

# Analysis of Surgical Wait Times in Nova Scotia

Carlo Carandang<sup>1</sup>, Gregory Horne<sup>2</sup>, William Wells<sup>1</sup>, Catherine Stokes<sup>2</sup>

Corresponding Author: Carlo Carandang, MD

Email: [carandangc@gmail.com](mailto:carandangc@gmail.com)

ORCID: [0000-0002-2008-9642](https://orcid.org/0000-0002-2008-9642)

*This project was supported by Saint Mary's University, MSc Program in Computing & Data Analytics and NSCC Institute of Technology*

The statistical analysis and code for this study is located at this repository:  
[https://github.com/Carlo-Carandang/Nova\\_Scotia\\_Surgical\\_Wait\\_Times](https://github.com/Carlo-Carandang/Nova_Scotia_Surgical_Wait_Times)

---

<sup>1</sup> Saint Mary's University, MSc Program in Computing & Data Analytics, 923 Robie Street, Halifax, Nova Scotia, Canada, B3H 3C3

<sup>2</sup> NSCC, Institute of Technology Campus, 5685 Leeds Street, Halifax, Nova Scotia B3K 2T3

# Analysis of Surgical Wait Times in Nova Scotia

## Abstract

**OBJECTIVES:** Surgical wait times in Nova Scotia are at crisis levels, as patients are waiting too long on the wait list. This can lead to worsening illness course, and in some cases premature death. This project started with the objective of addressing the surgical waitlist crisis, as people in Nova Scotia are waiting months to years before receiving their life-saving surgeries.

**METHODS:** This study involved obtaining a clinical dataset from an open data portal on the internet, loading the data into our analytical environment, exploring the data using analytical software, cleaning the data, determining feature variables (independent variables), and utilizing the data to train models for prediction of the outcome variable(s). R was utilized to perform exploratory, descriptive, and predictive statistical analysis on the surgical wait times dataset in Nova Scotia, for the time period 2014 to 2016.

**RESULTS:** A bivariate multiple linear regression model was constructed with two dependent variables (consult\_90th and surgery\_90th), representing the 90th percentiles for each instance of a surgical specialty's wait time, added together to give the combined surgical wait time and one independent variable.

**CONCLUSION:** Although we can give an estimate of average wait times in Nova Scotia based on specialty from the model, only the waittimes from the following specialties are statistically significant: General Surgery, Dental, Ophthalmology, Orthopaedic, and Otolaryngology (ENT) at 95% confidence. However, as the model is statistically significant with all the coefficients, then it can be assumed that the model can predict with 95% confidence the wait times from any specialty, even if the individual coefficient is not statistically significant.

Surgical wait times in Nova Scotia are at crisis levels, as patients are waiting too long on the wait list. This can lead to worsening illness course, and in some cases premature death. This project started with the objective of addressing the surgical waitlist crisis, as people in Nova Scotia are waiting months to years before receiving their life-saving surgeries. The Nova Scotia government website only gives numbers about wait times that are difficult to interpret. This research team wanted to look for correlations of the variables in the surgical wait times dataset. In fact, every province has a wait times dataset that is updated every 3 months, and post the data to the Canadian Institute for Health Information public-facing website. The data goes back to the beginning of the recordings, which is roughly around the early 2010s.

The data set for surgical wait times in Nova Scotia has a dozen columns and approximately 7000 rows when the dataset was obtained for the years 2014 to 2016. It's basically an

Excel spreadsheet, and is not in any relational database tables. It maps surgical wait times based on date, hospital, Health Region Zone, specialty, procedure, and doctor.

Research shows that surgical wait times are associated with significant morbidity and mortality risk<sup>1-6</sup>. Morbidity risk is the probability of the illness worsening, while mortality risk is the probability of dying from the illness. It is widely known that the surgical wait times in Nova Scotia are at crisis-levels currently. Patients in Nova Scotia are waiting too long on average to receive their procedure. However, it is not known how the data can inform the improvement of services, and hence morbidity and mortality. Therefore, it is important to analyze the wait times data using advanced statistical and data analytics to help provide more insights to help decrease morbidity and mortality from waiting for surgical procedures.

Although the most urgent surgeries get scheduled first, waiting too long for surgery can be detrimental to one's health because waiting too long is a burden on the patient and their family, waiting too long can lead to death, and waiting too long can lead to worsening of the illness. Indeed, while the most urgent surgeries get scheduled first, waiting too long for surgery can be detrimental to one's health.

