5-2024

Factors Predictive of the Development of Surgical Site Infection in Thyroidectomy, A Replication Study of Myssiorek (2018)

Kaitlyn M. Kenig
University of Nebraska Medical Center

Capstone Replication Analysis

Factors predictive of the development of surgical site infection in Thyroidectomy, A Replication Study of Myssiorek (2018)

Kaitlyn Kenig

Principal Investigator:

Kaitlyn Kenig
MPH Student with Emphasis in Biostatistics
College of Public Health
University of Nebraska Medical Center
kkenig@unmc.edu

Replicated Article Title: Factors predictive of the development of surgical site infection in thyroidectomy – An analysis of NSQIP database

Journal: International Journal of Surgery 60 (2018) 273-278

Table of Contents

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
AMR	Antimicrobial Resistance
ASA	American Society of Anesthesiologists
BMI	Body Mass Index
CPR	Cardiopulmonary Resuscitation
CPT	Current Procedural Terminology
GEE	Generalized Estimating Equations
GLMM	Generalized Linear Mixed Models
IQR	Inter Quartile Range
MEA	Measurement and Estimation Analysis
NSQIP	National Surgical Quality Improvement Program
PoDx	Post-operative Diagnosis
SAP	Statistical Analysis Plan
SSI	Surgical Site Infection
TCA	Theory of Change Analysis
VIF	Variance Inflation Factor

Abstract

The original study aimed to show that thyroidectomy does not result in surgical site infection (SSI) in most cases, and thus routine prescription of antibiotics is not necessary. The study looked to see what risk factors could predict the incidence of SSI. This would highlight those individuals who were at most risk of developing SSI, and then antibiotics would only be prescribed to these individuals instead of all or most individuals who undergo thyroidectomy.

This study used NSQIP data to look at incidence of SSI and look for risk factors that may be predictive of SSI. Only surgeries that were considered clean were included. To determine factors that could be significantly associated with SSI the study used Chi-square tests and logistic regression.

It was determined that being 80 or older, male, and a current smoker were all significant predictors of SSI after thyroidectomy as well as being ventilator dependent. If an individual did not have these risk factors, then incidence of SSI was relatively low and routine prescription of antibiotics was not recommended.

The replication analysis included pure replication, Measurement and Estimation Analysis (MEA) and Theory of Change Analysis (TCA). The goal of the pure replication was to reproduce results presented in the tables from the original paper using the NSQIP data and reported statistical methods. For the MEA, the goal was to check the model assumptions used in the univariate Chi-square analysis and logistic regression. The assumptions for the multivariate logistical regression were also checked using residual analysis. The MEA proved that model assumptions were met for both Chi-square and logistic regression. For the TCA, the objective was to determine if there was any interaction between gender and being a current smoker, as

both show strong statistical significance based on the multivariate analysis. It was determined that no interaction was present.

1. Introduction

This paper describes the statistical analysis methods and data presentations used in the replication analysis of the article *Factors predictive of the development of surgical site infection in thyroidectomy – An analysis of NSQIP database.* This study uses National Surgical Quality Improvement Program (NSQIP) data from 2012 to 2015 for Thyroidectomy to determine if there are any predictive factors of Surgical Site Infection SSI (Myssiorek et al., 2018).

Thyroid surgery is typically a very sterile procedure with low rates of SSI. Due to this, most thyroid surgeries do not require antibiotics to be routinely prescribed or recommended by international standards. Though SSI is rare with this procedure, if it does occur it can cause horrible consequences. This is why around 26% of surgeons still prescribe antibiotics almost always. Problems such as antimicrobial resistance though have increased the challenge in treating SSI.

This study was conducted to gain a better understanding of the incidence of SSI and the risk factors that predict SSI in thyroidectomy. With a better understanding, surgeons will be able to decide which patients were at the most risk and therefore only prescribe for those selected patients. This would reduce the risk of SSI and would avoid the issue of antibiotic resistance. Through this study the hope is to find the most effective use of health care resources in respect to SSI in thyroidectomy procedures.

There were three main goals or objectives of the study. The first is to determine the incidence of SSI after thyroidectomy with clean wounds. The second is to identify the pre-operative risk factors predictive of SSI in clean wound thyroidectomy performed in general and for those who specifically have a cancerous lesion. The third goal of the study is to guide thyroid surgeons in the use of selective prophylactic antibiotic use.

I chose to replicate this study for multiple reasons. The first being that I have done work in the past with adverse events (AEs) in relation to pharmaceutical products when I worked as a Pharmacovigilant Specialist. I thought it would be interesting to see AEs presented in relation to surgical procedures. The second reason was when I worked in a hematology laboratory I got to assist in several small procedures. One of these was a bone marrow biopsy. These are procedures that do not routinely require antibiotics to be prescribed as they are considered clean procedures. Before every bone marrow biopsy though we would mention the risk of infection and how to ensure the site was clean. This study reminded me of these procedures. Another reason I picked thyroidectomy specifically is because my mother-in-law just had this procedure and I wanted to know more about it.

2. Methods

2.1. Pure Replication

For the pure replication analysis, the data was pulled from NSQIP, which is a risk adjusted surgical outcomes database. The time interval examined matched the study and spanned from 2012 to 2015. Thyroidectomy cases with the Current Procedural Terminology (CPT) codes 60210, 60212, 60220, 60225, 60240, 60252, 60254, 60260 were identified. Cases with patients younger than 18, cases considered retrosternal thyroidectomy cases, and cases not classified as clean were excluded from the analysis. Superficial, deep, and organ-space infections were examined.

For the replication analysis tables 1 through 8 from the original study were considered. A univariate Chi-square analysis and logistic regression of thyroidectomy cases with and without SSI were performed to confirm statistically significant differences seen in table 1 and table 3. For this PROC FREQ with the Chi-square option in SAS was used. For the mean and standard deviation of age and Body Mass Index (BMI), PROC MEANS in SAS was used. Note for table 3 there were 307 post-operative diagnoses, these were sorted from highest to lowest frequency. Those with a frequency of more than 1% were analyzed and all other diagnoses were grouped as 'other.' These post-operative diagnoses were analyzed for the comparison of the two groups, they were not analyzed using multivariate logistic regression. Both tables 1& 3 were replicated entirely.

For tables 2 & 4, eligible variables were identified from tables 1 & 3 by looking at those variables with a p-value less than 0.2. For all eligible variables in the original analysis excluding ASA classification, a multivariate analysis was performed. This was done using PROC LOGISTIC in SAS. The logistic model in Table 4 is similar to table 2, however the model is adjusted by adding three types of procedures that were considered statistically different between

the two groups. Table 5 will not be replicated due to it being unclear if the factors presented in this table occurred prior to the SSI.

For the replication of tables 6 & 8, a subgroup analysis was performed on patients who had thyroidectomy for malignancy (CPT 60252 and 60254). This subgroup analysis followed similar steps as above for the full replication analysis while only considering two of the CPT codes. For table 6 & 7 the groups were contrasted using Chi-square, t-test, and non-parametric tests for pre-intra, and post operative characteristics. For table 6, the authors of the original paper compared each subgroup to all other subjects for continuous variables such as age and BMI. In the replication analysis, groupings remained together when computing the Chi-square test. Table 7 will not be replicated due to this table presenting factors that may have occurred after SSI. The multivariate logistical regression was run for the sub-group analysis similarly to above for the full analysis. For all tables being replicated we expected to see similar results to the original analysis. Rounding errors or differences may be seen as it is unclear what software was used in the original study.

2.1.1. Data Manipulation

For the data coming from the NSQIP database, the first step was to take the data from years 2012, 2013, 2014 and 2015 and subset with the appropriate CPT codes. During this step, the data was subset for only those with wound classification considered “1-Clean.”

Once the data used for the study was isolated further manipulation was needed to produce the demographic groups seen in table 1. The first step was to create a binomial variable for occurrence of SSI so that Chi-square analysis could be run. This was originally split between the following three variables: Occurrences of Organ Space SSI, Occurrences of Deep Incisional SSI,

and Superficial surgical site infection. For these three variables they were either marked as having the infection or "No Complication." If for a subject, all three SSI variables agreed that there was "No Complication," then the new variable for SSI was considered "No SSI." If even one of the three variables was marked as having SSI, the subject was considered in the "With SSI" group for the new SSI variable created.

In the data, age was a character variable with those over 90 being displayed as 90+. For the replication analysis, all those 90+ were converted to 90, and the age variable was then converted to numeric so that age group variables seen in table 1 could easily be produced. The original analysis compared each level for age against all other levels combined. Due to this, the pure replication analysis created a new variable for each age group versus one age group variable that has all the age groups represented. This will give a different Chi-square p-value for each age category. In the data there was no variable for BMI, however there was height and weight. The following calculation was used to create a variable for BMI: 703.0768*(weight)/(height*height). From these results, BMI variables seen in table 1 were created. Similarly to age, original analysis compared each level for BMI against all other levels combined. So, for the pure replication BMI variables were created for each BMI category versus one BMI group variable that has all the BMI categories represented. ASA classification variables also had to be redefined for the pure replication and a new variable created for each category as the original analysis compared each classification level against all other classification levels combined.

Other variables such as race, diabetes, and ones for principal anesthesia were manipulated to make binomial variables for the purpose of Chi-square analysis. The variable used for hypo-albuminemia was numeric and needed to be converted into a categorical variable for the use in table 1. No cut-off range for hypo-albuminemia was mentioned in the original

analysis paper, so those with albumin levels <2.5 mg/dL were considered to have hypo-albuminemia (Akirov et al., 2017).

Intra- and post- operative characteristic data in table 3 also needed to be manipulated some for the Chi-square analysis. CPT codes were made into binomial variables for analysis. The variable for deep vein thrombosis and the PoDX variables were also converted into binomial variables.

2.2. Measurement and Estimation Analysis

After results from the original analysis were verified, the model assumptions of the analyses used were checked. For Chi-square analysis the model must meet the following criteria: two categorical variables with two or more categories for each variable, independence of observations, and a relatively large sample size (*Libguides: SAS tutorials: Chi-square test of Independence*).

For logistic regression, the assumptions of appropriate outcome structure, observation independence, absence of multicollinearity, linearity of independent variables and log odds, and of a large sample size must be met (*Logistic and linear regression assumptions - lexjansen*). The assumption of multicollinearity was tested using the CORRB option in PROC LOGISTIC.

As well as checking model assumptions, Chi-square analysis was rerun for age groups, BMI groups, CPT Codes and ASA classification groups. For the rerun, these variables overall p-value is displayed as an additional column in the pure replication tables. After the Chi-square is rerun the logistic regression models will be rerun where all levels of age and BMI are included in the model as it is incorrect to not include all levels as the original analysis did.

2.3. Theory of Change Analysis

For this analysis, the original study was extended by looking at interactions in the multivariate logistic regression model as the original study did not comment on whether these were examined. Specifically, the theory of change analysis looked for an interaction between gender and being a current smoker. In table 2 of the analysis, both show strong significant differences of <0.001.

If an interaction was seen in the multivariate logistical regression, then two subgroup analyses would have been conducted. One for males who smoke and one for females who smoke. These subgroup analyses would have been conducted by comparing those with SSI and those without using Chi-square and multivariate logistic regression like the pure replication detailed above.

3. Results

3.1. Pure Replication

For the pure replication analysis of the tables seen in the article *Factors predictive of the development of surgical site infection in thyroidectomy – An analysis of NSQIP database*, many differences were seen between the original analysis and the replication results. The replicated tables with side by results of the original analysis and replication are represented below with the discrepancies highlighted.

In table 1, there were slight differences in the Chi-square p-values for most of the age groups. They were within .005, so these differences are not significant and do not change the interpretation. These differences may be due to rounding or typos. The frequencies and percentages for all age groups match, except for those over 80 years old. Replication analysis shows more individuals in this category than the original analysis. This discrepancy is related to when the age variable from the original data source was transformed into a numerical variable. Those 90+ were sorted into the over 80 group in the replication but were lost in the original analysis. This caused a difference in Chi-square p-value, but the results remain statistically significant in both the original analysis and replication analysis. For gender, there is a discrepancy in frequency count for those with no SSI. This is a typo in the original table, as the no SSI column matches the frequency in the total column even though there are males with SSI.

