DATA 623 Assignment #2

Jordan Swanson 10005366

Background/Rationale

The amount of time a patient spends in the hospital is a significant indication of the cost of their treatment to the healthcare system (Brownell, 1995). As we've seen with the COVID-19 pandemic, hospital resources are unfortunately finite, and efficient utilization of resources can be an economic way to reduce the burden and cost of a stay, all else being equal. Undoubtedly, it's an important goal of any healthcare system to discharge patients as soon as they've completed their course of treatment, but attempts to alleviate stresses with additional spending don't necessarily translate into better outcomes (Needham, 2003).

Length of stay (LOS) alone isn't a good indicator of the types of patients who could be treated with greater efficiency, as some medical issues will always require lengthier stays than others. On the other end of the spectrum, a diagnosis that would normally require a single days' stay in the hospital must be considered abnormal if a patient is admitted for several days. Of course, there are multiple factors that affect the complexity (and length of stay) of a patient, but a look at the correlation between length of stay and diagnosis codes might reveal the first areas for investigation. Codes that have little variation in LOS would seemingly have little that be changed in how they're treated to minimize LOS, in contrast with codes that have much greater variation.

Research Question/Objective

Among the most common diagnosis codes contained in AHS's Discharge Abstract Database (DAD), which single codes have the greatest variance in LOS?

Study Methods

The dataset used was the AHS DAD database available on the DATA623 shared drive, covering the period of April 1st, 2015, to March 31st, 2016. From that dataset, the ICD-10-CA diagnostic codes (Canadian Institute for Health Information, 2021) were extracted from the fields correlating to diagnoses (DXCODE1 … DXCODE25) and analyzed for frequency. To focus on the more common diagnostic codes, only the codes that appeared over 1000 times in the dataset will be considered, resulting in 13 unique codes that represent 18.1% of all codes present in the database. The codes, descriptions, and frequencies are included in Table 1.

To visually compare the different codes and their distributions, boxplots were made for each code with regards to the LOS (in days) associated with each code (Figure 1). To account for the potential effect of outliers, especially on the upper end of measurements, the natural logarithm of the LOS was used in the plot, with the raw values being available in Table 2.

Results

Table 1 shows the 13 most common codes and their frequency within all the codes listed in the dataset (*n*=134,866). Included among the codes are some common reasons for hospitalization, such as pregnancy (Z37000, O70101, O68001) and cardiac concerns (I500, I4890, I2510), though some descriptors are fairly vague (U989).

Looking at Figure 1, we can see that though some of the codes have relatively little variation in the LOS for patients diagnosed with them, specifically those related to pregnancy, many others have a much wider distribution of LOS. Perhaps not too surprisingly, the most common code, I100 (benign hypertension), has one of the widest distributions of any of the codes. Numerical values are also available in Table 2.

With the goal of reducing the hospital stays of patients, the focus should be on the codes that have the longest LOS. Looking at the maximum LOS for the codes related directly to pregnancy and delivery, even the longest stay is merely 46 days. Contrasted with a patient coded with N390 for a urinary tract infection, where even the mean value is nearly 27 days, though the descriptive statistics imply that that number is pulled upwards by some massive outliers. As expected, many of the codes have outliers that indicate nearly unbounded maximums, sometimes measuring in years. On the other end, all codes have a floor effect with the minimum possible LOS of a single day.

Perhaps the more important statistic to consider is the interquartile range (IQR), available in Table 2. As a measure of the spread of the data closer to the median, it's more resistant to outliers than the mean, and can be an indicator of variation for a greater portion of the data. Certainly, though patients coded for a urinary tract infection (N390) have some of the longest stays, it also represents by far the widest IQR, perhaps being a candidate for the ICD-10-CA code worth considering current treatment for with the goal of reducing LOS.

This investigation has been quite preliminary, and certainly has its limitations. To begin, the data is nearly six years old, and likely doesn't properly reflect current practices. Secondly, though this investigation has looked at each ICD-10-CA code in isolation, realistically patients can and do have multiple codes (and conditions) during their hospital stay that would significantly complicate the analysis. A further investigation would have to start by attempting to untangle the relationships between the codes before any meaningful conclusions could be drawn. One interesting angle that could also be explored in the future is the difference in distribution of each code's LOS when compared between facilities in the province. Surely

certain facilities receive the patients in the worst health, but there might be some interesting conclusions to be drawn there.