Ultimately, the researchers wanted to find the factors associated with surgical wait times, so that new insights can be obtained to decrease the wait times, and hence save lives and prevent illness deterioration.

## Methods

This study involved obtaining a clinical dataset from an open data portal on the internet, loading the data into our analytical environment, exploring the data using analytical software, cleaning the data, determining feature variables (independent variables), and utilizing the data to train models for prediction of the outcome variable(s).

The research team divided the work into the following tasks:

- Project topic selection (C.C., G.H., W.W., C.S.)
- Dataset selection (C.C., W.W.)
- Data cleaning (C.C., G.H.)
- Exploratory analysis and multiple linear regression in R (G.H., W.W.)
- Multiple linear regression and polynomial regression in Python (C.C. and G.H.)
- Visualization of data and graphs (G.H.)
- Research, report writing, and editing (C.C., G.H., W.W., C.S.)

Initially, the data had to be cleaned, as there were many missing values. This was performed manually, as the column and rows were populated haphazardly, and therefore had to inspect and note the missing values, row by row (12 columns and about 7000 rows). Subsequently, business and clinical rules were followed to fill in missing values, utilizing the clinical domain expertise of one of the authors (C.C.) who had practiced as a physician in Nova Scotia for several years. As an example of missing values, each of the surgical

procedures had to be classified into the surgical specialty to which it belonged, as many of the rows were missing this information. In addition, there were missing values for hospital, Health Region Zone, procedure performed by surgeon, and specialty of the surgeon. Any missing values not obtainable within the dataset were scraped from the internet to populate the missing values, such as hospital information and information on the surgeon.

Once the dataset was cleaned and missing values were replaced with inferred values from domain expertise and/or web scraping, exploratory statistical analysis was performed on the surgical wait times dataset in Nova Scotia. The research team was trying to determine if there were any correlations between the numerous independent variables (i.e. specialty, period, facility, etc.) and the dependent variable (surgical wait times). This exploratory analysis was the beginning of an ongoing, in-depth analysis of the Nova Scotia Surgical Wait Times, to help to improve patient care through data analysis and statistical modelling.

R was utilized to perform exploratory, descriptive, and predictive statistical analysis on the surgical wait times dataset in Nova Scotia, for the time period 2014 to 2016.

## Analytical Environment

### Software

Data analyses, visualizations, and production of this report were undertaken using software with the specifications detailed below.

- Anaconda v0.0.0
- Jupyter Notebook v0.0.0
- pdfTeX (TeXlive 2015) v3.14159265-2.6-1.40.16
- Pandoc v1.16.0.2
- R v3.4.4
- R Studio v1.1.383
- R Markdown v2

### Hardware

Data analyses, visualizations, and production of this report were undertaken using hardware with the specifications detailed below:

**Apple MacBook Air** Processor: 1.6 GHz Intel Core i5; Number of Processors: 1; Total Number of Cores: 2; L2 Cache (per Core): 256 KB; L3 Cache: 3 MB; Memory: 8 GB 1600 MHz DDR3; Operating System: Apple Mac OS X v10.0

**Apple Macbook Pro** Processor: 3.1 GHz Intel Core i7; Number of Processors: 1; Total Number of Cores: 2; L2 Cache (per Core): 256 KB; L3 Cache: 3 MB; Memory: 16 GB 1857 MHz DDR3; Graphics: Intel Iris Graphics 6100 1536 MB; Operating System: Apple Mac OS X v10.13

**Hewlett Packard Spectre 13** Processor: 2.5GHz Intel Core i7; Number of Processors: 1; Total Number of Cores: 4; Memory: 8 GB DDR3; Graphics: Intel HD Graphics 520; Operating System: Xubuntu Linux v16.04 LTS

## Acquire the Surgical Wait Times Data

The surgical wait times data was obtained from the Nova Scotia Government [Open Data Portal](#). Here are the steps to acquire the dataset:

- 1) Navigate to the Open Data Portal
- 2) Select the Data Catalogue
- 3) Search 'surgical wait times'
- 4) Select Surgical Wait Times
- 5) Select the Export option
- 6) Select CSV
- 7) Save the data file to the project subdirectory (nova\_scotia\_wait\_times/data\_analysis/data).
- 8) Load the data from a tab-separated-values formatted file with column headers

## Wrangle the Data

Once the data was loaded, the feature names of the data set were determined:

[1] Period [2] Specialty [3] Procedure [4] Provider  
[5] Zone [6] Facility [7] Year [8] Quarter  
[9] Consult\_Median [10] Consult\_90th [11] Surgery\_Median [12] Surgery\_90th