For BMI, there are also discrepancies between the original analysis and the replication regarding frequency counts. These differences have caused differences in Chi-square p-value as well. For those with a BMI from 19 to <25 kg/m^2, this difference in counts caused a change in statistical significance. For the row displaying BMI mean and standard deviation, there is a discrepancy here, however it is related to the original analysis displaying median instead of

mean. The original analysis also displays the inter quartile range (IQR) even though it claims to display standard deviation.

For the American Society of Anesthesiologists (ASA) classification variables, there are some slight rounding differences. For ASA variables that have a low frequency count though we see discrepancies in the p-value. The original analysis methods state that a univariate Chi-square analysis was performed on all variables with no mention of other tests being performed outside of multivariate logistic regression. However, the p-value displayed for those ASA variables with low frequency counts is the Fisher's Exact Test two-sided p-value and not the Chi-Square p-value. This difference in p-values displayed did not affect the significance interpretation for ASA classification variables, however.

In table 1, the row for principal anesthesia, other than general, has a discrepancy of 2 in the frequency count. This variable had to be manipulated for Chi-square analysis as there were subjects with general anesthesia under this variable. When excluding those who had general anesthesia there were also a couple individuals who were marked "None" and these were excluded in the replication analysis as they did not receive anesthesia. The original analysis only excluded those marked "general." The p-value for this variable is like the ASA classification scenario as the p-value displayed in the table is from the Fisher's Exact Test and not the Chi-square p-value. The original analysis uses the p-value obtained for Fisher's Exact Test throughout the study for variables both in table 1 and table 3 that have low frequency counts.

Another area where there are differences between the original analysis and the replication is related to hypo-albuminemia. This difference is due to there not being a categorical variable for hypo-albuminemia and it needing to be derived from albumin levels. The original analysis did not disclose the ranges they used to derive this variable and therefore this caused differences

in the results. The p-value from the replication remains not statistically significant, however it is less than 0.200 and therefore would have been included in the multivariate logistic regression.

Results for table 3 show that the replication analysis is similar to the original analysis. The differences for this table are either slight rounding differences or the p-value not being the Chi-square p-value. For operative time and length of hospital stay, the values displayed are median and IQR. As these variables are continuous variables and not categorical variables, the p-value displayed is from a t-test and not Chi-square.

For the multivariate logistic regression tables, there are many discrepancies between the original analysis and the replication analysis. The p-values seen, along with the odds ratios and confidence intervals (CI), do not align. This is likely due to differences seen in frequencies between the original analysis and replication analysis. When looking at the p-values to determine significance, however, the interpretation remains the same for Table 2. We see significant differences in those older than 80, in males, those with a BMI 40 to <50, those who are current smokers, and those who were ventilator dependent within 48 hours preceding surgery. For Table 4, we see a slight change in interpretation. We continue to see significant differences in those older than 80, in males, those with a BMI 40 to <50, those who are current smokers, and those who were ventilator dependent within 48 hours preceding surgery. Where it varies from the original analysis, is the replication shows that total thyroid lobectomy, unilateral, with or without isthmectomy (CPT 60220) is also statistically significant.

For the subgroup both Table 6 and Table 8 were replicated. In Table 6 we see similar differences to what was observed in Table 1. When looking at the age variables, when there was a small sample size the original analysis displayed Fisher's Exact p-value though there was no mention of this in their methods section. The group for those who were older than 80 remained

statistically significant when looking at incidence of SSI. The counts for this age group were slightly different in the replication than in the original analysis and this could be related to how the 90+ patients were handled when manipulating the variables. No differences were observed for males in the original subgroup analysis compared to the replication subgroup analysis.

When looking at BMI for Table 6 replication, the frequencies did not match the original analysis for any of the BMI categories. Results for those with BMI ranging from 19 to less than 25 kg/m^2 remained statistically significant when looking at Fisher's Exact p-value but not when using the Chi-square p-value. As the sample size is small, Fisher's Exact p-value is what should be considered. For ASA classification, the differences seen between the original analysis and the replication analysis included Fisher's Exact p-value being displayed when the sample size was low, and minor rounding issues. Results remained not statistically significant. Same trends for variables related to diabetes, race, anesthesia, wound infection, steroid use, weight loss, COPD, functional status, ascites, dialysis, hypertension, and disseminated cancer were seen between the two analyses. For hypo-albuminemia, the frequencies are way higher in the replication analysis than in the original analysis. As mentioned above, this is due to not using the same cut-off and the cut-off for the original analysis not being provided. The results remained not statistically significant, however. Frequencies for categories related to smoking status and history of CHF were consistent between the two analyses. Being a current smoker remains statistically significant. Having a history of CHF in 30 days before surgery did not show as statistically significant in the original analysis, but was showing statistical significance in the replication even when looking at Fisher's Exact test.

As CHF had a p-value of less than 0.200 in the original analysis, it was already included in the logistic regression model for the subgroup analysis. The replication of Table 8 was

consistent with the previous replications of multivariate regression tables. Meaning, the p-values, odds ratios, and 95% confidence intervals did not align. As far as interpretation, the p-value for the age category of those 80 or older remained statistically significant. Having a history of CHF had a statistically significant p-value in the original analysis but was not found to be statistically significant in the replication of the logistic model for the subgroup analysis. The opposite of this is true for males. This category was not found to be statistically significant in the original analysis but was found to be statistically significant in the replication.

3.2. Measurement and estimation analysis

The purpose of the MEA portion of the replication analysis was to check the model assumptions for each of the statistical tests chosen. First the replication analysis looked at model assumptions for univariate Chi-square analysis. For this analysis the data requirements are that there are two or more categorical variables with two or more categories for each variable. There is independence of observations, or no relationship between subjects in each group. The last assumption is that there is a relatively large sample size with frequencies of at least 5 for over 80% of the cells (*Libguides: SAS tutorials: Chi-square test of Independence*).

Some of the variables displayed in the tables did not meet the first assumption as they were not categorical variables. This includes age, BMI, length of hospital stays, and operative time. For these variables, t-test was used instead of Chi-square analysis. For the variables that did meet the assumption of being categorical variables, all observations were all independent of each other. Some of these variables did not meet the assumption of a large sample size, however. Though the sample size of the study was large, some of the frequencies in the two-by-two tables were low. For these variables, Chi-square was not appropriate and Fisher's Exact was used to determine the p-value and significance of the relationship between the variable and SSI.

Logistic regression was also used in the study. For this analysis, there is an assumption of appropriate outcome structure. This means that for binary logistic regression, the outcome variable must be binary. Similarly to Chi-square, there is an assumption of observation independence. The other assumptions for logistic regression include the assumption of a large sample size and the absence of multicollinearity (*Logistic and linear regression assumptions – lexjansen).*

The outcome structure is binary as our outcome variable is the occurrence of SSI. For this a subject either had SSI or they did not, so the assumption was met. The assumption of observation independence was also met, as well as the assumption of a large sample size. For the assumption of the absence of multicollinearity, this was tested by running a correlation analysis in PROC LOGSITIC. Looking at the correlation tables ran for each of the multivariate logistic regressions, this assumption was met as well. There does not appear to be a significant correlation between any of the variables.

As part of the MEA, variables for age, BMI, ASA Classification, functional health status prior to surgery and CPT codes were coded as a single variable versus indicator variables. Age remained statistically significant when coded this way with a Chi-square p-value less than 0.05. BMI, ASA Classification and CPT codes were all statistically significant as well. In the original Chi-square analysis, there were statistically significant groups within these variables and therefore were included in the multivariate logistic regression. The logistic regression in the original analysis, however, was not standard as it did not include all levels of the indicator variables. Logistic regression models were rerun using the single variables for age, BMI, and CPT codes.

Results for logistic regression in Table 2 using single variables for age and BMI show a p-value greater than 0.05 and therefore results are not statistically significant. When looking at each category for age with the reference group being those 18-29 years old, those over 70 had a statistically significant p-value. When looking at each category for BMI with the reference group being 19 to <25 kg/m2, BMI 40 to <50 kg/m2 was statistically significant. Previously, we had seen older age and higher BMI to be statistically significant predictors of SSI. Male gender, being a current smoker, and being ventilator dependent all remain statistically significant predictors. In the logistic regression model seen in Table 4, BMI is no longer statistically significant overall as the p-value is 0.051, however as seen above the individual groups compared to the reference groups show statistically significant results for BMI 40 to <50 kg/m2. Male gender, being a current smoker, and being ventilator dependent all remain statistically significant predictors for SSI in this model. CPT code is also statistically significant where it was not in the original analysis.

Table 1A: Demographics for the two groups, original versus replication results

	Panel A: original paper				Panel B: replication results				MEA results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Age, mean (SD), year	51.4 (18.8)	51.4 (14.5)	51.4 (14.8)	0.985	51.4 (14.8)	51.4 (14.5)	51.4 (14.8)	NA	NA
Age 18 to 29, year	4471 (7.8)	11 (4.8)	4482 (7.8)	0.092	4471 (7.8)	11 (4.8)	4482 (7.8)	0.092	0.0480
Age 30 to 39, year	8856 (15.5)	41 (18.0)	8897 (15.5)	0.304	8856 (15.5)	41 (18.0)	8897 (15.5)	0.301	
Age 40 to 49, year	11956 (20.9)	49 (21.5)	12005 (20.9)	0.838	11956 (20.9)	49 (21.5)	12005 (20.9)	0.833	
Age 50 to 59, year	13987 (24.5)	61 (26.8)	14048 (24.5)	0.429	13987 (24.5)	61 (26.8)	14048 (24.5)	0.425	
Age 60 to 69, year	11103 (19.4)	43 (18.9)	11146 (19.4)	0.823	11103 (19.4)	43 (18.9)	11146 (19.4)	0.828	
Age 70 to 79, year	5516 (9.7)	13 (5.7)	5529 (9.6)	**0.043**	5516 (9.7)	13 (5.7)	5529 (9.6)	**0.044**	
Age ≥ 80, year	1207 (2.1)	10 (4.4)	1217 (2.1)	**0.032**	1254 (2.2)	10 (4.4)	1264 (2.2)	**0.025**	
Gender, male	11539 (20.1)	70 (30.7)	11539 (20.1)	**<0.001**	11469 (20.1)	70 (30.7)	11539 (20.1)	**<0.001**	NA
BMI, mean (SD), kg/ m^2	28.9 (25.0-34.1)	30.8 (25.9-35.7)	28.9 (25.0-34.1)	**0.008**	30.1 (7.7)	31.7 (7.5)	30.2 (7.7)	NA	NA
BMI <19 kg/m^2	604 (1.5)	3 (1.8)	607 (1.5)	0.739	893 (1.6)	3 (1.3)	896 (1.6)	0.764	0.0148
BMI 19 to <25 kg/m^2	9690 (23.9)	30 (18.3)	9720 (23.9)	0.092	13573 (23.8)	41 (18.0)	13614 (23.7)	**0.041**	

