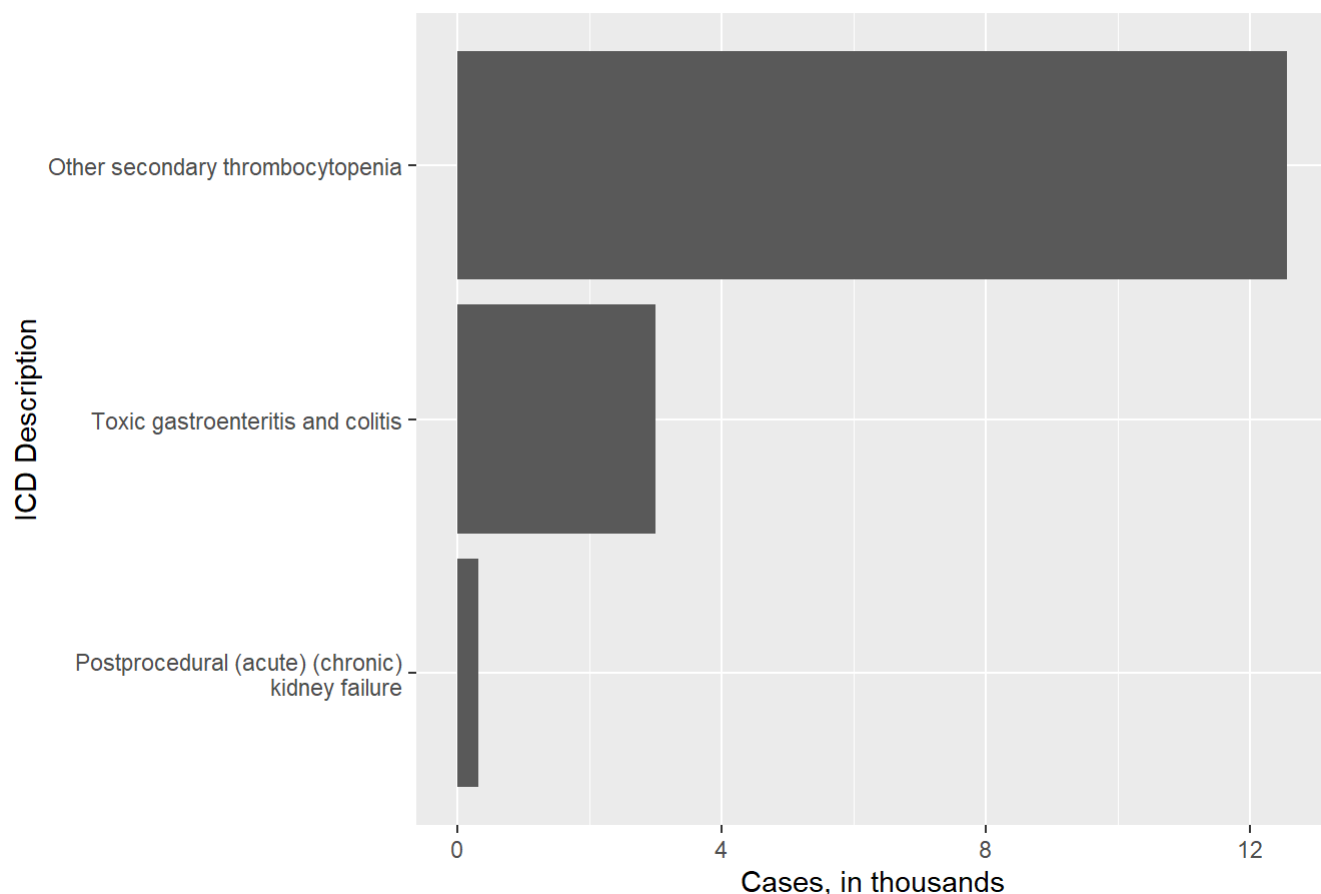
Most Common Category 'A' drug reactions



Most Common Category 'B' drug reactions



Most Common Category 'C' drug reactions



Target Prediction

The target variable was the absence or presence of at least one ADE during the inpatient stay. Just over two percent of admissions exhibited a relevant ICD code. Not all independent variables were used in the final model, notable removals included patient state of origin, provider, and secondary insurance type as these exhibited too many levels to consider and had missing values.