The dataset contained 6843 observations and 12 features. Viewing a sample of the dataset, it was determined that multiple variables were categorical, and the wait times were the only continuous variables. Here are the definitions of the 12 features in the dataset:

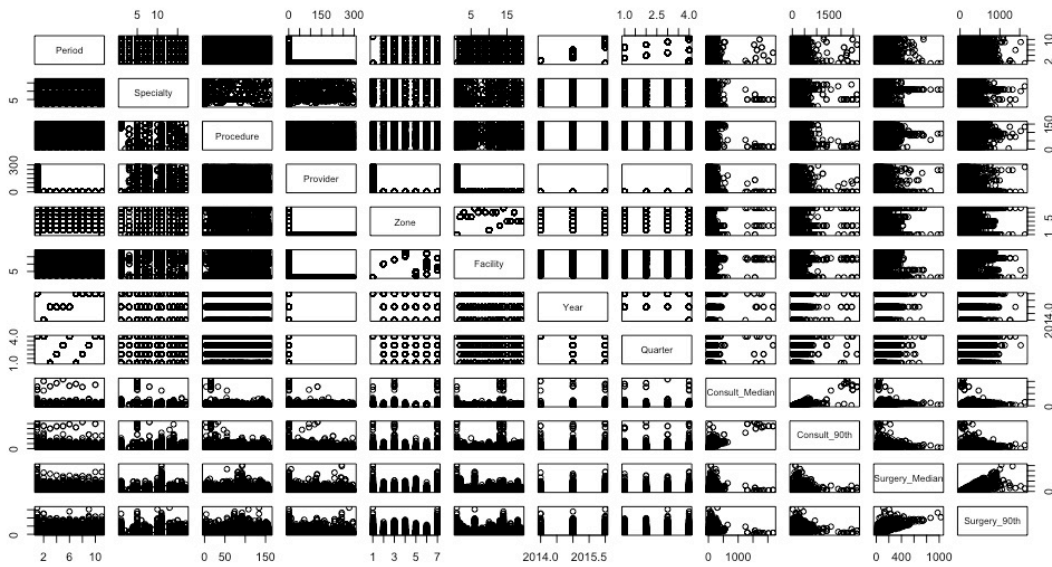
Feature	Definition
Period	Time period
Specialty	Surgical specialty
Procedure	Surgical procedure
Provider	Surgeon
Zone	Healthcare Zone, 1 to 4
Facility	Hospital
Year	Year
Quarter	Quarter (3-months)
Consult_Median	Maximum time that 50% of patients recently waited for consult
Consult_90th	Maximum time that 90% of patients recently waited for consult
Surgery_Median	Maximum time that 50% of patients recently waited for surgery

Surgery\_90th            Maximum time that 90% of patients recently waited for surgery

The surgical specialty types were parsed as follows: specialty, all specialties, cardiac, dental, general, neurosurgery, obstetrics/gynaecology, ophthalmology, oral and maxillofacial, oral maxillofacial, orthopaedic, otolaryngology (ent), plastic, thoracic, urology, vascular. In total, there are 15 surgical specialty types and 159 surgical procedure types.

Next was the scatterplots of the features paired against one another:

Figure. Features as pairs in scatterplots.



The exploratory analysis gave us an initial signal that specialty may be correlated with wait times, as the scatterplots of the various combinations of the variables showed that the specialties showed a ‘banded’ pattern when it was paired and plotted versus wait times, and some specialties appeared to have greater variation in wait times than others. The ‘banded’ pattern occurred due to the specialty variable being categorical (discrete datapoints). Because of this banded pattern, specialty was chosen as the independent variable, and consult and surgery times were chosen as the dependent variable. Therefore, after this exploration of the dataset for features to study, only observations with surgical procedure type ‘all’ were extracted:

feature	missing_count	nonmissing_count
consult_90th	12	284
consult_median	12	284
facility	296	0
period	0	296
procedure	0	296
provider	0	296

quarter	296	0
specialty	0	296
surgery_90th	0	296
surgery_median	0	296
year	296	0
zone	296	0

The minimum, maximum, average, standard deviation, and total combined wait days by consultation and surgery specialty. Average wait days is the median number of days on the surgical wait list:

specialty	minimum	maximum	average	sigma	total	observations
cardiac	66	198	157	49	702	5
dental	148	1032	327	319	7006	16
general	65	2234	177	298	14432	56
neurosurgery	155	949	252	236	3081	10
obstetrics/gynaecology	64	882	199	149	9573	41
ophthalmology	115	2875	392	497	16779	33
oral maxillofacial	171	620	421	159	4332	11
orthopaedic	162	1365	662	318	26539	38
otolaryngology (ent)	136	1081	390	258	11910	25
plastic	151	738	372	186	5598	15
thoracic	73	449	179	134	1307	6
urology	61	819	219	170	6002	22
vascular	112	685	307	242	2151	6

Rows with missing values were removed before training the prediction model.