Table 1A: Demographics for the two groups, original versus replication results

	Panel A: original paper				Panel B: replication results				MEA results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
BMI 25 to <30 kg/m^2	12484 (30.8)	46 (28.0)	12530 (30.8)	0.446	17335 (30.3)	64 (28.1)	17399 (30.3)	0.458	
BMI 30 to <35 kg/m^2	8801 (21.7)	38 (23.2)	8839 (21.7)	0.652	12481 (21.8)	49 (21.5)	12530 (21.8)	0.898	
BMI 35 to <40 kg/m^2	4861 (12.0)	23 (14.0)	4884 (12.0)	0.425	6909 (12.1)	36 (15.8)	6945 (12.1)	0.088	
BMI 40 to <50 kg/m^2	3275 (8.1)	22 (13.4)	3297 (8.1)	**0.012**	4694 (8.2)	32 (14.0)	4726 (8.2)	**0.001**	
BMI≥ 50 kg/m^2	737 (1.8)	2 (1.2)	739 (1.8)	0.773	1032 (1.8)	3 (1.3)	1035 (1.8)	0.579	
BMI> 35 kg/m^2	8771 (21.6)	47 (28.7)	8818 (21.7)	**0.030**	12635 (22.1)	71 (31.1)	12706 (22.2)	**0.001**	**NA**
ASA class 1	4331 (7.6)	9 (4.0)	4340 (7.6)	**0.039**	4331 (7.6)	9 (4.0)	4340 (7.6)	0.039	<.0001
ASA class 2	35347 (61.9)	123 (54.2)	35470 (61.8)	**0.016**	35347 (61.9)	123 (54.0)	35470 (61.8)	**0.014**	
ASA class 3	16662 (29.2)	93 (41.0)	16755 (29.2)	**<0.001**	16662 (29.2)	93 (40.8)	16755 (29.2)	**<0.001**	
ASA class 4	722 (1.3)	1 (0.4)	723 (1.3)	0.540	722 (1.3)	1 (0.4)	723 (1.3)	0.265	
ASA class 5	4 (< 0.1%)	1 (0.4)	5 (< 0.1%)	**0.020**	4 (< 0.1%)	1 (0.4)	5 (< 0.1%)	**<0.001**	
Diabetes	7316 (12.8)	42 (18.4)	7358 (12.8)	**0.011**	7316 (12.8)	42 (18.4)	7358 (12.8)	0.011	NA

Table 1A: Demographics for the two groups, original versus replication results

	Panel A: original paper				Panel B: replication results				MEA results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Race, White	39854 (69.7)	148 (64.9)	40002 (69.7)	0.113	39854 (69.7)	148 (64.9)	40002 (69.7)	0.113	NA
Race, Black	7362 (12.9)	24 (10.5)	7386 (12.9)	0.289	7362 (12.9)	24 (10.5)	7386 (12.9)	0.289	NA
Principal anesthesia, other than general	175 (0.3)	0	175 (0.3)	1.000	173 (0.3)	0	173 (0.3)	0.4054	1.000
Open wound/Wound infection	101 (0.2)	0	101 (0.2)	1.000	101 (0.2)	0	101 (0.2)	0.525	1.000
Steroid use	1319 (2.3)	6 (2.6)	1325 (2.3)	0.746	1319 (2.3)	6 (2.6)	1325 (2.3)	0.746	NA
Weight loss > 10% in the last 6 months	284 (0.5)	2 (0.9)	286 (0.5)	0.315	284 (0.5)	2 (0.9)	286 (0.5)	0.416	0.315
History of severe COPD	1253 (2.2)	9 (3.9)	1262 (2.2)	0.071	1253 (2.2)	9 (4.0)	1262 (2.2)	0.071	NA
Current Smoker	8277 (14.5)	55 (24.1)	8332 (14.5)	**<0.001**	8277 (14.5)	55 (24.1)	8332 (14.5)	**<0.001**	NA
History of CHF in 30 days before surgery	144 (0.3)	2 (0.9)	146 (0.3)	0.115	144 (0.3)	2 (0.9)	146 (0.3)	0.062	0.115
History of ascites	28 (< 0.1%)	0	28 (< 0.1%)	1.000	28 (< 0.1%)	0	28 (< 0.1%)	0.738	1.000
Currently on dialysis (pre-operative)	228 (0.4)	1 (0.4)	229 (0.4)	0.599	228 (0.4)	1 (0.4)	229 (0.4)	0.925	0.599
Hypertension	21755 (38.1)	97 (42.5)	21852 (38.1)	0.165	21755 (38.1)	97 (42.5)	21852 (38.1)	0.165	NA

Table 1A: Demographics for the two groups, original versus replication results

	Panel A: original paper				Panel B: replication results				MEA results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Disseminated Cancer	512 (0.9)	5 (2.2)	517 (0.9)	0.057	512 (0.9)	5 (2.2)	517 (0.9)	**0.039**	0.057
Hypo-albuminemia	1350 (5.8)	5 (4.7)	1355 (5.8)	0.610	34079 (59.6)	122 (53.5)	34201 (59.6)	0.060	NA
Transfusion in 72 h before surgery	11 (< 0.1%)	0	11 (< 0.1%)	1.000	11 (< 0.1%)	0	11 (< 0.1%)	0.834	1.000
Ventilator dependent within 48 h preceding surgery	31 (0.1)	2 (0.9)	33 (0.1)	**0.008**	31 (0.1)	2 (0.9)	33 (0.1)	**<0.0001**	0.008
Totally dependent functional status before surgery	26 (0)	0	26 (0)	1.000	26 (0.1)	0	26 (0.1)	0.747	0.385
Partially dependent functional status before surgery	247 (0.4)	1 (0.4)	248 (0.4)	0.625	247 (0.4)	1 (0.4)	248 (0.4)	0.988	
Independent functional status before surgery	56581 (99.5)	224 (99.6)	56805 (99.5)	1.000	56581 (99.0)	224 (98.3)	56805 (99.0)	0.240	

Note: Categorical variables, n (%)

Table 2A: Multivariate logistic regression to predict surgical site infection, pre-operative; original versus replication results

	Panel A: original paper			Panel B: replication results		
Variable	P-value	Odds Ratio	95% Confidence interval	P-value	Odds Ratio	95% Confidence interval
Age 18–29	0.082	0.480	0.210–1.097	0.168	1.544	0.833-2.861
Age 70–79	0.269	0.701	0.373–1.317	0.050	1.779	1.000-3.163
Age≥80	**0.022**	2.382	1.131–5.016	**0.041**	0.502	0.259-0.972

Table 2A: Multivariate logistic regression to predict surgical site infection, pre-operative; original versus replication results

	Panel A: original paper			Panel B: replication results		
Variable	P-value	Odds Ratio	95% Confidence interval	P-value	Odds Ratio	95% Confidence interval
Gender, male	**<0.001**	2.028	1.457–2.823	**<0.001**	0.574	0.431-0.765
BMI 19 to <25	0.416	0.843	0.558–1.272	0.300	1.204	0.847-1.712
BMI 40 to <50	**0.017**	1.768	1.106–2.826	**0.005**	0.575	0.389-0.849
Diabetes	0.077	1.462	0.959–2.230	0.117	0.750	0.523-1.075
Race, White	0.123	0.775	0.560–1.071	0.107	1.253	0.952-1.648
COPD	0.702	0.820	0.296–2.271	0.467	0.774	0.387-1.545
Current smoker	**<0.001**	1.825	1.259–2.645	**<0.001**	0.540	0.396-0.737
CHF	0.158	2.959	0.657–13.337	0.365	0.510	0.119-2.188
Hypertension	0.281	0.825	0.582–1.170	0.840	1.031	0.768-1.383
Disseminated Cancer	0.316	1.809	0.568–5.768	0.089	0.459	0.187-1.127
Ventilator dependent within 48h preceding surgery	**<0.001**	14.524	3.109–67.835	**0.003**	0.103	0.023-0.459

Table 2B: MEA Multivariate Logistic Regression Model

Variable	Odds Ratio	95% Confidence Interval	P-value	Overall p-value
Age 30-39 vs. 18-29	1.769	0.907-3.451	0.232	0.053
Age 40-49 vs. 18-29	1.486	0.768-2.877	0.863	
Age 50-59 vs. 18-29	1.512	0.784-2.914	0.756	
Age 60-69 vs. 18-29	1.332	0.668-2.656	0.597	
Age 70-79 vs. 18-29	0.840	0.363-1.943	**0.034**	
Age≥80 vs. 18-29	2.997	1.217-7.382	**0.013**	

Table 2B: MEA Multivariate Logistic Regression Model

Variable	Odds Ratio	95% Confidence Interval	P-value	Overall p-value
BMI < 19 kg/m^2 vs. 19 to <25 kg/m^2	1.046	0.322-3.392	0.768	0.050
BMI 25 to <30 kg/m^2 vs. 19 to <25 kg/m^2	1.126	0.757-1.675	0.654	
BMI 30 to <35 kg/m^2 vs. 19 to <25 kg/m^2	1.193	0.780-1.824	0.920	
BMI 35 to <40 kg/m^2 vs. 19 to <25 kg/m^2	1.584	0.999-2.511	0.164	
BMI 40 to <50 kg/m^2 vs. 19 to <25 kg/m^2	2.093	1.292-3.392	**0.006**	
BMI≥ 50 kg/m^2 vs. 19 to <25 kg/m^2	0.836	0.255-2.743	0.463	
Gender, male	0.564	0.423-0.753	**<0.001**	**<0.001**
Diabetes	0.752	0.523-1.082	0.125	0.125
Race, White	1.249	0.949-1.644	0.113	0.113
COPD	0.752	0.375-1.506	0.421	0.421
Current smoker	0.549	0.401-0.750	**<0.001**	**<0.001**
CHF	0.507	0.117-2.185	0.362	0.362
Hypertension	0.977	0.730-1.362	0.985	0.985
Disseminated Cancer	0.447	0.182-1.099	0.079	0.079
Ventilator dependent within 48h preceding surgery	0.100	0.023-0.447	**0.003**	**0.003**

Table 3A: Intra- and post-operative characteristics., original versus replication results

	Panel A: original paper				**Panel B: replication results**				**MEA Results**
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Operative time, median (IQR), minute	102 (74–140)	118.5 (86.0–166.8)	102 (74–140)	**<0.001**	102 (74–140)	118.5 (86.0–166.5)	102 (74–140)	**<0.001**	NA
Elective surgery	55766 (97.6)	219 (96.1)	55985 (97.6)	0.131	55766 (97.6)	219 (96.1)	55985 (97.6)	0.131	NA
Partial thyroid lobectomy, unilateral; with or without isthmusectomy-60210	3386 (5.9)	16 (7.0)	3402 (5.9)	0.486	3386 (5.9)	16 (7.0)	3402 (5.9)	0.486	**0.002**
Partial thyroid lobectomy, unilateral; with contralateral subtotal lobectomy, including isthmusectomy-60212	335 (0.6)	4 (1.8)	339 (0.6)	**0.047**	335 (0.6)	4 (1.8)	339 (0.6)	**0.022**	**0.002**
Total thyroid lobectomy, unilateral; with or without isthmusectomy-60220	16616 (29.1)	44 (19.3)	16660 (29.0)	**0.001**	16616 (29.1)	44 (19.3)	16660 (29.0)	**0.001**	**0.002**
Total thyroid lobectomy, unilateral; with contralateral subtotal lobectomy, including isthmusectomy-60225	1081 (1.9)	9 (3.9)	1090 (1.9)	**0.044**	1081 (1.9)	9 (4.0)	1090 (1.9)	**0.023**	**0.002**

Table 3A: Intra- and post-operative characteristics., original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Thyroidectomy, total or complete-60240	26693 (46.7)	107 (46.9)	26800 (46.7)	0.948	26693 (46.7)	107 (46.9)	26800 (46.7)	0.948	**0.002**
Thyroidectomy, total or subtotal for malignancy; with limited neck dissection-60252	5589 (9.8)	31 (13.6)	5620 (9.8)	0.053	5589 (9.8)	31 (13.6)	5620 (9.8)	0.053	**0.002**
Thyroidectomy, total or subtotal for malignancy; with radical neck dissection-60254	912 (1.6)	4 (1.8)	916 (1.6)	0.787	912 (1.6)	4 (1.8)	916 (1.6)	0.849	**0.002**
Thyroidectomy, removal of all remaining thyroid tissue following previous removal of a portion of thyroid-60260	2531 (4.4)	13 (5.7)	2544 (4.4)	0.332	2531 (4.4)	13 (5.7)	2544 (4.4)	0.352	**0.002**
Transfusion intra-operative, post-operative	93 (0.2)	2 (0.9)	95 (0.2)	0.055	93 (0.2)	2 (0.9)	95 (0.2)	**0.008**	0.055
Occurrence myocardial infarction	32 (0.1)	0	32 (0.1)	1.000	32 (0.1)	0	32 (0.1)	0.721	1.000
Occurrence sepsis	39 (0.1)	18 (7.9)	57 (0.1)	**<0.001**	39 (0.1)	18 (7.9)	57 (0.1)	**<0.001**	<.0001
Occurrence Septic Shock	8 (0)	4 (1.8)	12 (0)	**<0.001**	8 (0)	4 (1.8)	12 (0)	**<0.001**	<.0001
Occurrence pneumonia	107 (0.2)	8 (3.5)	115 (0.2)	**<0.001**	107 (0.2)	8 (3.5)	115 (0.2)	**<0.001**	<.0001