Appendix

ICD-10-CA Coding	Description	Frequency
I100	Benign hypertension	5984
Z37000	Single live birth, pregnancy resulting from both spontaneous ovulation and conception	4588
I500	Congestive heart failure	1499
N390	Urinary tract infection	1491
I4890	Atrial fibrillation	1443
E1152	Type 2 diabetes mellitus with certain circulatory complications	1335
E119	Type 2 diabetes mellitus without (mention of) complications	1285
N179	Acute renal failure, unspecified	1252
O70101	Second degree perineal laceration during delivery, delivered, with or without mention of antepartum condition	1233
I2510	Atherosclerotic heart disease of native coronary artery	1162
U989	Unspecified place of occurrence	1058
O68001	Labour and delivery complicated by fetal heart rate anomaly, delivered, with or without mention of antepartum condition	1027
J189	Pneumonia, unspecified	1010

Table 1: Descriptions and frequency of common ICD-10-CA codes

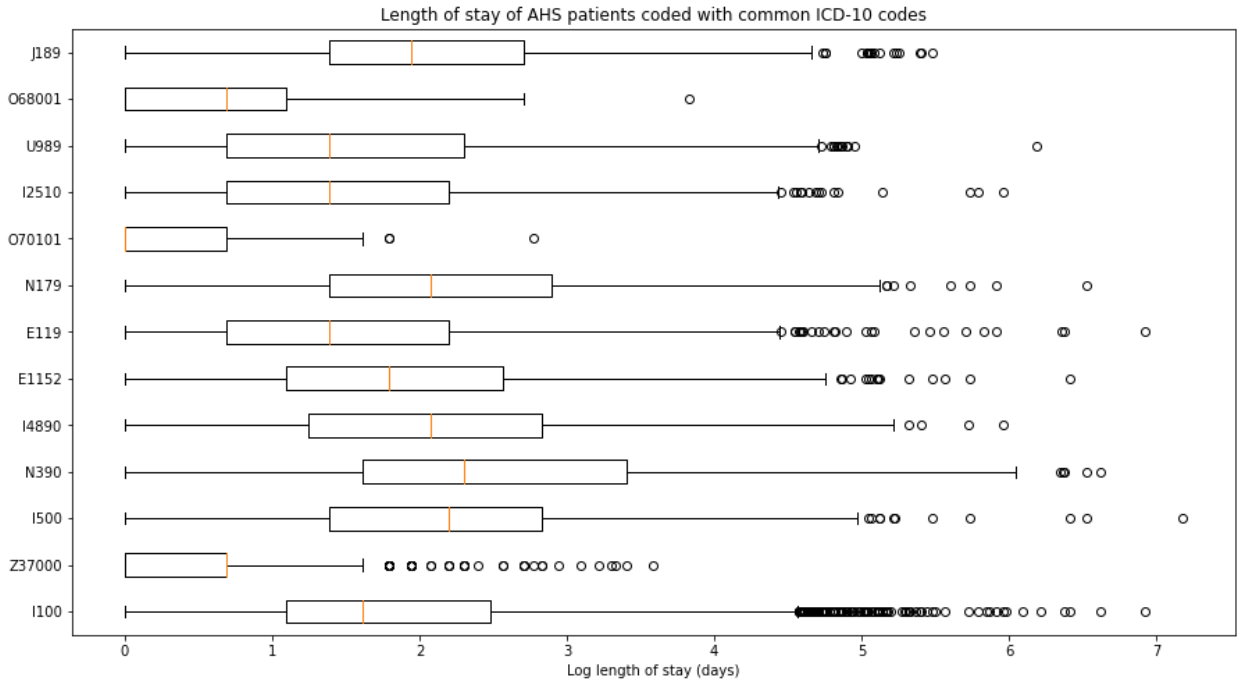

Figure 1: Boxplots of common ICD-10-CA codes

Coding	Count	Mean	SD	Min	25%	50%	75%	Max	IQR
I100	5984	13.8	32.9	1	3	5	12	1016	9
Z37000	4588	2.0	1.5	1	1	2	2	36	1
I500	1499	17.8	46.9	1	4	9	17	1303	13
N390	1491	26.9	52.5	1	5	10	30	748	25
I4890	1443	16.7	27.3	1	3.5	8	17	387	13.5
E1152	1335	14.1	29.3	1	3	6	13	606	10
E119	1285	12.6	44.1	1	2	4	9	1016	7
N179	1252	17.6	32.9	1	4	8	18	679	14
O70101	1233	1.7	0.9	1	1	1	2	16	1
I2510	1162	9.4	22.3	1	2	4	9	387	7
U989	1058	11.6	24.7	1	2	4	10	487	8
O68001	1027	2.2	1.9	1	1	2	3	46	2
J189	1010	15.6	26.3	1	4	7	15	239	11

Table 2: Descriptive statistics of length of stay for common ICD-10-CA codes

References

Brownell, M.D. & Roos, N.P. (1995). Variation in length of stay as a measure of efficiency in Manitoba hospitals. *Canadian Medical Association Journal* 152(5). 675-682.

Canadian Institute for Health Information. (2021). *Canadian Coding Standards for ICD-10-CA and CCI.* Accessed January 23, 2022. https://secure.cihi.ca/estore/productSeries.htm?pc=PCC189.

Needham, D. M., Anderson, G., Pink, G. H., McKillopy, I., Tomlinson, G. A., & Detsky, A. S. (2003). A province-wide study of the association between hospital resource allocation and length of stay. *Health Services Management Research* 16(3). 155-166.