## Visualise the Data

After wrangling the data, the data was visualized:

Figure 1.

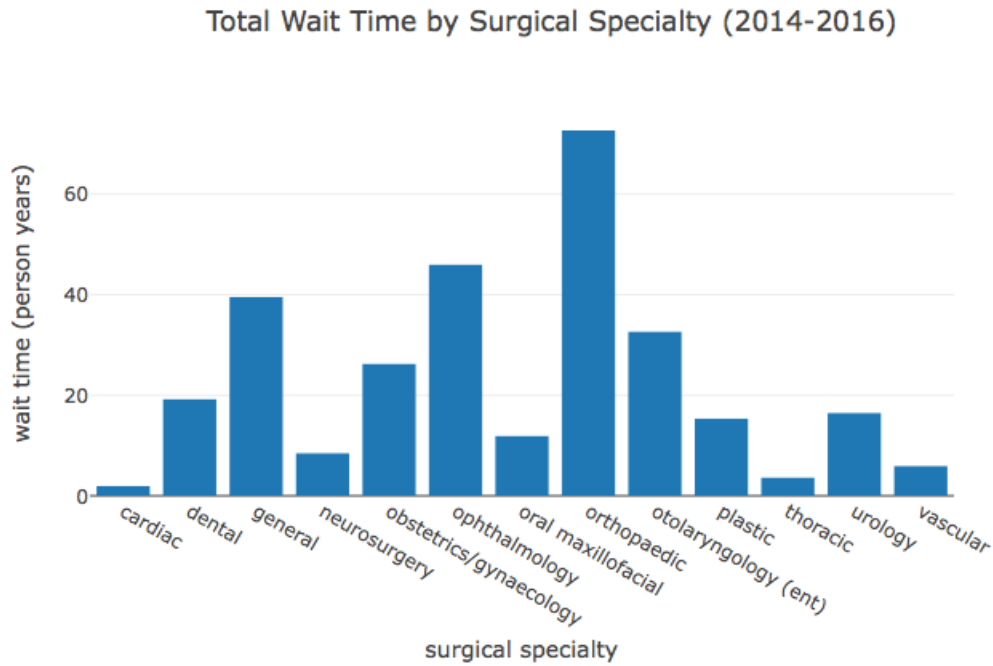


Figure 1 shows the total wait time in person years, grouped by surgical specialty. Cardiac surgery has the lowest value, while orthopaedic has the highest.

Figure 2.