Table 3A: Intra- and post-operative characteristics., original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Occurrence unplanned intubation	184 (0.3)	9 (3.9)	193 (0.3)	**<0.001**	184 (0.3)	9 (4.0)	193 (0.3)	**<0.001**	<.0001
Occurrence ventilator >48 h	91 (0.2)	8 (3.5)	99 (0.2)	**<0.001**	91 (0.2)	8 (3.5)	99 (0.2)	**<0.001**	<.0001
Occurrence deep vein thrombosis	30 (0.1)	1 (0.4)	31 (0.1)	0.116	30 (0.1)	1 (0.4)	31 (0.1)	**0.012**	0.116
Occurrence pulmonary emboli	31 (0.1)	2 (0.9)	33 (0.1)	**0.008**	31 (0.1)	2 (0.9)	33 (0.1)	**<0.0001**	0.008
Occurrence wound disruption	22 (< 0.1%)	1 (0.4)	23 (< 0.1%)	0.088	22 (< 0.1%)	1 (0.4)	23 (< 0.1%)	**0.003**	0.088
Urinary Tract Infection	138 (0.2)	0	138 (0.2)	1.000	138 (0.2)	0	138 (0.2)	0.456	1.000
Cardiac arrest	18 (0)	1 (0.4)	19 (0)	0.073	18 (0)	1 (0.4)	19 (0)	**0.0007**	0.073
Cerebrovascular accident	13 (0)	0	13 (0)	1.000	13 (0)	0	13 (0)	0.820	1.000
Acute renal failure	4 (0)	2 (0.9)	6 (0)	**<0.001**	4 (0)	2 (0.9)	6 (0)	**<0.001**	0.0002
Progressive renal failure	17 (0)	0	17 (0)	1.000	17 (0)	0	17 (0)	0.795	1.000
Return to the operating room	717 (1.3)	50 (21.9)	767 (1.3)	**<0.001**	717 (1.3)	50 (21.9)	767 (1.3)	**<0.001**	<.0001
Length of hospital stay	1 (1-1)	1 (1–2)	1 (1-1)	**<0.001**	1 (1-1)	1 (1–2)	1 (1-1)	**<0.001**	NA

Table 3A: Intra- and post-operative characteristics., original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
PoDX- malignant neoplasm of thyroid gland	17312 (30.3)	78 (34.2)	17390 (30.3)	0.199	17312 (30.3)	78 (34.2)	17390 (30.3)	0.199	NA
PoDX-Nontoxic multi-nodular goiter	10874 (19.0)	42 (18.4)	10916 (19.0)	0.815	10874 (19.0)	42 (18.4)	10916 (19.0)	0.815	NA
PoDx- Nontoxic uni-nodular goiter	6520 (11.4)	15 (6.6)	6535 (11.4)	**0.022**	6520 (11.4)	15 (6.6)	6535 (11.4)	**0.022**	NA
PoDx-Benign neoplasm of thyroid gland	5063 (8.9)	9 (3.9)	5072 (8.8)	**0.009**	5063 (8.9)	9 (4.0)	5072 (8.8)	**0.009**	NA
PoDX-Toxic diffuse goiter without thyrotoxicosis or storm	2197 (3.8)	12 (5.3)	2209 (3.9)	0.267	2197 (3.8)	12 (5.3)	2209 (3.9)	0.267	NA
PoDx-Goiter unspecified	1751 (3.1)	7 (3.1)	1758 (3.1)	0.996	1751 (3.1)	7 (3.1)	1758 (3.1)	0.996	NA
PoDx-Chronic lymphocytic thyroiditis	1235 (2.2)	4 (1.8)	1239 (2.2)	0.822	1235 (2.2)	4 (1.8)	1239 (2.2)	0.673	1.000
PoDx-Unspecified nontoxic nodular goiter	790 (1.4)	1 (0.4)	791 (1.4)	0.273	790 (1.4)	1 (0.4)	791 (1.4)	0.223	0.384
PoDx-Toxic multinodular goiter without thyrotoxicosis or storm	831 (1.5)	5 (2.2)	836 (1.5)	0.394	831 (1.5)	5 (2.2)	836 (1.5)	0.356	0.394
PoDx- neoplasm of uncertain behavior of other and unspecified endocrine glands	916 (1.6)	3 (1.3)	919 (1.6)	0.806	916 (1.6)	3 (1.3)	919 (1.6)	0.730	1.000

Table 3A: Intra- and post-operative characteristics., original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (57143)	With SSI (228)	Total (57371)	P-value	No SSI (57143)	With SSI (228)	Total (57371)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
PoDx-Other	4358 (7.6)	29 (12.7)	4387 (7.6)	**0.004**	4358 (7.6)	29 (12.7)	4387 (7.6)	**0.004**	NA
PoDx-Unknown	5296 (9.3)	23 (10.1)	5319 (9.3)	0.670	5296 (9.3)	23 (10.1)	5319 (9.3)	0.670	NA

Note: Values represent n (%), PoDx, post-operative diagnosis.

Table 4A: Multivariate logistic regression to predict surgical site infection, original versus replication results

	Panel A: original paper			Panel B: replication results		
Variable	P-value	Odds Ratio	95% Confidence interval	P-value	Odds Ratio	95% Confidence interval
Age 18–29	0.081	0.479	0.209–1.094	0.159	1.558	0.841-2.888
Age 70–79	0.280	0.706	0.376–1.327	0.055	1.756	0.988-3.123
Age≥80	**0.020**	2.413	1.147–5.078	**0.038**	0.496	0.256-0.962
Gender, male	**<0.001**	2.044	1.468–2.846	**<0.001**	0.570	0.428-0.759
BMI 19 to <25	0.447	0.852	0.564–1.287	0.330	1.191	0.838-1.694
BMI 40 to <50	**0.017**	1.773	1.109–2.836	**0.005**	0.574	0.388-0.847
Diabetes	0.082	1.453	0.953–2.216	0.132	0.758	0.529-1.087
Race, White	0.126	0.776	0.561–1.074	0.115	1.247	0.947-1.640
COPD	0.662	0.795	0.285–2.220	0.473	0.776	0.388-1.551
Current smoker	**<0.002**	1.797	1.239–2.606	**<0.001**	0.549	0.402-0.750

Table 4A: Multivariate logistic regression to predict surgical site infection, original versus replication results

	Panel A: original paper			Panel B: replication results		
Variable	P-value	Odds Ratio	95% Confidence interval	P-value	Odds Ratio	95% Confidence interval
CHF	0.160	2.931	0.654–13.140	0.361	0.510	0.120-2.165
Hypertension	0.255	0.816	0.575–1.158	0.774	1.044	0.778-1.401
Disseminated Cancer	0.345	1.749	0.549–5.576	0.113	0.484	0.197-1.189
Ventilator dependent within 48h preceding surgery	**<0.001**	13.469	2.858–63.462	**0.006**	0.122	0.027-0.551
Partial thyroid lobectomy, unilateral; with contralateral subtotal lobectomy, including isthmusectomy-60212	0.092	2.729	0.850–8.762	0.076	0.400	0.145-1.099
Total thyroid lobectomy, unilateral; with or without isthmusectomy-60220	0.120	0.743	0.511–1.081	**0.006**	1.597	1.145-2.228
Total thyroid lobectomy, unilateral; with contralateral subtotal lobectomy, including isthmusectomy-60225	0.189	1.739	0.762–3.969	0.072	0.538	0.274-1.057

Table 4B: MEA Multivariate Logistic Regression Model

Variable	Odds Ratio	95% Confidence Interval	P-value	Overall P-value
Age 30-39 vs. 18-29	1.791	0.918-3.494	0.259	0.053
Age 40-49 vs. 18-29	1.512	0.781-2.930	0.887	
Age 50-59 vs. 18-29	1.542	0.799-2.975	0.768	
Age 60-69 vs. 18-29	1.369	0.686-2.734	0.624	
Age 70-79 vs. 18-29	0.873	0.377-2.020	**0.039**	
Age≥80 vs. 18-29	3.114	1.262-7.681	**0.011**	
BMI < 19 kg/m^2 vs. 19 to <25 kg/m^2	1.046	0.322-3.396	0.775	0.051
BMI 25 to <30 kg/m^2 vs. 19 to <25 kg/m^2	1.113	0.748-1.656	0.622	

Table 4B: MEA Multivariate Logistic Regression Model

Variable	Odds Ratio	95% Confidence Interval	P-value	Overall P-value
BMI 30 to <35 kg/m^2 vs. 19 to <25 kg/m^2	1.182	0.773-1.807	0.896	
BMI 35 to <40 kg/m^2 vs. 19 to <25 kg/m^2	1.574	0.993-2.496	0.167	
BMI 40 to <50 kg/m^2 vs. 19 to <25 kg/m^2	2.083	1.285-3.377	**0.006**	
BMI≥ 50 kg/m^2 vs. 19 to <25 kg/m^2	0.838	0.255-2.752	0.471	
Gender, male	0.570	0.426-0.761	**<0.001**	**<0.001**
Diabetes	0.760	0.528-1.093	0.139	0.139
Race, White	1.253	0.951-1.652	0.109	0.109
COPD	0.751	0.374-1.508	0.421	0.421
Current smoker	0.550	0.402-0.753	**<0.001**	**<0.001**
CHF	0.502	0.118-2.141	0.352	0.352
Hypertension	1.002	0.733-1.370	0.989	0.989
Disseminated Cancer	0.488	0.197-1.207	0.121	0.121
Ventilator dependent within 48h preceding surgery	0.116	0.026-0.524	**0.005**	**0.005**
CPT Code 60210 vs. 60260	0.884	0.424-1.845	0.680	**0.008**
CPT Code 60212 vs. 60260	2.109	0.674-6.595	0.093	
CPT Code 60220 vs. 60260	0.519	0.279-0.966	**<0.001**	
CPT Code 60225 vs. 60260	1.539	0.654-3.620	0.147	
CPT Code 60240 vs. 60260	0.756	0.424-1.349	0.070	
CPT Code 60252 vs. 60260	1.059	0.551-2.033	0.686	
CPT Code 60254 vs. 60260	0.705	0.228-2.182	0.468	