Training data were oversampled to have 50% of 100,000 samples having an ADE, with the rest having no ADE code. Models were trained through 5-fold cross-validation and optimized for performance to maximize accuracy. No hyperparameter tuning was employed; defaults from the associated packages were used.

Performance

Overall performance measured by F1 and F2 scores was low. All four models had scores less than 0.5, which is the routine cutoff for an acceptable result from a balanced dataset. F1 values were between 0.03 and 0.14. Considering the target variable is highly unbalanced, this result is expected. The low F1 values owe largely to the poor precision of the models; each one had difficulty differentiating between admissions featuring an adverse drug event versus the comparatively high rate of false positives. Detailed performance data is available in Table 3. F2 scores are applicable in this domain as false negatives are more potentially harmful in a healthcare setting.

Looking at each specific model, SGB had the highest F-scores and precision, with logistic and neural nets having similar performance. Implementation of the RF model was flawed and suffered from across-the-board inferior performance on the test data. It is possible that this model was incompletely trained and did not reach the necessary complexity to predict the outcome. None of the metrics were over 10% for RF.

Selected Model Performance for the Detection of Adverse Drug Events

Model	Accuracy	Precision	Recall	F1	F2
GLM	77.6	6.7	71.9	0.12	0.24
SGB	77.2	7.4	82.4	0.14	0.27
Random Forest	2.2	5.6	2.2	0.03	0.03
Neural Net	67.9	5.6	85.6	0.11	0.22

Conclusions and Future Directions

Several descriptions considered suggest that patients having adverse drug events are distinct from the general population, and that the target group overlaps with known groups with greater medical needs. Charges for an admission with an ADE are nearly twice that of one without, but this is driven by length of stay and the relationship diminishes when considering costs per day.

The population studied appears to be representative of a usual hospital setting judging by the most common admitting diagnoses and procedures performed. For ADE's ICD classification succeeds in capturing the strongest 'A'-category events. However, the C category and in particular secondary thrombocytopenia may suffer from less precision to being correlated with a drug-related problem. Category B ADE's may suffer from the opposite problem – highly specific but too few cases.

Implementation of machine learning models relied on the default hyperparameters being appropriate for the dataset; future iterations of this research should consider hyperparameter tuning, especially for maximum depth of an RF model.

Putting these results into clinical practice is the true purpose of this exercise, and while F1 and F2 scores were not adequate, there was a class imbalance problem with less than five percent of cases having ADE's. However, clinical practice in hospital pharmacy already requires dozens of screening questions as part of medication reconciliation. A one-in-twenty chance of identifying a risk is potentially an improvement over the 'warning fatigue' that clinical decision-making tools currently exhibit. Likewise, maximization of recall would help ensure no cases are missed. There is a distinct possibility that the original dataset did not detect all adverse events, and that some of the false positives were misclassified originally. Supporting this claim would require retrospective reviews of such cases.

HCUP data could be extended in several ways. At the time of writing, 48 states and the District of Columbia have data available, as well as several different admission populations to consider. Additionally, there are uniform records available over the past few decades.

The CHARGES file in this study was not utilized to its maximum, as it was only considered as a total charge per admission and not further detailed into hospital cost centers or related to related ICD-10 procedures. For HCUP data especially, not much information is given about the specific medications administered while hospitalized, but the intensity of drug therapy could be estimated by searching revenue codes related to the pharmacy department.

This dataset was not organized in a way that made it intuitive to determine a complete picture of a case's hospital admission. Diagnoses and procedures were separated over several dozen columns, with no established reason for their order apart from admitting diagnosis. Real-life implementation of medical records is ordered differently.

Schema for each patient may point to a different database for each admission, with diagnoses linked via a patient-specific key. The organization of the original HCUP files necessitates the agency to provide access to one or two flat files – CORE and CHARGE - with the CORE file in 'wide' format and the charges listed in 'tall' manner. For medical organizations with access to their own records, querying the original sources would provide higher fidelity records to analyze.

Beyond refinement of the existing methods, there are several future directions that this type of research may head. Similar analyses can be performed on datasets from electronic medical records and pharmacy or insurance claims data. These types of records would allow correlation of therapy compliance with the risk of adverse events. Employment of a case-control methodology could help distinguish differences between patients with ADE's while compensating for the class imbalance problem.