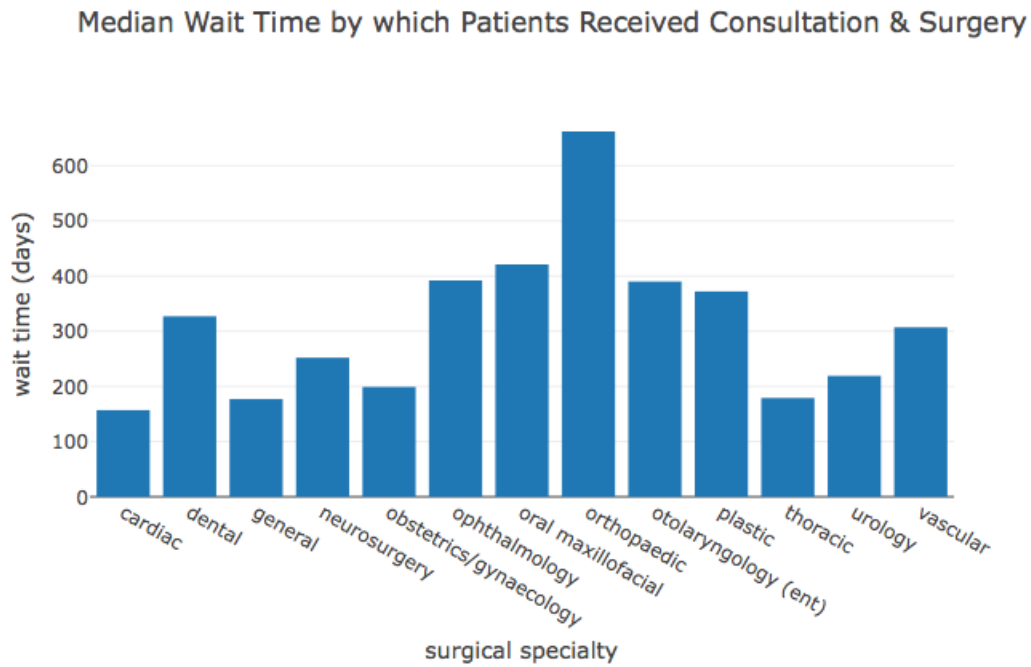




Figure 2 shows the median total wait time in days, grouped by surgical specialty. Cardiac surgery has the lowest value, while orthopaedic has the highest.

Figure 3.

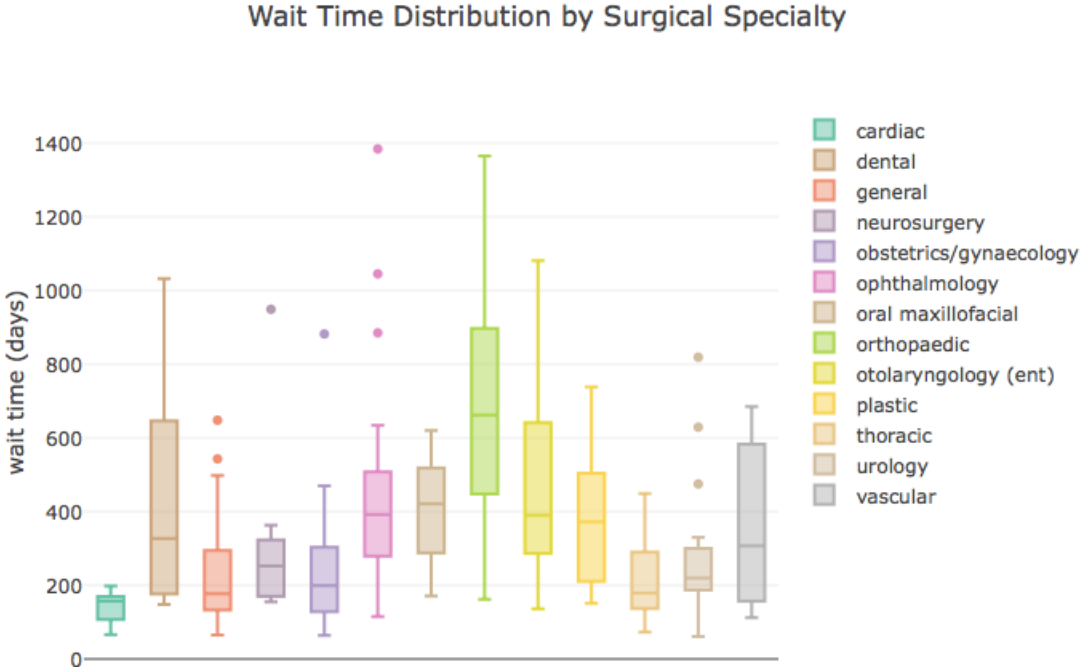


Figure 3 shows the wait time distribution in days, grouped by surgical specialty. Notice cardiac surgery and thoracic surgery at the lower end, and orthopaedic at the higher end.

Figure 4.