Table 6A: Subgroup Analysis, pre-operative risk factors for SSI; original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (6501)	With SSI (35)	Total (6536)	P-value	No SSI (6501)	With SSI (35)	Total (6536)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Age, mean (SD), year	48.7 (15.2)	52.5 (14.0)	48.7 (15.2)	0.133	48.7 (15.2)	52.5 (14.0)	48.7 (15.2)	NA	**<.0001**
Age 18 to 29, year	762 (11.7)	1 (2.9)	763 (11.7)	0.117	762 (11.7)	1 (2.9)	763 (11.7)	0.103	0.030
Age 30 to 39, year	1215 (18.7)	6 (17.1)	1221 (18.7)	0.815	1215 (18.7)	6 (17.1)	1221 (18.7)	0.815	
Age 40 to 49, year	1402 (21.6)	6 (17.1)	1408 (21.5)	0.525	1402 (21.6)	6 (17.1)	1408 (21.5)	0.526	
Age 50 to 59, year	1446 (22.2)	12 (34.3)	1458 (22.3)	0.088	1446 (22.2)	12 (34.3)	1458 (22.3)	0.088	
Age 60 to 69, year	1064 (16.4)	6 (17.1)	1070 (16.4)	0.902	1064 (16.4)	6 (17.1)	1070 (16.4)	0.902	
Age 70 to 79, year	488 (7.5)	1 (2.9)	489 (7.5)	0.515	488 (7.5)	1 (2.9)	489 (7.5)	0.297	
Age ≥ 80, year	123 (1.9)	3 (8.6)	126 (1.9)	**0.029**	124 (1.9)	3 (8.6)	127 (1.9)	**0.004**	
Gender, male	1761 (27.1)	14 (40.0)	1775 (27.2)	0.087	1761 (27.1)	14 (40.0)	1775 (27.2)	0.087	NA
BMI, median (IQR), kg/ m2	28.3 (24.7– 33.6)	29.8 (26.5– 36.6)	28.3 (24.7– 33.6)	0.184	28.3 (24.6 – 33.5)	30.3 (26.4 – 36.3)	28.3 (24.6 – 33.5)		**<.0001**
BMI <19 kg/m2	68 (1.5)	1 (4.0)	69 (1.5)	0.313	119 (1.8)	1 (2.9)	120 (1.8)	0.652	0.378
BMI 19 to <25 kg/m2	1185 (25.7)	2 (8.0)	1187 (25.6)	**0.043**	1670 (25.7)	4 (11.4)	1674 (25.6)	0.0539	

Table 6A: Subgroup Analysis, pre-operative risk factors for SSI; original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (6501)	With SSI (35)	Total (6536)	P-value	No SSI (6501)	With SSI (35)	Total (6536)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
BMI 25 to <30 kg/m2	1505 (32.6)	10 (40.0)	1515 (32.6)	0.432	2088 (32.1)	12 (34.3)	2100 (32.1)	0.784	
BMI 30 to <35 kg/m2	912 (19.8)	5 (20.0)	917 (19.8)	1.000	1314 (20.1)	9 (25.7)	1323 (20.2)	0.4191	
BMI 35 to <40 kg/m2	516 (11.2)	3 (12.0)	519 (11.2)	0.754	720 (11.1)	3 (8.6)	723 (11.1)	0.638	
BMI 40 to <50 kg/m2	357 (7.7)	3 (12.0)	360 (7.8)	0.438	490 (7.5)	5 (14.3)	495 (7.6)	0.132	
BMI≥ 50 kg/m2	66 (1.4)	1 (4.0)	67 (1.4)	0.305	95 (1.5)	1 (2.9)	96 (1.5)	0.4936	
BMI> 35 kg/m2	931 (20.2)	7 (28.0)	938 (20.2)	0.331	1305 (20.1)	9 (25.7)	1314 (20.1)	0.406	NA
ASA class 1	526 (8.1)	2 (5.7)	528 (8.1)	1.000	526 (8.1)	2 (5.7)	528 (8.1)	0.607	0.843
ASA class 2	4131 (63.6)	21 (60.0)	4152 (63.6)	0.656	4131 (63.5)	21 (60.0)	4152 (63.5)	0.664	
ASA class 3	1763 (27.2)	12 (34.3)	1775 (27.2)	0.344	1763 (27.1)	12 (34.3)	1775 (27.2)	0.342	
ASA class 4	72 (1.1)	0	72 (1.1)	1.000	72 (1.1)	0	72 (1.1)	0.531	
ASA class 5	0	0	0	1.000	0	0	0		
Diabetes	759 (11.7)	5 (14.3)	764 (11.7)	0.596	759 (11.7)	5 (14.3)	764 (11.7)	0.632	0.596
Race, White	4914 (75.6)	23 (65.7)	4937 (75.5)	0.175	4914 (75.6)	23 (65.7)	4937 (75.5)	0.175	NA

Table 6A: Subgroup Analysis, pre-operative risk factors for SSI; original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (6501)	With SSI (35)	Total (6536)	P-value	No SSI (6501)	With SSI (35)	Total (6536)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Race, Black	326 (5.0)	2 (5.7)	328 (5.0)	0.695	326 (5.0)	2 (5.7)	328 (5.0)	0.850	0.695
Principal anesthesia, other than general	21 (0.3)	0	21 (0.3)	1.000	21 (0.3)	0	21 (0.3)	0.736	1.000
Open wound/Wound infection	11 (0.2)	0	11 (0.2)	1.000	11 (0.2)	0	11 (0.2)	0.808	1.000
Steroid use	136 (2.1)	0	136 (2.1)	1.000	136 (2.1)	0	136 (2.1)	0.387	1.000
Weight loss > 10% in the last 6 months	26 (0.4)	0	26 (0.4)	1.000	26 (0.4)	0	26 (0.4)	0.708	1.000
History of severe COPD	106 (1.6)	2 (5.7)	108 (1.7)	0.113	106 (1.6)	2 (5.7)	108 (1.7)	0.059	0.113
Current Smoker	807 (12.4)	9 (25.7)	816 (12.5)	**0.034**	807 (12.4)	9 (25.7)	816 (12.5)	**0.018**	**0.034**
History of CHF in 30 days before surgery	18 (0.3)	1 (2.9)	19 (0.3)	0.097	18 (0.3)	1 (2.9)	19 (0.3)	**0.005**	0.097
History of ascites	3 (< 0.1)	0	3 (< 0.1)	1.000	3 (< 0.1)	0	3 (< 0.1)	0.899	1.000
Currently on dialysis (pre-operative)	24 (0.4)	0	24 (0.4)	1.000	24 (0.4)	0	24 (0.4)	0.719	1.000
Hypertension	2052 (31.6)	13 (37.1)	2065 (31.6)	0.479	2052 (31.6)	13 (37.1)	2065 (31.6)	0.479	NA
Disseminated Cancer	178 (2.7)	1 (2.9)	179 (2.7)	0.623	178 (2.7)	1 (2.9)	179 (2.7)	0.966	0.623

Table 6A: Subgroup Analysis, pre-operative risk factors for SSI; original versus replication results

	Panel A: original paper				Panel B: replication results				MEA Results
Variable	No SSI (6501)	With SSI (35)	Total (6536)	P-value	No SSI (6501)	With SSI (35)	Total (6536)	P-value	P-value *Fisher's Exact is displayed if Chi-square model assumptions are not met
Hypo-albuminemia	121 (4.5)	0	121 (4.5)	1.000	3815 (58.7)	19 (54.3)	3834 (58.7)	0.598	NA
Totally dependent functional status before surgery	3	0	3 (0)		3 (0.1)	0	3 (0.1)	0.899	0.969
Partially dependent functional status before surgery	25 (0.4)	0	25 (0.4)	1.000	25 (0.4)	0	25 (0.4)	0.713	
Independent functional status before surgery	6455 (99.3)	35 (100)	6490 (99.3)	1.000	6455 (99.3)	35 (100)	6490 (99.3)	0.618	

Note: Categorical variables, n (%)

Transfusion in 72 h before surgery and ventilator dependent within 48 h preceding surgery returned no value for the SSI group. Therefore, not mentioned in the table.

Table 8A: Multivariate pre-operative risk factors for SSI in the cancer subgroup; original versus replication results

	Panel A: original paper			Panel B: replication results		
Variable	P-value	Odds Ratio	95% Confidence interval	P-value	Odds Ratio	95% Confidence interval
Age 18–29	0.993	0	0 – Not-computable	0.195	1.503	0.811 – 2.784
Age 50–59	0.090	2.054	2.054 0.894–4.722	0.599	0.923	0.684 – 1.245
Age≥80	**0.038**	5.083	1.096–23.575	**0.016**	0.452	0.236 – 0.864
Gender, male	0.155	1.800	0.801–4.043	**<0.001**	0.590	0.444 – 0.784
BMI 19 to <25	0.106	0.301	0.070–1.291	0.094	1.339	0.952 – 1.883

Table 8A: Multivariate pre-operative risk factors for SSI in the cancer subgroup; original versus replication results

	Panel A: original paper			**Panel B: replication results**		
Variable	P-value	Odds Ratio	95% Confidence interval	P-value	Odds Ratio	95% Confidence interval
Race, White	0.036	0.419	0.185–0.945	0.071	1.286	0.979 – 1.691
COPD	0.611	1.720	0.213–13.901	0.389	0.741	0.374 – 1.466
Current smoker	0.139	2.073	0.789–5.444	**<0.001**	0.527	0.386 – 0.718
CHF	**0.019**	12.463	1.505–103.198	0.170	0.372	0.090 – 1.526

Table 8B: MEA Multivariate Logistic Regression Model

Variable	Odds Ratio	95% Confidence Interval	P-value	Overall P-value
Age 30-39 vs. 18-29	1.772	0.909-3.457	0.337	**0.047**
Age 40-49 vs. 18-29	1.510	0.782-2.915	0.962	
Age 50-59 vs. 18-29	1.573	0.823-3.004	0.807	
Age 60-69 vs. 18-29	1.432	0.733-2.797	0.699	
Age 70-79 vs. 18-29	0.918	0.408-2.067	**0.045**	
Age≥80 vs. 18-29	3.397	1.425-8.095	**0.005**	
BMI < 19 kg/m^2 vs. 19 to <25 kg/m^2	1.049	0.323-3.402	0.714	**0.017**
BMI 25 to <30 kg/m^2 vs. 19 to <25 kg/m^2	1.140	0.767-1.694	0.539	
BMI 30 to <35 kg/m^2 vs. 19 to <25 kg/m^2	1.230	0.809-1.871	0.883	
BMI 35 to <40 kg/m^2 vs. 19 to <25 kg/m^2	1.664	1.059-2.613	0.147	

Table 8B: MEA Multivariate Logistic Regression Model

Variable	Odds Ratio	95% Confidence Interval	P-value	Overall P-value
BMI 40 to <50 kg/m^2 vs. 19 to <25 kg/m^2	2.222	1.393-3.545	**0.004**	
BMI≥ 50 kg/m^2 vs. 19 to <25 kg/m^2	0.940	0.290-3.050	0.560	
Gender, male	0.553	0.415-0.737	**<.001**	**<.001**
Race, White	1.268	0.964-1.667	0.089	0.089
COPD	0.705	0.353-1.408	0.322	0.322
Current smoker	0.545	0.398-0.745	**<0.001**	**<0.001**
CHF	0.388	0.094-1.605	0.191	0.191

3.3. Theory of change analysis

The TCA looks to expand the original analysis and give more robust results. In the original paper, looking for interactions was not mentioned in relation to logistic regression. Looking at table 2, we see that both gender and being a current smoker are strongly significant. If an interaction was determined, then a subgroup analysis would be further investigated. However, a significant interaction was not seen between the two variables. When an interaction was run in the PROC LOGISTIC model for table 2, we see a p-value of 0.358. The logistic regression model for TCA analysis was run including all levels of age and BMI. This result was obtained by using overall categorical variables for both.

4. Discussion

When closely examining the results of the paper, the original authors did not treat the categorical variables with multiple levels as grouped as part of the same variable. They did not perform overall tests of significance for these variables and treated them as indicator variables in the multivariable logistic regression models, allowing the levels of the categorical variables to be selected out of the model, creating an odd reference group for those remaining. As part of the MEA analysis, the correct models were run treating the variables (such as age group and BMI) as single categorical variables rather than individual indicator variables. These sensitivity analyses did show differences from the original models run in the paper. When logistic regression was rerun with single variables including all levels for these variables, we see that age and BMI are not statistically significant predictors of SSI. Male gender, being a current smoker, and being ventilator dependent all remain statistically significant predictors.

For the pure replication analysis, overall, when the replication was performed as described in the original paper, Chi-square analysis are consistent between the replication and original analysis, as seen in the pure replication of the tables. The original analysis was thorough in checking model assumptions and altering the test used if assumptions were not met. Where the original analysis could be improved is to detail this more in the methods section. As they claim that the p-value seen is from a univariate Chi-square test, but this is not the p-value displayed in all cases which is misleading.

A couple of the tables in the original study involved intra- and post-operative risk factors and were not replicated as part of the pure replication analysis. They were not analyzed in the multivariate logistic regression replication analysis as it is unclear whether these factors occurred prior to the SSI, and they may have even been caused by SSI and not the other way around.