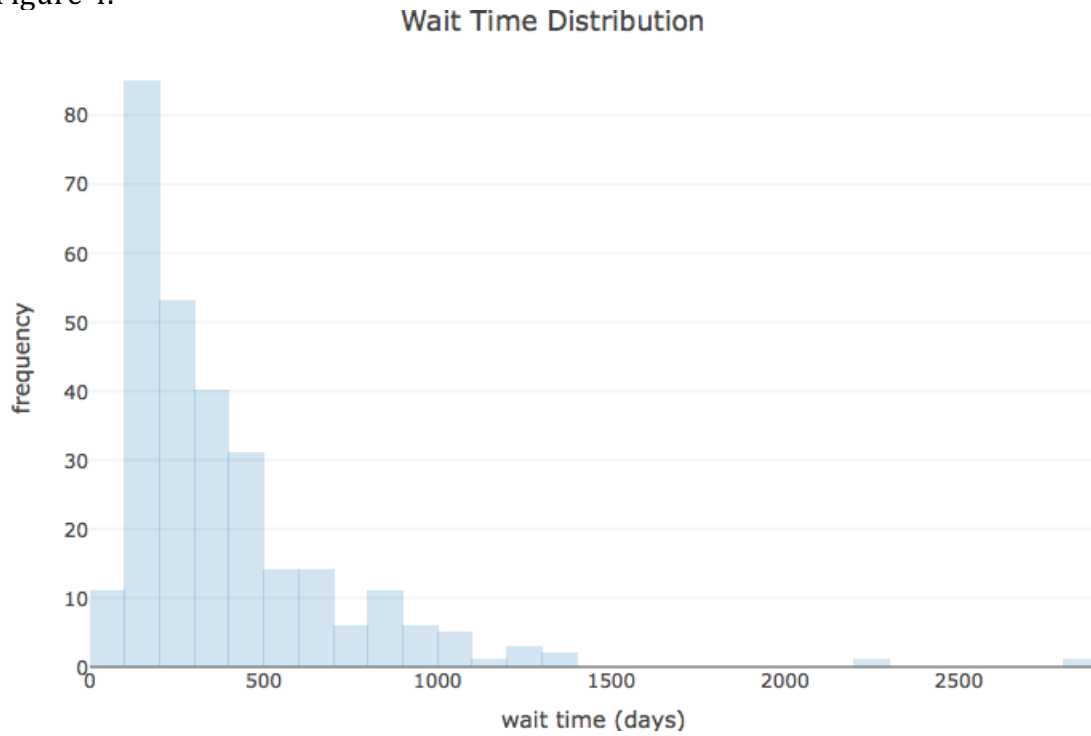


Figure 4 shows the histogram of the frequencies of the wait times. Most people are waiting between 100 to 200 days for their surgical procedure.

Figure 5.

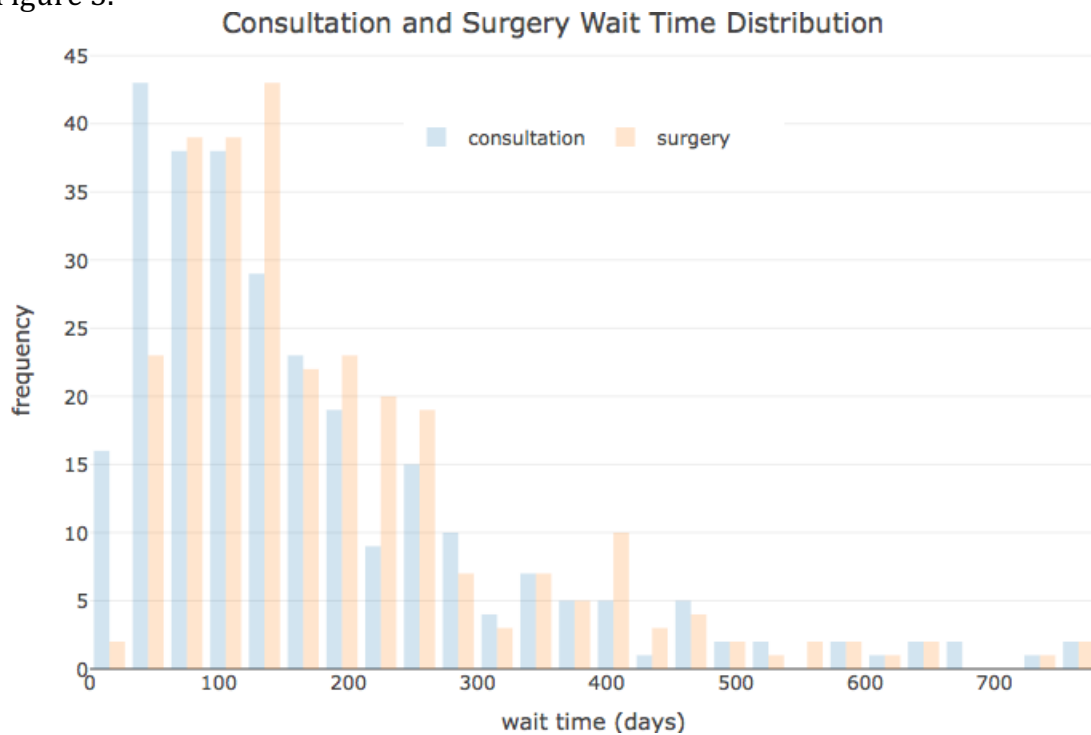


Figure 5 shows the histogram of the frequencies of the wait times, grouped by consultation wait times and surgery wait times. Most people are waiting between 50 days and 200 days for a consultation, and between 50 to 250 days for a surgical procedure.

It was determined from the dataset that the total wait time to receive a surgical procedure is the sum of the consultation wait time and the surgical wait time. The domain expert who had worked in the system (C.C.) for several years interpreted the total wait time for a surgical procedure as the sum of the following wait times:

\* Wait time to see a family doctor, who will determine if a surgery consultation is needed (family doctor wait time) \* Wait time to consult with a surgeon (consult wait time) \* Wait time for the surgical procedure (surgery wait time)

So the total wait time to receive a surgical procedure is as follows:

**Total Wait Time = Family Doctor Wait Time + Consult Wait Time + Surgery Wait Time**

But this dataset is missing the Family Doctor Wait Time, so the estimated wait times for surgical procedures in this analysis is missing the Family Doctor Wait Time. Therefore, any estimated prediction of wait times from this study is an underestimate of the real wait time.

## Results

### Build the Model

Prior to building the statistical model the baseline factor was set to 'general surgery' instead of the default 'cardiac surgery' to determine the impact, if any, on the linear regression model with regards to the null hypothesis.

A bivariate multiple linear regression model was constructed with two dependent variables (consult\_90th and surgery\_90th), representing the 90th percentiles for each instance of a surgical specialty's wait time, added together to give the combined surgical wait time and one independent variable.

**A summary of the multiple linear regression model is as follows:**

Call: `lm(formula = specialty90 ~ specialty)`

Residuals:

Min	1Q	Median	3Q	Max	
-536.39	-144.57	-66.16	70.76	2366.55	
Coefficients:		Estimate	Std. Error	t value	Pr(>)
(Intercept)		257.71	38.80	6.643	1.68e-10 ***
specialtycardiac		-117.31	135.51	-0.866	0.387417
specialtydental		180.16	82.30	2.189	0.029447 *

specialtyneurosurgery	50.39	99.67	0.506	0.613607
specialtyobstetrics/gynaecology	-24.23	59.68	-0.406	0.685083
specialtyophthalmology	250.74	63.71	3.935	0.000106 ***
specialtyoral maxillofacial	136.10	95.75	1.421	0.156338
specialtyorthopaedic	440.68	61.02	7.222	5.19e-12 ***
specialtyotolaryngology (ent)	218.69	69.83	3.131	0.001930 **
specialtyplastic	115.49	84.41	1.368	0.172388
specialtythoracic	-39.88	124.72	-0.320	0.749385
specialtyurology	15.10	73.05	0.207	0.836358
specialtyvascular	100.79	124.72	0.808	0.419727

Signif. codes: 0 | \*\*\* 0.001 | \*\* 0.01 | \* 0.05 | ' 0.1 | ' 1

Residual standard error: 290.3 on 271 degrees of freedom

Multiple R-squared: 0.2383, Adjusted R-squared: 0.2046

F-statistic: 7.066 on 12 and 271 DF, p-value: 3.522e-11

**From the above summary of the model, the following statistical measures were extracted:**

degrees of freedom: 13 and 271

p-value of the model:  $3.522271910^{-11}$

residual standard error: 290.3315348

F-statistic: 7.0658736

F-critical: 1.7583858

**An analysis of variance (ANOVA) of the linear regression model was used to validate the model:**

Analysis of Variance Table

Response: specialty90

Measure	Df	Sum Sq	Mean Sq	F value	Pr(>F)
specialty	12	7147193	595599	7.0659	3.522e-11 ***
Residuals	271	22843240	84292		

## Interpretation

The residual standard error (standard error of the estimate) is 290.33, where it is the standard deviation of the variation of observations around the regression line (it is the standard deviation of the regression model). So waittimes average can be estimated as  $\pm 2(290.33) = 580.66$ . As this value is large when compared to the average waittimes in the

sample, then the variation of observed y values from the regression line is also large. For future research, we should look for other variables which can explain more of the variation in waittimes (ie. costs for procedures, facilities funding, staffing levels, etc.).

$R^2$  is 0.2383, but we do have to utilize the adjusted  $R^2$ , as we have multiple independent variables. The adjusted  $R^2$  is 0.2046, where 20.46% of the variation in waittimes is explained by the variation in specialty, taking into account the sample size and number of independent variables.