These suspect risk factors include sepsis, septic shock, pneumonia, wound disruption, return to operating room, and length of hospital stay.

From the replication analysis, we can confirm that the results of a low significance of SSI is seen for clean procedures. The significance of this, is that routine prescribing of antibiotics for this procedure is not necessary. The study confirms that smoking plays a role in predicting SSI as well as age and gender. Other variables show significance when looking at Chi-square analysis, but don't when looking at logistic regression.

For the tables in the original analysis that represented p-values and odds ratios obtained from multivariate logistic regression, we did not see values that were consistent in the replication analysis. This is likely because BMI was included in the multivariate logistic regression model. As the frequencies for BMI did not match the original analysis, we see differences in the results obtained. Though numbers did not align perfectly for the multivariate logistic regression analysis, we see consistencies in variables that are predictive of surgical site infection between the original analysis and the replication analysis. Based on the logistic regression analysis from both the original and replication analysis we can see that the pre-operative risk factors such as age for those 80 and older, male gender, BMI 40 to <50, being a current smoker, and being ventilator dependent are all predictors of SSI. If individuals have these risk factors, then antibiotics being prescribed as a preventive measure for SSI should then be considered.

As for the subgroup analyses tables, the same number of clean thyroidectomy cases with and without SSI, were identified in the replication analysis as the original analysis. Other than these values and a handful of other frequencies, results were not consistent between the replication analysis and the original analysis. This could be due to the small sample of cases with

SSI for the subgroup. Further research on this subgroup would be needed to confirm any findings.

There is no mention that model assumptions were checked in the original paper. However, when conducting the measurement and estimation analysis, it was clear that the investigators of the study did check model assumptions for both Chi-square analysis and the multivariate logistic regression.

The original study also did not mention interactions being checked between variables, but when the TCA was run, no interactions were seen. To expand on the study more, a future analysis could include recursive partitioning or 'tree analysis' using SAS HPSPLIT.

Limitations seen in the replication analysis were BMI calculations used were not provided and hypo-albuminemia cut-off levels were not provided. Methods also did not detail fully what values were displayed if model assumptions were not met. Other limitations of the study were provided in the original paper and include important variables not being provided in the database. One such variable is the use of prophylactic antibiotics. If most of these procedures did in fact have antibiotics prescribed, then many of the risk factors of SSI could be masked. Due to this, these results should not be taken as stand-alone results and further research is needed to prove which risk factors are most indicative of prophylactic antibiotic use for thyroidectomy. The results are not null, however. They are consistent with previous studies completed in this field and should be considered along with other studies done on prophylactic antibiotic use for thyroidectomy. One of the reasons that even though the study has a major limitation it should not be discarded is due to the importance of the aim of the study and research on limiting unnecessary use of antibiotics.

5. Public Health Implications

Replication and reproducibility are crucial parts of scientific research. It allows us to verify results published in the original study by reproducing the original results using the original data. Replication research also allows us to examine the sensitivity or robustness of the analysis and extend on the analysis from the original study. One of the reasons this is so important is that many studies are found to not be reproducible, meaning the conclusions drawn may not be accurate. In many cases, all it takes is one article to influence policy or common practices and decision making in a hospital.

The significance of the study I chose to replicate is they are looking to find the most effective use of health care resources in respect to surgical site infection or SSI in thyroidectomy procedures. There were three main goals or objectives of the study. The first is to determine the incidence of SSI after thyroidectomy with clean wounds. The second is to identify the pre-operative risk factors predictive of SSI in clean wound thyroidectomy performed in general and for those who specifically have a cancerous lesion. The third goal of the study is to guide thyroid surgeons in the use of selective prophylactic antibiotic use.

As one of the main goals is to influence thyroid surgeons common prescribing practices, we can see why replicating this analysis is so important. Though SSI is rare with this procedure, if it does occur it can cause horrible consequences so if influencing results are misleading, we could see adverse outcomes. The original study wants to show that thyroidectomy does not result in SSI in most cases, and thus routine prescription of antibiotics is not necessary. This is important to prevent Antimicrobial Resistance (AMR), which results when bacteria become resistant to antibiotics. This is important to try and prevent as the more AMR seen the harder it is to treat infections (*Antimicrobial resistance).*

This is a global concern, as if we do not have effective antibiotics, we are not able to treat common infections or perform many surgeries safely. If we over prescribe antibiotics, we will see more and more development of AMR (*Antimicrobial resistance).* That is why studies such as this one are greatly important. If we can determine the procedure does not end in SSI and thus antibiotics are not needed, we can prevent the over prescription of antibiotics and slow the development of AMR. This allows antibiotics to be effective for a longer time and allows more time for research in developing new antibiotics as well. We need more studies that help prevent the unnecessary prescribing of antibiotics.

References

Akirov, A., Masri-Iraqi, H., Atamna, A., & Shimon, I. (2017b). Low albumin levels are associated with mortality risk in hospitalized patients. *The American Journal of Medicine*, *130*(12). https://doi.org/10.1016/j.amjmed.2017.07.020

Libguides: SAS tutorials: Chi-square test of Independence. Chi-Square Test of Independence - SAS Tutorials - LibGuides at Kent State University. (n.d.). https://libguides.library.kent.edu/SAS/ChiSquare

Logistic and linear regression assumptions - lexjansen. (n.d.). https://www.lexjansen.com/wuss/2018/130_Final_Paper_PDF.pdf

Myssiorek, D., Ahmed, Y., Parsikia, A., Castaldi, M., & McNelis, J. (2018a). Factors predictive of the development of surgical site infection in thyroidectomy – an analysis of NSQIP database. *International Journal of Surgery*, *60*, 273–278. https://doi.org/10.1016/j.ijsu.2018.11.013

World Health Organization. (n.d.). *Antimicrobial resistance*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance#:~:text=As%20a%20result%20of%20drug,through%20genetic%20changes%20in%20pathogens.

Appendix

SAS Code for isolating NSQIP data used in the study:

```
*********************************************************;
* prepare data for Kaitlyn                               *;
* Replicating https://doi.org/10.1016/j.ijsu.2018.11.013 *;
*                                                        *;
* eligibility                                            *;
* 2012 until 2015,
* thyroidectomy cases with the CPT codes
* 60210, 60212, 60220, 60225, 60240, 60252, 60254, 60260.
* age of patient was =18 years.
* excluded all cases not classified as Class 1 (clean).
* 1/19/24                                                *;
*********************************************************;

*get data;

libname ssd 'C:\Users\lmsmith\OneDrive - University of Nebraska Medical
Center\U\Capstone_course\NSQIPdata';

data year2012;
   set ssd.acs_nsqip_puf12;
   if CPT in ('60210', '60212', '60220', '60225', '60240', '60252', '60254',
'60260');
   if WNDCLAS='1-Clean';
   year=2012;
run;
data year2013;
   set ssd.acs_nsqip_puf13;
   if CPT in ('60210', '60212', '60220', '60225', '60240', '60252', '60254',
'60260');
   if WNDCLAS='1-Clean';
   year=2013;
run;
data year2014;
   set ssd.acs_nsqip_puf14;
   if CPT in ('60210', '60212', '60220', '60225', '60240', '60252', '60254',
'60260');
   if WNDCLAS='1-Clean';
   year=2014;
run;
data year2015;
   set ssd.acs_nsqip_puf15_v2;
   if CPT in ('60210', '60212', '60220', '60225', '60240', '60252', '60254',
'60260');
   if WNDCLAS='1-Clean';
   year=2015;
run;

libname sse 'C:\Users\lmsmith\OneDrive - University of Nebraska Medical
Center\U\Capstone_course\Students\Kaitlyn Kenig';
data sse.thyroid;
   set year2012 year2013 year2014 year2015;
run;
```

```
proc freq data=thyroid;
tables sex;

run;
```

SAS Code for demographic data manipulation:

```
/* Replicating https://doi.org/10.1016/j.ijsu.2018.11.013 *;
*                                                        *;
* eligibility                                            *;
* 2012 until 2015,
* thyroidectomy cases with the CPT codes
* 60210, 60212, 60220, 60225, 60240, 60252, 60254, 60260.
* age of patient was >=18 years.
* excluded all cases not classified as Class 1 (clean)*/

libname thyroid "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access = readonly;

proc contents data = thyroid.thyroid;
run;

proc freq data = thyroid.thyroid;
table FNSTATUS2;
run;

/* Create dataset */
data SSI;
set thyroid.thyroid (keep = CaseID CPT ORGSPCSSI WNDINFD SUPINFEC SEX AGE
HEIGHT WEIGHT ASACLAS DIABETES RACE_NEW ANESTHES WNDINF STEROID WTLOSS HXCOPD
SMOKE HXCHF ASCITES DIALYSIS HYPERMED DISCANCR PRALBUM TRANSFUS VENTILAT
FNSTATUS2);

/* Create functional status variables */
if FNSTATUS2 = "Independent" then FSIND = "Yes";
else FSIND = "No";

if FNSTATUS2 = "Partially Dependent" then FSPD = "Yes";
else FSPD = "No";

if FNSTATUS2 = "Totally Dependent" the FSTD = "Yes";
else FSTD = "No";

/* Create variable for hypo-albuminemia */
/* Albumin levels for hypoalbuminemia were not defined - got ranges from
https://www.amjmed.com/action/showPdf?pii=S0002-9343%2817%2930800-8 */
if PRALBUM < 2.5 then HYPALBU = "Yes";
else HYPALBU = "No";

/* Create binomial variable for SSI */
if ORGSPCSSI = "No Complication" then SSI = "No";
if WNDINFD = "No Complication" then SSI = "No";
if SUPINFEC = "No Complication" then SSI = "No";

if WNDINFD = "Deep Incisional SSI" then SSI = "Yes";
if ORGSPCSSI = "Organ/Space SSI" then SSI = "Yes";
if SUPINFEC = "Superficial Incisional SSI" then SSI = "Yes";
```

```
/* Create Race variables */
if RACE_NEW = "White" then WHITE = "Yes";
else WHITE = "No";

if RACE_NEW = "Black or African American" then BLACK = "Yes";
else BLACK = "No";

/* Turn age in numeric */
if AGE = "90+" then AGE2 = 90;
else AGE2 = input(AGE, 8.);

/* Calculate BMI */
if WEIGHT > 0 and HEIGHT > 0 then BMI = 703.0768*(WEIGHT)/(HEIGHT*HEIGHT);
else BMI = . ;

/* Create binomial variable for Diabetes */
if DIABETES = "NO" then DIABETES1 = "No";
else DIABETES1 = "Yes";

/* Create variable for principal anethesia other than general */
if ANESTHES = "General" then P_ANESTHES = "No";
else if ANESTHES = "None" then P_ANESTHES = "No"; /*these were considered as
principal anethesia in original paper */
else P_ANESTHES = "Yes";

run;

/* Check N counts from table 1 header */
proc freq data = ssi;
table ssi;
run;

/* Check counts for P_ANESTHES */
proc freq data = ssi;
table P_ANESTHES;
run;

data ssi2;
set ssi (keep = FNSTATUS2 CaseID CPT SSI AGE2 BMI SEX ASACLAS DIABETES1 WHITE
BLACK P_ANESTHES WNDINF STEROID WTLOSS HXCOPD SMOKE HXCHF ASCITES DIALYSIS
HYPERMED DISCANCR HYPALBU TRANSFUS VENTILAT FSIND FSTD FSPD);

/* Create Age Groups */
if 18 <= AGE2 <= 29 then AGE18 = "Yes";
else AGE18 = "No";

if 30 <= AGE2 <= 39 then AGE30 = "Yes";
else AGE30 = "No";

if 40 <= AGE2 <= 49 then AGE40 = "Yes";
else AGE40 = "No";

if 50 <= AGE2 <= 59 then AGE50 = "Yes";
else AGE50 = "No";
```

```
if 60 <= AGE2 <= 69 then AGE60 = "Yes";
else AGE60 = "No";

if 70 <= AGE2 <= 79 then AGE70 = "Yes";
else AGE70 = "No";

if AGE2 => 80 then AGE80 = "Yes";
else AGE80 = "No";

/* Create BMI Groups */
if 0 < BMI < 19 then BMI19 = "Yes";
else BMI19 = "No";

if 19 <= BMI < 25 then BMI25 = "Yes";
else BMI25 = "No";

if 25 <= BMI < 30 then BMI30 = "Yes";
else BMI30 = "No";

if 30 <= BMI < 35 then BMI35 = "Yes";
else BMI35 = "No";

if 35 <= BMI < 40 then BMI40 = "Yes";
else BMI40 = "No";

if 40 <= BMI < 50 then BMI50 = "Yes";
else BMI50 = "No";

if BMI => 50 then BMI50p = "Yes";
else BMI50p = "No";

if BMI > 35 then BMI35p = "Yes";
else BMI35p = "No";

/* Create ASA Class Variables */
if ASACLAS = "1-No Disturb" then ASA1 = "Yes";
else ASA1 = "No";

if ASACLAS = "2-Mild Disturb" then ASA2 = "Yes";
else ASA2 = "No";

if ASACLAS = "3-Severe Disturb" then ASA3 = "Yes";
else ASA3 = "No";

if ASACLAS = "4-Life Threat" then ASA4 = "Yes";
else ASA4 = "No";

if ASACLAS = "5-Moribund" then ASA5 = "Yes";
else ASA5 = "No";

if SSI = "Ye" then AGEMEAN = "51.4 (14.5)";
if SSI = "No" then AGEMEAN = "51.4 (14.8)";

if SSI = "Ye" then BMIMED = "30.8 (26.0-36.8)";
if SSI = "No" then BMIMED = "29.0 (24.9-34.2)";
```

```
/* Categorical variables for groups */
if 18 <= age2 <= 29 then agegroup = "Age 18-29";
if 30 <= age2 <= 39 then agegroup = "Age 30-39";
if 40 <= age2 <= 49 then agegroup = "Age 40-49";
if 50 <= age2 <= 59 then agegroup = "Age 50-59";
if 60 <= age2 <= 69 then agegroup = "Age 60-69";
if 70 <= age2 <= 79 then agegroup = "Age 70-79";
else if age2 >= 80 then agegroup = "Age 80 or more";

if 0 < BMI < 19 then bmigroup = "BMI <19";
if 19 <= BMI < 25 then bmigroup = "BMI 19 to < 25";
if 25 <= BMI < 30 then bmigroup = "BMI 25 to < 30";
if 30 <= BMI < 35 then bmigroup = "BMI 30 to < 35";
if 35 <= BMI < 40 then bmigroup = "BMI 35 to < 40";
if 40 <= BMI < 50 then bmigroup = "BMI 40 to < 50";
else if  BMI >= 50 then bmigroup = "BMI >= 50";

run;

proc means data = ssi2 std mean median qrange p25 p75 prt kurtosis; /* table
1 shows mean and SD is collected for BMI, but actually median and range */
class SSI;
var BMI AGE2;
run;

proc means data = ssi2 std mean median qrange p25 p75 prt; /* table 1 shows
mean and SD is collected for BMI, but actually median and range */
var BMI AGE2;
run;

/* still struggling to get p-value between two means for BMI and AGE2 */

libname sse "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone";
data sse.demographics;
set ssi2;
run;
```