Next, we want to determine if the model is significant. Hypothesis testing is set up as follows:

**Hypotheses:  $H_0: (\beta_1) = (\beta_2) = \dots = (\beta_k) = 0$  (no linear relationship)**

**$H_A$ : at least one  $(\beta_i) \neq 0$  (at least one independent variable affects y)**

**F statistic =  $(SSR/k)/(SSE/(n-k-1))$**

Since  $F = 7.07$  is in the rejection region (it is greater than the  $F(\text{critical}) = 1.76$ ), we reject the null hypothesis at  $\alpha = 0.05$ , and accept the alternative hypothesis that at least one independent variable affects total surgical wait times at 95% confidence. According to the model the  $H_0$  (null hypothesis) should be rejected in favour of  $H_A$  (alternative hypothesis). We determine that the model is significant.

In addition, we conclude that the regression model does explain a significant portion of the variation in waittimes. This conclusion is also confirmed by the F-statistic with a p-value =  $3.522271910 \times 10^{-11}$ , and therefore it is less than the alpha value of 0.05, and we can also conclude from this p-value that the model does explain a significant portion of the variation in waittimes.

An analysis of variance (ANOVA) of the linear regression model was used to validate the model. For example, the degrees of freedom, sum of squares, and mean squared can easily be retrieved for both the specialty and residuals parameters, and the F-statistic for the specialty parameter. The values of the F-statistic (7.0658736) and F-critical (1.7880147) indicate the  $H_0$  (null hypothesis) should be rejected in favour of  $H_A$  (alternative hypothesis).

Next, we extract the coefficients so the linear regression model equation can be constructed. Here is the prediction model:

**waittimes = 257.71 - 117.31(Cardiac Surgery) + 180.16(Dental) + 50.39(Neurosurgery) - 24.23(Obstetrics/Gynaecology) + 250.74(Ophthalmology) + 136.10(Oral Maxillofacial) + 440.68(Orthopaedic) + 218.69(Otolaryngology (ENT)) + 115.49(Plastic Surgery) - 39.88(Thoracic Surgery) + 15.10(Urology) + 100.79(Vascular Surgery)**

General Surgery is the default, where all the other specialty variables are 0. The model predicts the following for average wait times by specialty:

Although we can give an estimate of average wait times in Nova Scotia based on specialty from the model, only the waittimes from the following specialties are statistically significant: General Surgery, Dental, Ophthalmology, Orthopaedic, and Otolaryngology (ENT) at 95% confidence. However, as the model is statistically significant with all the coefficients, then it can be assumed that the model can predict with 95% confidence the wait times from any specialty, even if the individual coefficient is not statistically significant.

## Conclusion

In this project, we started with the analysis for specialty vs wait times, as the exploratory analysis gave us an initial signal that specialty may be correlated with wait times, as the scatterplots of the various combinations of the variables showed that the specialties showed a 'banded' pattern when it was paired and plotted versus wait times, and some specialties appeared to have greater variation in wait times than others. The 'banded' pattern occurred due to the specialty variable being categorical (discrete datapoints).

We used multiple linear regression for specialty vs. wait times, and found that our model explained 20% of the variation in the dependent variable (wait times). Although we can give an estimate of average wait times in Nova Scotia based on specialty from the model, only the wait times from the following specialties are statistically significant: General Surgery, Dental, Ophthalmology, Orthopaedic, and Otolaryngology (ENT) at 95% confidence. However, as the model is statistically significant with all the coefficients, then it can be assumed that the model can predict with 95% confidence the wait times from any specialty, even if the individual coefficient is not statistically significant.

In summary, our model for specialty vs. total wait times (consult plus surgery wait times) was statistically significant, as indicated by our F-statistic with a  $p < 0.05$ . This F-statistic analysis was performed with all of the specialties included in the model, without backward elimination.

## References

1. Coughlin SMP, Guidolin K, Fortin D, Frechette E, Malthaner R, Inculet R. Is it safe to wait? The effect of surgical wait time on survival in patients with non-small lung cell cancer. *Can J Surg.* 2015;58(6):414–18.
2. Malaisrie SC, McDonald E. Mortality while waiting for aortic valve replacement. *Ann Thorac Surg.* 2014;98(5):1564–70.
3. Kulkarni GS, Urbach DR. Longer wait times increase overall mortality in patients with bladder cancer. *J Urol.* 2009;182(4):1318–24.
4. Rexius H, Brandrup-Wognsen G. Mortality on the waiting list for coronary artery bypass grafting: Incidence and risk factors. *Ann Thorac Surg.* 2004;77(3):769–74.
5. Koomen EM, Hutten BA. Morbidity and mortality in patients waiting for coronary artery bypass surgery. *Eur J Cardiothorac Surg.* 2001;19(3):260–5.
6. Everhart JE, Lombardero M. Increased waiting time for liver transplantation results in higher mortality. *Transplantation.* 1997;64(9):1300–6.