SAS code for Table 1 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 1: Demographic Chi-Square */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demo;
set capstone.demographics;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table1.pdf";

proc freq data = demo;
tables SSI*AGE18 SSI*AGE30 SSI*AGE40 SSI*AGE50 SSI*AGE60 SSI*AGE70 SSI*AGE80
SSI*BMI19 SSI*BMI25 SSI*BMI30 SSI*BMI35 SSI*BMI40  SSI*BMI50 SSI*BMI50p
SSI*BMI35p SSI*ASA1 SSI*ASA2 SSI*ASA3 SSI*ASA4 SSI*ASA5 SSI*SEX SSI*DIABETES1
```

```
SSI*WHITE SSI*BLACK SSI*P_ANESTHES SSI*WNDINF SSI*STEROID SSI*WTLOSS
SSI*HXCOPD SSI*SMOKE SSI*HXCHF SSI*ASCITES SSI*DIALYSIS SSI*HYPERMED
SSI*DISCANCR SSI*HYPALBU SSI*TRANSFUS SSI*VENTILAT SSI*FSIND SSI*FSTD
SSI*FSPD /chisq;
run;

proc means data = demo std mean median qrange p25 p75 prt kurtosis; /* table
1 shows mean and SD is collected for BMI, but actually median and range */
class SSI;
var BMI AGE2;
run;

proc means data = demo std mean median qrange p25 p75 prt; /* table 1 shows
mean and SD is collected for BMI, but actually median and range */
var BMI AGE2;
run;

/* MEA */

proc freq data = demo;
tables SSI*agegroup SSI*bmigroup SSI*ASACLAS SSI*FNSTATUS2 / chisq;
run;


ODS PDF CLOSE;
```

SAS code for Table 2 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 2: Multivariate Logistic Regression, pre-op */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demo;
set capstone.demographics;
run;

data demo2;
set demo;

if AGE18 = "No" then AGE18 = 0;
else if AGE18 = "Yes" then AGE18 = 1;

if AGE70 = "No" then AGE70 = 0;
else if AGE70 = "Yes" then AGE70 = 1;

if AGE80 = "No" then AGE80 = 0;
else if AGE80 = "Yes" then AGE80 = 1;

if BMI25 = "No" then BMI25 = 0;
else if BMI25 = "Yes" then BMI25 = 1;

if BMI50 = "No" then BMI50 = 0;
else if BMI50 = "Yes" then BMI50 = 1;
```

```
if SEX = "female" then SEX = 0;
else if SEX = "male" then SEX = 1;

if DIABETES1 = "No" then DIABETES1 = 0;
else if DIABETES1 = "Ye" then DIABETES1 = 1;

if WHITE = "No" then WHITE = 0;
else if WHITE = "Yes" then WHITE = 1;

if HXCOPD = "No" then HXCOPD = 0;
else if HXCOPD = "Yes" then HXCOPD = 1;

if SMOKE = "No" then SMOKE = 0;
else if SMOKE = "Yes" then SMOKE = 1;

if HXCHF = "No" then HXCHF = 0;
else if HXCHF = "Yes" then HXCHF = 1;

if HYPERMED = "No" then HYPERMED = 0;
else if HYPERMED = "Yes" then HYPERMED = 1;

if DISCANCR = "No" then DISCANCR = 0;
else if DISCANCR = "Yes" then DISCANCR = 1;

if VENTILAT = "No" then VENTILAT = 0;
else if VENTILAT = "Yes" then VENTILAT = 1;

run;


ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table2.pdf";

proc logistic data = demo2 order=freq descending;
class AGE18 AGE70 AGE80 BMI25 BMI50 SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF
HYPERMED DISCANCR VENTILAT;
model SSI = AGE18 AGE70 AGE80 BMI25 BMI50 SEX DIABETES1 WHITE HXCOPD SMOKE
HXCHF HYPERMED DISCANCR VENTILAT / CORRB;
run;

ODS PDF CLOSE;
/* Kaitlyn Kenig Capstone */
/* MEA Analysis */
/* Table 2: Multivariate Logistic Regression, pre-op */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demo;
set capstone.demographics;
run;


ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\MEA_Table2.pdf";

proc logistic data = demo order=freq descending;
class AGEGROUP (REF="Age 18-29");
```

```
class BMIGROUP (REF="BMI 19");
class SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF HYPERMED DISCANCR VENTILAT;
model SSI = AGEGROUP BMIGROUP SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF HYPERMED
DISCANCR VENTILAT / clodds=wald CORRB;
run;

ODS PDF CLOSE;
```

SAS code for Intra- and post op data:

```
/* Replicating https://doi.org/10.1016/j.ijsu.2018.11.013 *;
*                                                        *;
* eligibility                                            *;
* 2012 until 2015,
* thyroidectomy cases with the CPT codes
* 60210, 60212, 60220, 60225, 60240, 60252, 60254, 60260.
* age of patient was >=18 years.
* excluded all cases not classified as Class 1 (clean)*/

/* code for intra and post op characteristics */

libname thyroid "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access = readonly;

proc contents data = thyroid.thyroid;
run;

proc freq data = thyroid.thyroid;
table PODIAGTX ELECTSURG;
run;

/* Create dataset with intra and post op characteristics*/
data op;
set thyroid.thyroid (keep = CaseID CPT ORGSPCSSI WNDINFD SUPINFEC OPTIME
ELECTSURG OTHBLEED CDMI OTHSYSEP OTHSESHOCK OUPNEUMO REINTUB NFAILWEAN OTHDVT
PULEMBOL DEHIS URNINFEC CDARREST CNSCVA OPRENAFL RENAINSF RETURNOR TOTHLOS
PODIAGTX);

/* Create binomial variable for SSI */
if ORGSPCSSI = "No Complication" then SSI = "No";
if WNDINFD = "No Complication" then SSI = "No";
if SUPINFEC = "No Complication" then SSI = "No";

if WNDINFD = "Deep Incisional SSI" then SSI = "Yes";
if ORGSPCSSI = "Organ/Space SSI" then SSI = "Yes";
if SUPINFEC = "Superficial Incisional SSI" then SSI = "Yes";

/* Make elective surgery binomial */
if ELECTSURG = "Yes" then ELECTSURG1 = "Yes";
else ELECTSURG1 = "No";

/* make CPT codes binomial */
if CPT = "60210" then OP60210 = "Yes";
else OP60210 = "No";

if CPT = "60212" then OP60212 = "Yes";
else OP60212 = "No";
```

```
if CPT = "60220" then OP60220 = "Yes";
else OP60220 = "No";

if CPT = "60225" then OP60225 = "Yes";
else OP60225 = "No";

if CPT = "60240" then OP60240 = "Yes";
else OP60240 = "No";

if CPT = "60252" then OP60252 = "Yes";
else OP60252 = "No";

if CPT = "60254" then OP60254 = "Yes";
else OP60254 = "No";

if CPT = "60260" then OP60260 = "Yes";
else OP60260 = "No";

/* Make DVT binomial */
if OTHDVT = "No Complication" then DVT = "No";
else DVT = "Yes";

/* Post op diagnosis variables */
if PODIAGTX = "MALIGNANT NEOPLASM OF THYROID GLAND" then DX1MAL = "Yes";
else DX1MAL = "No";

if PODIAGTX = "NONTOXIC MULTINODULAR GOITER" then DX2NON = "Yes";
else DX2NON = "No";

if PODIAGTX = "NONTOXIC UNINODULAR GOITER" then DX3NON = "Yes";
else DX3NON = "No";

if PODIAGTX = "BENIGN NEOPLASM OF THYROID GLANDS" then DX4BEN = "Yes";
else DX4BEN = "No";

if PODIAGTX = "TOXIC DIFFUSE GOITER WITHOUT THYROTOXIC CRISIS OR STORM" then
DX5TOX = "Yes";
else DX5TOX = "No";

if PODIAGTX = "GOITER UNSPECIFIED" then DX6GOI = "Yes";
else DX6GOI = "No";

if PODIAGTX = "CHRONIC LYMPHOCYTIC THYROIDITIS" then DX7CHR = "Yes";
else DX7CHR = "No";

if PODIAGTX = "UNSPECIFIED NONTOXIC NODULAR GOITER" then DX8UNS = "Yes";
else DX8UNS = "No";

if PODIAGTX = "TOXIC MULTINODULAR GOITER WITHOUT THYROTOXIC CRISIS OR STORM"
then DX9TOX = "Yes";
else DX9TOX = "No";

if PODIAGTX = "NEOPLASM OF UNCERTAIN BEHAVIOR OF OTHER AND UNSPECIFIED
ENDOCRINE GLANDS" then DX10NEO = "Yes";
else DX10NEO = "No";
```

```
if PODIAGTX in ("NULL" "MALIGNANT NEOPLASM OF THYROID GLAND" "NONTOXIC
MULTINODULAR GOITER" "NONTOXIC UNINODULAR GOITER" "BENIGN NEOPLASM OF THYROID
GLANDS" "TOXIC DIFFUSE GOITER WITHOUT THYROTOXIC CRISIS OR STORM" "GOITER
UNSPECIFIED" "CHRONIC LYMPHOCYTIC THYROIDITIS" "UNSPECIFIED NONTOXIC NODULAR
GOITER" "TOXIC MULTINODULAR GOITER WITHOUT THYROTOXIC CRISIS OR STORM"
"NEOPLASM OF UNCERTAIN BEHAVIOR OF OTHER AND UNSPECIFIED ENDOCRINE GLANDS")
then DXOTHER = "No";
else DXOTHER = "Yes";

if PODIAGTX = "NULL" then DXUNK = "Yes";
else DXUNK = "No";

run;

data op2;
set op (keep = CaseID CPT OPTIME ELECTSURG1 SSI OP60210 OP60212 OP60220
OP60225 OP60240 OP60252 OP60254 OP60260 OTHBLEED CDMI OTHSYSEP OTHSESHOCK
OUPNEUMO REINTUB NFAILWEAN DVT PULEMBOL DEHIS URNINFEC CDARREST CNSCVA
OPRENAFL RENAINSF RETURNOR TOTHLOS DX1MAL DX2NON DX3NON DX4BEN DX5TOX DX6GOI
DX7CHR DX8UNS DX9TOX DX10NEO DXOTHER DXUNK);

run;


libname sse "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone";
data sse.opdata;
set op2;
run;
```

SAS code for Table 3 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 3: Intra- and post-operative characteristics Chi-Square */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data opdata;
set capstone.opdata;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table3.pdf";

proc freq data = opdata;
tables SSI*CPT SSI*ELECTSURG1 SSI*OP60210 SSI*OP60212 SSI*OP60220 SSI*OP60225
SSI*OP60240 SSI*OP60252 SSI*OP60254 SSI*OP60260 SSI*OTHBLEED SSI*CDMI
SSI*OTHSYSEP SSI*OTHSESHOCK  SSI*OUPNEUMO SSI*REINTUB SSI*NFAILWEAN SSI*DVT
SSI*PULEMBOL SSI*DEHIS SSI*URNINFEC SSI*CDARREST SSI*CNSCVA SSI*OPRENAFL
SSI*RENAINSF SSI*RETURNOR SSI*DX1MAL SSI*DX2NON SSI*DX3NON SSI*DX4BEN
SSI*DX5TOX SSI*DX6GOI SSI*DX7CHR SSI*DX8UNS SSI*DX9TOX SSI*DX10NEO
SSI*DXOTHER SSI*DXUNK/chisq;
run;

proc means data = opdata std mean median qrange p25 p75 prt kurtosis;
```

```
class SSI;
var OPTIME TOTHLOS;
run;

proc means data = opdata std mean median qrange p25 p75 prt;
var OPTIME TOTHLOS;
run;


ODS PDF CLOSE;
```

SAS code for Table 4 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 4: Multivariate Logistic Regression, predict infection */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demo;
set capstone.demographics;
run;

data opdata;
set capstone.opdata;
run;

proc sort data = opdata;
by CaseID;
run;

proc sort data = demo;
by CaseID;
run;

data table4;
merge opdata demo;
by CaseID;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table4.pdf";

proc logistic data = table4 descending;
class AGE18 AGE70 AGE80 SEX BMI25 BMI50 DIABETES1 WHITE HXCOPD SMOKE HXCHF
HYPERMED DISCANCR VENTILAT OP60212 OP60220 OP60225;
model SSI = AGE18 AGE70 AGE80 SEX BMI25 BMI50 DIABETES1 WHITE HXCOPD SMOKE
HXCHF HYPERMED DISCANCR VENTILAT OP60212 OP60220 OP60225 / CORRB;
run;


ODS PDF CLOSE;
/* Kaitlyn Kenig Capstone */
/* Table 4: Multivariate Logistic Regression, predict infection */
/* MEA Analysis */
```

```
libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demo;
set capstone.demographics;
run;

data opdata;
set capstone.opdata;
run;

proc sort data = opdata;
by CaseID;
run;

proc sort data = demo;
by CaseID;
run;

data table4;
merge opdata demo;
by CaseID;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\MEATable4.pdf";

proc logistic data = table4 descending;
class AGEGROUP (REF="Age 18-29");
class BMIGROUP (REF="BMI 19");
class SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF HYPERMED DISCANCR VENTILAT CPT;
model SSI = AGEGROUP BMIGROUP SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF HYPERMED
DISCANCR VENTILAT CPT / CORRB;
run;


ODS PDF CLOSE;
```

SAS code for Table 5 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 5: Multivariate Logistic Regression, post op complications */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data opdata;
set capstone.opdata;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table5.pdf";

proc logistic data = opdata descending;
class OTHSYSEP OTHSESHOCK OUPNEUMO REINTUB NFAILWEAN DVT PULEMBOL DEHIS
CDARREST OPRENAFL RETURNOR;
model SSI = OTHSYSEP OTHSESHOCK OUPNEUMO REINTUB NFAILWEAN DVT PULEMBOL DEHIS
CDARREST OPRENAFL RETURNOR TOTHLOS / CORRB;
```

```
run;


ODS PDF CLOSE;
```

SAS code for Table 6 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 6: Demographic Chi-Square, Sub group analysis */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demosub;
set capstone.demographics;
where CPT in ("60252" "60254");
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table6.pdf";

/* table header counts for subgroup SSI */
proc freq data = demosub;
table ssi;
run;

data demosub2;
set demosub;

if 18 <= age2 <= 29 then agegroup = "Age 18-29";
if 30 <= age2 <= 39 then agegroup = "Age 30-39";
if 40 <= age2 <= 49 then agegroup = "Age 40-49";
if 50 <= age2 <= 59 then agegroup = "Age 50-59";
if 60 <= age2 <= 69 then agegroup = "Age 60-69";
if 70 <= age2 <= 79 then agegroup = "Age 70-79";
else if age2 >= 80 then agegroup = "Age 80 or more";

if 0 < BMI < 19 then bmigroup = "BMI <19";
if 19 <= BMI < 25 then bmigroup = "BMI 19 to < 25";
if 25 <= BMI < 30 then bmigroup = "BMI 25 to < 30";
if 30 <= BMI < 35 then bmigroup = "BMI 30 to < 35";
if 35 <= BMI < 40 then bmigroup = "BMI 35 to < 40";
if 40 <= BMI < 50 then bmigroup = "BMI 40 to < 50";
else if  BMI >= 50 then bmigroup = "BMI >= 50";

run;

/*Chi square */
proc freq data = demosub2;
tables SSI*FNSTATUS2 SSI*AGEGROUP SSI*BMIGROUP SSI*BMI35p SSI*ASACLAS SSI*SEX
SSI*DIABETES1 SSI*WHITE SSI*BLACK SSI*P_ANESTHES SSI*WNDINF SSI*STEROID
SSI*WTLOSS SSI*HXCOPD SSI*SMOKE SSI*HXCHF SSI*ASCITES SSI*DIALYSIS
SSI*HYPERMED SSI*DISCANCR SSI*HYPALBU SSI*TRANSFUS SSI*VENTILAT SSI*FSIND
SSI*FSTD SSI*FSPD /chisq;
run;

proc freq data = demosub;
```

```
tables SSI*AGE18 SSI*AGE30 SSI*AGE40 SSI*AGE50 SSI*AGE60 SSI*AGE70 SSI*AGE80
SSI*BMI19 SSI*BMI25 SSI*BMI30 SSI*BMI35 SSI*BMI40  SSI*BMI50 SSI*BMI50p
SSI*BMI35p SSI*ASA1 SSI*ASA2 SSI*ASA3 SSI*ASA4 SSI*ASA5 SSI*SEX SSI*DIABETES1
SSI*WHITE SSI*BLACK SSI*P_ANESTHES SSI*WNDINF SSI*STEROID SSI*WTLOSS
SSI*HXCOPD SSI*SMOKE SSI*HXCHF SSI*ASCITES SSI*DIALYSIS SSI*HYPERMED
SSI*DISCANCR SSI*HYPALBU SSI*TRANSFUS SSI*VENTILAT SSI*FSIND SSI*FSTD
SSI*FSPD /chisq;
run;

proc means data = demosub std mean median qrange p25 p75 prt kurtosis;
class SSI;
var BMI AGE2;
run;

proc means data = demosub std mean median qrange p25 p75 prt;
var BMI AGE2;
run;

ODS PDF CLOSE;
```

SAS code for Table 7 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 7: Intra- and post-operative characteristics Chi-Square sub group*/

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data subopdata;
set capstone.opdata;
where CPT in ("60252" "60254");
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table7.pdf";

proc freq data = subopdata;
tables SSI*ELECTSURG1 SSI*OTHBLEED SSI*CDMI SSI*OTHSYSEP SSI*OTHSESHOCK
SSI*OUPNEUMO SSI*REINTUB SSI*NFAILWEAN SSI*DVT SSI*PULEMBOL SSI*DEHIS
SSI*URNINFEC SSI*CDARREST SSI*CNSCVA SSI*OPRENAFL SSI*RENAINSF SSI*RETURNOR
SSI*DX1MAL SSI*DX2NON SSI*DX3NON SSI*DX4BEN SSI*DX5TOX SSI*DX6GOI SSI*DX7CHR
SSI*DX8UNS SSI*DX9TOX SSI*DX10NEO SSI*DXOTHER SSI*DXUNK/chisq;
run;

proc means data = subopdata std mean median qrange p25 p75 prt kurtosis;
class SSI;
var OPTIME TOTHLOS;
run;

proc means data = subopdata std mean median qrange p25 p75 prt;
var OPTIME TOTHLOS;
run;


ODS PDF CLOSE;
```

SAS code for Table 8 values:

```
/* Kaitlyn Kenig Capstone */
/* Table 8: Multivariate Logistic Regression, pre-op for subgroup analysis*/

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data subdemo;
set capstone.demographics;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\Table8.pdf";

proc logistic data = subdemo descending;
class AGE18 AGE50 AGE80 SEX BMI25 WHITE HXCOPD SMOKE HXCHF;
model SSI = AGE18 AGE50 AGE80 SEX BMI25 WHITE HXCOPD SMOKE HXCHF / CORRB;
run;

ODS PDF CLOSE;
/* Kaitlyn Kenig Capstone */
/* Table 8: Multivariate Logistic Regression, pre-op for subgroup analysis*/
/* MEA Analyis */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data subdemo;
set capstone.demographics;
run;

ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\MEATable8.pdf";

proc logistic data = subdemo descending;
class AGEGROUP (REF="Age 18-29");
class BMIGROUP (REF="BMI 19");
class AGEGROUP SEX BMIGROUP WHITE HXCOPD SMOKE HXCHF;
model SSI = AGEGROUP SEX BMIGROUP WHITE HXCOPD SMOKE HXCHF / CORRB;
run;


ODS PDF CLOSE;
```

SAS code for TCA:

```
/* Kaitlyn Kenig Capstone */
/* TCA: Multivariate Logistic Regression, pre-op
Check for interaction between gender and current smoker */

libname capstone "V:\_Sandbox\Users\Kaitlyn Kenig\Capstone" access =
readonly;

data demo;
set capstone.demographics;
run;
```

```
ODS PDF file="V:\_Sandbox\Users\Kaitlyn Kenig\Capstone\TCA.pdf";

proc logistic data = demo;
class AGEGROUP BMIGROUP SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF HYPERMED
DISCANCR VENTILAT;
model SSI = AGEGROUP BMIGROUP SEX DIABETES1 WHITE HXCOPD SMOKE HXCHF HYPERMED
DISCANCR VENTILAT SMOKE*SEX / CORRB;
run;

ODS PDF CLOSE;
```